I am very happy to greet all the participants of the international scientific Conference on security issues and counter terrorism. The Conference Program contains a wide range of reports from various scientific fields — mathematical, computer, philosophical, political, and psychological. All of them are trying to find an answer to the same question: How to make the life of a present-day person more secure?

XXI century brought new global processes, challenges and threats to the mankind. A blistering development of information and telecommunication technologies leads to a mass information space conversion. There are some negatives and some positives. Scientific analysis of these "pluses" and "minuses" and a scientific justification of the secure behavior mechanisms in the information space are among the main objectives of our Conference.

I hope that two days of the Conference will be productive, they will create a new impetus for scientific research, and next year we will meet within the walls of the Moscow State University to discuss obtained results.

Academician V. A. Sadovnichy,
Conference Chairman,
Rector of Lomonosov Moscow State University.

Contents

Greetings

Plenary Talks

Subject Session "Mathematical Problems of Information Security"

Subject Session "Mathematical and software support of computer systems security"

Advanced Research Workshop-Round Table Discussion "Unconventional Information Warfare and the War on Terror: Critical Issues and International Cooperation"

Round-Table Discussion "Comprehensive Security in the Fuel and Energy Industrial Complex"

General Information

The Fourth All-Russian Scientific Conference "Mathematics and Security of Information Technologies"

Organizers: Lomonosov Moscow State University, Academy of Cryptography of the Russian Federation.

Advanced Research Workshop-Round Table Discussion "Unconventional Information Warfare and the War on Terror: Critical Issues and International Cooperation"

Organizers: Lomonosov Moscow State University, Cambridge Security Program, Cambridge University, United Kingdom with the help of the Security Council of the Russian Federation and with the assistance of the State University of New York and the Heinrich-Heine-University, Düsseldorf.

Round Table Discussion "Comprehensive Security in the Fuel and Energy Industrial Complex"

Organizers: Lomonosov Moscow State University, Joint-Stock Company "Gazprom" with the help of the Security Council of the Russian Federation.

Workshop-Meeting of the leaders of projects in the priority direction "Security and Countering Terrorism " of Federal Special-purpose Scientific and Technical Program "Research and Development on the Priority Directions in Development of Science and Technology"(FSSTP) during the years 2002–2006

Organizers: Lomonosov Moscow State University, Working group of the Scientific coordination Council of FSSTP.

Conference Co-Chairmen:

- V. A. Sadovnichy, Rector of Lomonosov Moscow State University;
- V. P. Sherstyuk (Security Council of the Russian Federation);
- N. N. Andreev, President of the Academy of Cryptography of the Russian Federation;
- S. K. Ushakov, Depute Chairman of Joint-Stock Company "Gazprom".

Advanced Research Workshop-Round Table Discussion "Unconventional Information Warfare and the War on Terror: Critical Issues and International Cooperation":

- Rafal Rohozinski, Co-Director, University of Cambridge, United Kingdom;
- V. V. Sokolov, Co-Director, Deputy Director of the IISI of MSU;
- A. V. Belyaeva, Coordinator, Citizens Initiative for Internet Policy Fund;
- Robert Gosende, Coordinator, State University of New York;
- Jan von Knop, Coordinator, Heinrich-Heine-University, Düsseldorf.

Co-Chairmen of Round Table Discussion "Comprehensive Security in the Fuel and Energy Industrial Complex":

- V. N. Pozharsky, Head of Security Service Department of Joint-Stock Company "Gazprom";

- A. I. Efimov, Head of Security Service Department of Joint-Stock Company "Gazprom";
- Yu. G. Popov, Head of Division of Security Service Department of Joint-Stock Company "Gazprom".

Organization Committee:

- V. V. Yashchenko, Chairman of the Organization Committee, Deputy Director of the IISI of MSU.
- V. N. Sachkov, Vice-President of the Cryptography Academy of the Russian Federation;
- S. F. Khomyakov, First Deputy Director-General of Security Service of Joint-Stock Company "Gazprom";
- V. N. Pozharsky, Head of Security Service Department of Joint-Stock Company "Gazprom".

Conference Secretariat:

- R. A. Sharyapov, Executive Secretary of the Organization Committee, IISI of MSU;
- V. I. Solodovnikov (Academy of Cryptography of the Russian Federation);
- Yu. V. Malinin (Academy of Information Systems);
- M. I. Anokhin;
- G. V. Baranova;
- T. A. Bratash;
- M. E. Semina;
- A. V. Sokolova.

Greetings

Rector of Lomonosov Moscow State University Acad. V. A. Sadovnichy

Allow me to open the International Scientific Conference on Issues of Security and Counter Terrorism. First of all I would like to introduce members of presidium — the main organizers of the conference:

- Assistant Secretary of the Security Council of the Russian Federation, Director of the Institute of Information Security Issues of Moscow State University, V. P. Sherstyuk;
- Vice-President of the Academy of Cryptography of the Russian Federation, V. N. Sachkov;
- Deputy Chief of the Security Service for the Joint Stock company "Gazprom", S. F. Khomyakov;
- Director of the Program on Security Studies of Cambridge University, R. Rohozinski.

In the presidium there are also heads of the Russian Federation governmental structures:

- Head of the Federal Agency on Information Technologies, V. G. Matyukhin;
- Vice-President of the State Duma Committee on Security, V. V. Dyatlenko;
- First Deputy Head of the Federal Service for Technical and Export Control, B. V. Nazarov.

In this conference hall there are scientists and experts from governmental structures, defense and policing services, institutions of Higher education and scientific organizations, from industry and business. Representatives of 15 foreign countries have arrived at the conference, among them, the universities of Cambridge, State of New York, Düsseldorf, the Academy of Science of China, the research centers of Canada and Switzerland, as well as specialists from the majority of the CIS countries. Russia is represented by leading scientists and experts on security issues from more than 60 institutions of Higher education, the Research Centers of Academies of Sciences and Industries, developers of technologies and means of security. The major feature of our conference is its interdisciplinary character. We shall discuss the pressing problems of humanity — security, counter-terrorism, and formation of the global in-

formation space. For solving these problems we need methods of various sciences — mathematics, physics, chemistry, biology, psychology, sociology, law, etc. Therefore, the conference program includes reports by representatives of various scientific schools. I can see in this conference hall many active participants of our Interdisciplinary Interdepartmental Seminar on Scientific Issues of Information Security which has been working at the Moscow State University since March, 2001 and includes participants from different structures. We have already conducted 19 sessions, published the most important reports last year, and issued them again for this conference, so that all the participants received a copy. An integral part of our conference is the fourth all-Russian conference "Mathematics and Information Technologies Security" (MITS-2005). It has already become a tradition to hold such conferences at the Moscow State University at the end of October. All participants today have received materials of the conference MITS-2004. The program of our conference includes about 40 mathematical reports in the fields of cryptography and computer security.

Dear Colleagues,

We perfectly well understand that the information space has no borders and therefore no country alone can solve the problem of providing a secure global information space on its own. The necessity of international scientific cooperation in the field of information security is mentioned in the Doctrine of Information Security of the Russian Federation approved by the President of the Russian Federation V. V. Putin in September 2000, and in the USA National Strategy to Secure Cyber space, approved by the United States President G. W. Bush in February 2003, and in the similar conceptual documents of a number of other countries. Scholars of the Moscow University actively participate in various international conferences and projects, together with foreign colleagues develop new ideas on providing information security. Recently, new mechanisms of the international scientific cooperation have started to function — NATO's Program "Security Through Science" and Scientific Board "Russia-NATO". This year Scientific Board "Russia-NATO" has defined five priorities of its work, one of them, the war on cyber-terrorism. The expert from Russia at this Scientific Board on the problems of cyber-terrorism is one of the Deputy Directors of the Institute of Information Security Issues of the Moscow State University. Last week in Brussels, on the basis of his report, our suggestions on the projects for 2006 were considered and basically approved.

Since November last year, in accordance with President Putin's assignment, the federal special scientific and technical program (FSSTP)

"Research and Development on Priority Directions of Science and Technology" has been realized. One of the priority directions is "Security and counterterrorism". The working expert group in this field is headed by V. P. Sherstjuk. Research managers of 13 projects on security and counterterrorism that are already being realized within FSSTP, will conduct a joint workshop at our conference.

On the initiative of the Joint Stock Company "Gazprom" management we will conduct a round table "Comprehensive Security in the Fuel and Energy Industrial Complex" in the framework of this conference.

We shall witness very interesting reports, discussions, section work, and round tables. I would like to thank all guests who have arrived to the Moscow University for our conference. I would like to thank all participants of the conference for accepting our invitation to contribute to interesting discussions and interesting sessions. Allow me to wish successful work to our conference and to declare it opened.

Head of the Russian Federation Agency for Information Technologies V. G. Matyukhin

Dear Colleagues,

Allow me to greet you on behalf the Russian Federation Agency for Information Technologies and Scientific Society.

Issues of information security are multi-dimensional and sometimes acquire new facets, the ones that we have not yet fully realized. As we know, it has become a general policy to actively introduce information technologies into the process of ruling over a country and its economy with the purpose of dramatic increase of management efficiency.

The Federal Agency has developed principles of the information infrastructure necessary for creating "the electronic government". The matters that lie in its basis are connected with creation of a unified cyber space and the effective realization of the digital signature, questions of creation memory on the basis of distribution of active storages, and creation of a unified national system of identification.

All the three tasks I consider as the cores. Studying of these very fields, in our opinion, creates the information interaction indispensable for the unified management, and for the legal significance of control systems that are being created. The formation of certifying centers system is being actively conducted. In June there was a presentation of the root certifying center that had been developed in the framework of the "Electronic Russia" program. And I hope that the scientific and technical questions have all been solved. The ones that remained unsolved were matters of a regulatory legal character and to deal with them we need more the political will than intellect.

As to matters of creation the distributed storages, we are working now jointly with the Moscow State University on the development of a two-cluster system model, with the purpose of defining main directions of work, and preparation of some program documents. As to information space, "social cards" are well familiar to everybody: schoolchild's card, student's card. At the present moment we are experiencing a real boom in the development of the identification elements space. It has to be

organized and regulated. There is a need to understand access rights, basing on the safety issues, and certainly taking into consideration the necessity of the biometric passport creation.

All of these three tasks which have to be solved for realization of the electronic government in the form of rules, regulations, corresponding architecture, have one weak point — it is the question of information security. Additional work has to be done in all three directions, work of fundamental character. The ones of the paramount importance that I would identify are the questions of information security organization, territorial distribution of information storages, and the question of access to information with security accreditation. The systems are of a national level, and the level of protection should correspond. I would like to wish participants of the conference fruitful exchange of opinions, and development of uniform positions in solving such global issues, as the issues of information security.

First Deputy Director of Federal Service for Technical and Export Control B. V. Nazarov

Dear participants and guests of the conference,

Allow me on behalf of the Director of Federal Service for Technical and Export Control S. I. Grigorov to greet you warmly and to wish you successful work.

Today one of most quickly developing sectors of world economy are information technologies. The peculiarity of the present stage of information technologies development is their extraordinarily high integration into all spheres of human activity. Along with the unequivocal benefits brought up by these processes, there also arise a whole spectrum of new threats, challenges and risks to national security. These new global realities make traditional forms and modes of work employed by security and law enforcement structures outdated. For essentially new approaches, other levels of professionalism and equipment are required. It is impossible to struggle against terrorism and crime of the 21st century by methods and means of the 20th century. Thus, consolidation of efforts of both state, and non-state actors is required. In these conditions, it is obvious that maintenance of information security obtains vital importance. Special emphasis in modern conditions is given to responding to the threats connected with terrorist actions. This is the main topic of our conference. The realities of today show, that while solving the matters of state's information security, we must put safe functioning of information-telecommunication systems of crucial objects in the forefront. It is logical because unauthorized influence on the specified systems can have the most catastrophic consequences: cause disorganization of government; impair significant damage; lead to technogenic and ecological accidents, including ones with a great quantity of human casualties and severe material damage. These challenges and threats have a global, international character. Practically all industrially developed countries pay utmost attention to questions of information security. An example

of it is the USA "National Strategy for Critical Information Infrastructure Protection" proceeded with "The National Strategy of Cyberspace Protection". In the Russian Federation the base for the activities in this sphere are "The Concept of National Security" and "The Doctrine of National Security". One of the agencies which implement the requirements of these documents is the Federal Service for Technical and Export Control, which I represent.

It is necessary to note, that within the framework of the administrative reform conducted under the supervision of President Putin, the authority and functions of the Service have been expanded in view of new threats in the field of information security. Priorities in our work for immediate perspective are the formation of a regulatory legal and methodical framework concerning maintenance of key systems of the informational infrastructure, development of a typical model of security threats, as well as the development of frequent models of threats for concrete types of categories of information systems, and on their basis — working out the requirements and norms on key systems protection.

It is truly a pleasure for me to note the international format of our meeting. As the basis for our international interaction in addressing the threats to information security serve intergovernmental agreements with the republics of Kazakhstan, Belarus and Ukraine on cooperation in the field of information protection.

In conclusion, allow me to express belief that the results of the conference will give a new impulse to the development of conceptual approaches and practical measures in solution of security and counterterrorism problems.

Vice-President of Academy of Cryptography of the Russian Federation V. N. Sachkov

Dear colleagues, Dear Guests,

Allow me on behalf of the Academy of Cryptography to welcome warmly all the participants of the conference and to wish you successful and fruitful work, and new creative contacts in the field of cooperation for providing information security.

Active development of telecommunication information systems in the last decades has played an important role in defining the ways of the further development of modern cryptography. The computerization of communication and management systems has essentially complicated, and has put forward a number of new problems for information protection. The character of threats has become significantly more complicated due to the increase in the opportunities of access to control facilities, processing, transition of information, and variety of means of influence, with the purpose of changing the essence and disrupting normal work of systems. In this connection, along with traditional means of cryptographic protection, there occurred a necessity to create new methods of providing information security.

The issues of providing integrity and identification of information correspondents, protection of communication streams in the computerized telecommunication systems have demanded essential expansion in the subject of cryptography and its methods of research.

Under the conditions of aggressive computer and program environment and the destructive influence of virus programs special attention in research work is now paid to the development of protective methods for the computerized information systems. A revolutionizing role in the development of modern cryptography was played by the occurrence of asymmetric enciphering, which has essentially enriched traditional methods of symmetric enciphering.

It is necessary also to note the further development of the symmetric enciphering due to the substantial growth of information processing speed, sophistication of output schemes and key distribution, etc. New opportunities for steadfast cryptographic enciphering of information were created due to the appearance of quantum cryptography.

Research in this area is being conducted in the Russian Academy of Sciences, at Moscow State University and in the Academy of Cryptography of the Russian Federation. Traditionally a great role in the development of cryptography is played by high-efficiency computing systems which substantially define parameters of perspective cryptographic algorithms and means of information protection.

A prominent aspect in the development of cryptography as a scientific field of knowledge is the adjacent area of the humanitarian issues of information security. At the initiative of the Rector of Moscow State University, V. A. Sadovnichy, this area has been actively developed in the last several years at Moscow State University. The interdisciplinary seminar, established under the supervision of V. A. Sadovnichy, is an important forum for discussing these questions that has played a significant role at this conference.

Certainly, the most pressing question of today is the struggle against international terrorism, including the counteraction to cyber- and information terrorism. Discussion of these matters with fellow scholars from different countries will help to search for the ways of counteraction to this international evil, and also will promote consolidation of scientists and experts working in this area.

In conclusion, allow me to welcome once again cryptologists and other experts in the field of information security, scientists, scholars, and guests of the conference. I would like to wish everybody success in the work at the conference and round tables, and to express hope for further international cooperation in this area.

Vice-President of the State Duma Committee for Security V. V. Dyatlenko

Dear Participants and Guests of the conference,

On behalf of the State Duma Committee for Security allow me to welcome you on the occasion of the opening of our conference, one of the most significant scientific forums, devoted to the issues of security and counterterrorism.

The beginning of the 21st century is marked by serious intensification of terrorist organizations' activity. In the opinion of some experts, in the modern world there are about 500 terrorist organizations and groups of various extremist orientations. We have faced challenges, to which we do not possess yet a symmetric answer. Terrorism has stepped into the hi-tech sphere; therefore issues of providing information security are addressed with special attention. Danger of terrorism consists also in the fact that organizing its subversive activities in the hi-tech sphere, it provokes authorities to reciprocal violence that in turn destabilizes democratic institutes of the society and thus promotes violation of human rights and freedoms of a citizen. One of the primary tasks that are being carried out now by legislators is the creation of a legal regulatory framework that will provide optimum coexistence of coercion through force and economic measures. We believe in the principle, that effective struggle against terrorism in all its facets is possible only in case that the civil society joins it.

For this purpose a uniform national strategy of counteraction to terrorism and providing information security should be developed. In this difficult task we rely on the methodological help of science.

Allow me to wish you fruitful work and new creative decisions.

First Deputy General Director of Joint Stock Company "Gazprom" S. F. Khomyakov

Dear colleagues,

Allow me on behalf of the "Gazprom" management to congratulate participants, guests and organizers of the conference on its opening. A practical scientific conference, devoted to security issues and countering terrorism, from our point of view, is extremely urgent at the present moment.

This is the reason why "Gazprom" has initiated a special section which would give an opportunity to discuss a burning question for us. I would like to focus on two aspects of this question. "Gazprom" is a huge organization which has stretched its enterprises from the Arctic Ocean up to our western borders.

Matters of information security are urgent for us in a very real, practical plane. Management of information streams in all our huge facilities is an essential object for our protection, and therefore we hope that scientific approaches which will be developed here will help us to work correctly in this direction.

The next aspect, on which I would like to focus specially, is double folded: the security issues in general and issues of countering terrorism in particular. I have to express here a practical point of view, in spite of the scientific character of the conference. I have to speak about economic parameters here. In the coming New Year, shares of "Gazprom" will be quoted on foreign markets, and western investors will possibly purchase them. It is important for them, how well we can manage our risks, and one of the essential risks is terrorism. I could give you sad statistics of the unfortunate increase of terrorist activity on the pipelines, on the pumping stations, and on our other holdings. This is the sad reality. In this connection we would hope for scientific conclusions at this conference that would help us to predict and reduce further occurrences of terrorist risks in the future.

I would like once again to congratulate the participants of the conference on its opening and to express hope for the further fruitful cooperation.

Head of Division of Security Service Department of Joint Stock Company “Gazprom” Yu. G. Popov

Ladies and Gentlemen,

Allow me, on behalf of the Chairman of Gazprom, to greet the participants and the organizers of this conference, and to read the welcome address. Carrying out such an event is very important for uniting representatives of law enforcement, the gas, oil and chemical branches, science and industry, and security services. It is very important to mention that among other topics, we will dwell on such subjects as the legal control and legal procedure of legislative guideline development and the conceptual basis for the development of this particular branch, taking into consideration domestic and foreign approaches. The conference will allow us to share scientific/technical and organizational approaches from different organizations, generalize accumulated experiences, and view assumed methods of problem solution. Everything mentioned above will, in their part, contribute to the creation and development of a new unified security system for different installations, mechanisms of state protection assessment, and this conference will promote further search and realization of new ideas and methods of protection. It goes without saying, that strengthening mutual understanding and interaction between government authorities, non-government organizations and business circles for fighting against terrorism favors the growth of homeland industry, business, and economy as a whole.

One of the most important courses for realization of new policy in the field of increasing protection of the critically important objects is the creation of new systems, designed to provide security to these objects under the complex conditions of threat aggravation, rising danger of natural disasters, and other technical threats. This includes: legislative base improvement, liability for breakage (including criminal, managerial, civil and disciplinary liability of persons guilty of set standards), requirements and regulations violation, failure to carry out necessary work, and property damage (under property damage we understand objects

and persons, members of society); authority, responsibility distinction and public authorities interaction; organization of patterns of ownership; conceptualized elaboration of coordinated scientific and technical policy in the field of creation new security technologies and methods of their application; formation of critical object monitoring systems; installation of a comprehensive warning system, threat suppression and negative manifestations; control of object's defense condition; creation of licensing system, meaning obligatory licensing, certification, and avowal for potentially dangerous objects and critical objects; inspection measures of dangerous and critical objects; creation of all-Russian roster of industrial installations depending on their social hazard and terrorist vulnerability based on unified normative and legal documents, criteria, classification, etc.; development of criteria and mechanisms for assessment of realization and providing minimal risks; risk management and stabilization of potentially dangerous objects; prevention of terrorist activities and sabotage; elaboration of tools and defensive means, as well as improvement and modernization of the standard technical base; improvement of selection, training, certification system, access of managers and specialists to potentially dangerous and critical objects based on today's training and technical elaborations including elements of psychological preparation and comprehensive influence of the human factor on the security level.

We hope that during the round-table discussion we will examine these approaches, ideas, and methods, designed to solve the mentioned issues.

Plenary Talks

Relevant Issues of Global Information Infrastructure Security

V. P. Sherstyuk, A. A. Streltsov

Dear participants of the conference,
Dear Guests,
Ladies and Gentlemen,

First of all I would like to express deep gratitude to all of you who accepted our invitation and have gathered today in this remarkable building which in accordance with serious problems that are being solved here, is justly referred to as the Intellectual center.

Today counter-terrorism and security issues have become a major focus of Russian and World policy. Developing methods of the effective solution to these problems have formed an independent scientific study. To develop such methods, the efforts both of home and foreign specialists are required, humanitarians and experts in science. It is difficult to overestimate active participation in this work at classical universities, and other scientific and educational institutions. They are capable of uniting the efforts of various scholars, of creating new understanding of terrorism as a phenomenon in our lives, and of offering methods to increase the efficiency of counterterrorism.

Lomonosov Moscow State University has accumulated certain experience in the organization of such cooperation in the field of information infrastructure security. Under the supervision of the University's rector, Academician V. Sadovnichy, the University united scientists and interested federal executive authorities, built cooperation with a consortium of institutes and academies in NATO countries that are engaged in studying security issues, and established creative contacts with the State University of New York.

Last week an agreement on cooperation was signed with the Henry Heine University (Düsseldorf). From this point of view we find very useful the discussion which took place in the framework of the conference "Security and Private Life in the Information Society" which was held last week in Düsseldorf. Many participants from that conference

are present today in this hall. I am glad to welcome them, I am glad to welcome Doctor von Knop, one of organizers of this conference. I hope that we will find an opportunity to continue discussions concerning counter-terrorism. This speech has a more limited purpose, that being the priority directions in international cooperation in the field of global information infrastructure security.

For a long time, this issue has been the center of attention for Russian political leadership. It has found reflection in the Doctrine of Information Security of the Russian Federation, approved by President Putin in September, 2000. In this document, the information infrastructure security is among the components of national interests in the information sphere. Searching for ways to solve this problem becomes especially important in connection with Russia's participation in forming the global society stipulated by the Okinava Charter of Global Information Community (2000).

It is understandable that the global infrastructure has become an important factor in society's development. In the economic sphere, it promotes the formation of a new sector connected with computerization of knowledge, creation of modern information technologies, development of the information industry products and devices, expansion of the electronic trade sphere, and structural reorganization in the labor market.

According to available data, over the past years internal current expenses for research and development have increased by 20–25% annually. The volume of communication services increases annually by more than 30% and in 2004 was 3% of the Gross National Product. Export of Russian technologies to foreign countries increased: in 2003 it was 23 billion rubles, in 2004 it is already more than 30 billion rubles.

According to certain data, the volume of information technologies in the market will reach 600 billion rubles. In the spiritual sphere, the information infrastructure promotes a wider realization of a person's rights and freedoms in the field of information activity, including creation and distribution of mass information, and freedom of expressing opinions, of religious beliefs, ideas, and speech.

In some estimations, by 2010 the number of active users of internet network in Russia will be more than 26 million people. Almost all federal bodies of the government are already represented in this system. Paperless technology is being realized in the function of the government, based on the global information infrastructure and the means of digital signature, which have been intensively introduced in Russia.

In the social sphere, the information infrastructure actively influences the development of educational system and social security. In the political sphere, it represents the subjects of political life and qualitative new conditions for public communications development, for conducting political struggle and for realization of the executive power and certain functions of the state, including its security and defensive capabilities. Thus, the role of the global information infrastructure in realization of citizens' interests in both the society, and the state, grows. On the other hand, the factors that create a threat to information infrastructure security multiply. Among core threats, it is necessary to list terrorism.

Terrorism, more and more, is becoming an ordinary part of life. The war that terrorism started is a special war, conducted by non-state formations against a nation, a society, and a state, for the achievement of certain political objectives. In this war, the opponent has significant financial and manpower resources; actively uses freedom of information guaranteed to citizens of democratic states for realization of terrorist acts; organizes actions both from territory of home country, and from territory of other countries. Targets of terrorist attacks can be the information and telecommunication systems of crucially important objects: energy, transportation, financial, and other infrastructures in the society, hydro-electric and atomic power stations and other ecologically dangerous industrial enterprises, etc. This problem was discussed during our conference at the round table on the topic "Comprehensive Security in the Fuel and Energy Industrial Complex".

An essential threat to the security of the global information infrastructure is represented by computer crime, which objectively creates conditions for spreading terrorist activity to the information sphere. Modern information technologies give criminal organizations and individuals significant opportunities to access information and telecommunication systems, to use information resources stored and processed in these systems, and to carry out illegal actions based on the information received. As the result interests of a person, society and the state can all suffer. So, in the experts' opinions, general damage to the economy caused by computer crimes in 2004 has almost doubled in comparison with 2003, and has reached more than $400 billion US. Taking into account the high level of latency of computer crimes, there are grounds to believe that the real damage makes a much greater sum.

Countering security threats to the national information infrastructures, which are segments of global information infrastructures, is the primary concern of governments. At the same time, some aspects of this problem are already the center of attention in the international community. This fact was confirmed during the discussion of the Russian draft Resolution on International Information Security by the first commit-

tee of the UN General Assembly, which was completed last week. This project provides for, in particular, the continuation of threat analysis in the information security sphere and possible joint measures by the international community on their neutralization. 163 countries expressed support for the project.

Fruitful discussion between scientists of the Russian Academy of Sciences and national academies of the USA concerning counteraction to the threat of computer terrorism has begun the process of strengthening the scientific cooperation between the two countries. The terrorist acts that have occurred recently, show the necessity to further expand this cooperation. Terrorism as a phenomenon in the modern world is still insufficiently studied, but it is possible to identify the general features of the threats generated by it.

First, terrorist acts can be carried out by both citizens of the state under attack, and citizens of other states. Second, the time and place of terrorist acts are seldom predictable. Third, unlike other criminal actions, acts of terrorism are committed, as a rule, for achieving certain political objectives.

In these conditions creating a system of counteraction to terrorism in general, and to computer terrorism in particular, demands significant efforts both from separate states, and the international community as a whole. In the process of global information infrastructure development, it becomes clearer that the success of each country in protecting the national sectors of this infrastructure, to a certain degree, will depend on the other countries' successes in this area.

The life of the international community as a whole, and each country separately, will be more protected from the threat of computer terrorism when effective forms of various state interactions are found, methods of responding to the threats, means of revealing, suppressing and eliminating the consequences of these threats are developed and introduced. It is possible to specify some basic directions for countering the computer terrorism threat.

The organizational direction deals with the development and realization of a system of special actions by the executive branch of government and citizens. They should be directed to hinder the realization of computer terrorism acts and increase efficiency of investigative actions concerning the facts of preparation and realization of such acts, and also on minimization of the negative consequences inflicted by these acts. One such important action is the perfection of a system of requirements to information security in information and telecommunication systems of crucially important to social infrastructure objects, and their certification. Another no less important action is the development of similar requirements for modern information technology products and the creation of a voluntary certification system for these products. It would increase the level of trust between developers and users of these technologies. No less important is the development of an audit system of state and non-state information and telecommunication systems and crucially important objects of infrastructure.

The legal direction for counteraction to the computer terrorism threat consists in the development and realization of legal mechanisms of public relations regulation connected with the exposure to the threat. These mechanisms should assist, on one hand, the timely revelation and neutralization of the threat, decreasing the social danger of its consequences, and on the other hand, should not reduce the state's guarantees of an individual's and a citizen's rights and freedoms. Within the framework of the given direction, they should improve the national legislation concerning application of modern information technologies in the various spheres of the society's activities and the state's functions, development of information infrastructure and executive authorities' activity on counteraction to threats of computer terrorism, and interaction of state and non-state organizations in realization of this activity. Along with this, we have achieved certain results, specifically the interstate agreements concerning cross-border terrorist acts. The agreements deal with revealing the preparatory stages of terrorist acts against crucially important objects of national information infrastructures, and taking investigatory actions in case such acts are realized.

Technical direction for the counteraction to the computer terrorism threat is connected with the systematic use of protection means concerning the information of individuals, commercial and non-commercial organizations, and governmental structures.

The personnel direction of counteraction to computer terrorism is connected with the development of educational programs capable of training qualified experts in the legal, organizational, and technical fields. In the Russian Federation, special attention is paid to this direction of activity. Today more than 100 institutions of Higher education carry out training in countering computer terrorism, providing about 2500 experts annually. At the same time, the demand for such experts in both state and non-state organizations, by estimations of the Ministry of Education, is approximately 5000 people.

An important reserve of efficiency in the increase of joint activity on counteraction to computer terrorism would be the expansion of in-

ternational cooperation in this area, creation of common educational standards for certain specialities.

Development of international cooperation in the field of counteraction to computer terrorism will be hardly an easy matter. Certain problems are created by differences in political objectives, social and economic situations, and the absence of a uniform terminological base. Nevertheless, there seems to be no reasonable alternative to international cooperation.

In conclusion, I would like to thank you once again for your participation in the work of our conference. I hope that the forthcoming discussion will allow us to deepen mutual understanding in the field of security, and to work out concrete agreements on joint steps that need to be taken to strengthen the security of our countries, Europe, and the whole world.

The International Cooperation in the Field of Information Security

A. S. Kremer

Dear members of the presidium, Dear participants of the conference,

One example of the international cooperation in the field of providing information security on info-communicational networks and systems about which I would like to inform you, is a new project of a research commission on information security by the International union of telecommunication. The project is called “Basic Level of Information Security of Network Operators”. The decision to open the project was made under the initiative of the Russian Federation administration of communication on April 8th, 2005. The project is categorized as “highest priority”. It should develop a series of recommendations by the International Union of Telecommunication and should be completed in 2008.

International standardization is an important factor in the representation of interests for regulators, operators and equipment manufacturers. Therefore, it is no wonder that besides Russian specialists, the project was joined by representatives of the USA, Canada, Japan, China, Korea and Brazil. Let me familiarize you with the 5 primary tasks of the project, and also with the working plans for the current and next years.

The first task: “Base level” means use of one-type criteria by operators of interacting networks. Use of criteria should depend on the network type (for example, wire or wireless) and on the accepted mode of regulation (for example, the demanded safety level is one of the conditions for granting the license or a condition of connecting to the other operator’s network).

The second task: The developed series of recommendations should be a reflection of balance among prospective operator expenses, expected result, and an opportunity of its estimation; balance of the national right and the established network practice of self-regulation; balance of

interests of users, operators and regulators. A series of recommendations should promote the creation of a demonstrative base and effective application.

The third task: It is necessary to organize interaction among the experts of the International Union of Telecommunication and experts of other international organizations working in the field of standardization such as ISA, ETC, ATF and others. The standard is only operating when there is the appendix. For providing state structures and business with practical recommendations on technical equipment, the existing standards will be analyzed within the framework of the project from the point of view of their efficiency, the current status and prospective directions of modernization.

The fourth task: Harmonization of the various languages in which experts speak about security is of utter importance. These are the languages of lawyers, insurance agents, appraisers or evaluators, technologists, law-applying bodies and standards makers.

The fifth task: The basic criterion of ensuring information security must be the protected object's ability to perform its basic functional tasks during infringement.

At present moment it is possible to speak about evolutionary changes in the conceptual models of setting and solving problems in electro communication. The basic attributes of change in the paradigm of communication are:

- the separation in network architecture of logic formation and the essence of services from the level of the information transfer;
- gradual transformation of voice services from being the basic product of communications into one of many appendices working atop IP;
- gradual transformation of communication services from a directly consumed product into a means of access to non-telecommunication services;
- and transition from a *providing* telecommunication branch to a system creating info-communication branch.

Changes to the communication paradigm should find reflection in the changes to the paradigm of the information security system as an integral component of communication networks. It is expressed in transition from protection of the information to providing network information security maintenance, in translation of information security from a technical problem into a problem of the society as a whole. Confirmation of the necessity of such translation is the development of an internet network in which each user, to a certain degree, becomes an operator.

Now some words about the organizational work. For carrying out the project within the framework of the research commission on information security by the International Union of Telecommunication, in October of current year the so-called focus-group was formed. This organizational form, in conformity with the rules of the International Union of Telecommunication, gives an opportunity of participation in the project to the organizations, non-members of The International Electro Communication Union (IECU). The nearest working plans in the framework of this project include:

- preparation of the report on the existing state of affairs in the investigated area;
- ideas on the structure of the base level security for various types of communication operators;
- carrying out focus-group seminars for consolidation of objectives with participation of the interested parties (the seminar will take place in March of the next year in Moscow);
- work on coordination and distribution of questionnaires on behalf of the sector for standardization of the International Union of Telecommunication, with the goal of collecting information;
- preparation of a review on the poll's results;
- the analysis of the base level security standards, and possible gaps in the identified standards;
- the description of business-appendices pertaining to the requirements of the base level security;
- the analysis of the evaluation methods for information network security;
- and last, the formation of a preliminary variant in base level security and suggestions for the recommendations release by the standardization sector of the IECU.

Unconventional Information Warfare: Challenge Determination

R. Rohozinski

Rector Sadovnichy,
Fellow scholars,
Ladies and Gentlemen,

It is indeed a rare pleasure for me to be here at Lomonosov University presenting on this topic. Indeed our two universities, Cambridge and Lomonosov, have had the rare opportunity of breeding the generations of leaders in both government and security sectors, on whose shoulders it has fallen to deal with issues of both national security and national development.

Before I start, a few words about the Cambridge Program. In the last 5 years, as the world has increasingly picked up its tempo of globalization new security actors forced countries east, west, north, south to address new forms of threats. It was recognized within the security services and other responsible agencies in the west, that the competencies that have served them well during the Cold War in terms of assessing the nature of the threat and reacting accordingly were no longer appropriate to actors, which were neither state based nor organizations at all, and yet whose capabilities approached those of nation states in the degree of fear and the destruction that they can cause. As a result of the need to help reconceptualize the nature of security in this new era, the University of Cambridge, together with the government and its other partners constituted a new program, which will help "think outside the box" about these issues and help guide those in responsible organs in formulating appropriate policy responses. In the last four years the Cambridge Security Program has convened cross agency meetings which included the Departments of Defense, Internal Security as far as Intelligence to help officers within these institutions and policy makers understand better the nature of the threat that they were looking at. The task was to help them understand the context in which the threat had emerged rather than helping them with their daily work, which is the operational and

they knew much better. The Advanced Network Research Group, which is the part of the Cambridge Security Program specifically, looks at the intersection between technical networks and technical systems such as the Internet, global communication networks, financial information networks and process control networks and what might be called human agency. Basically what we are interested in exploring is how individuals motivated in different ways can and do use these networks to achieve their political aims which may be contrary to the interests of security of the state and to the international order. Our Program works internationally. We work with UK institutions as well as US and others. And a part of acting as facilitator for discussions we are also involved in active research in a number of areas including information operations both offensive and defensive as well as the evolution of boarders and boundaries in cyberspace.

It is truly an honor for me that we have been able to hold a NATO Advanced Research Workshop jointly with a Conference on Information Security. This is basically for three reasons, I would say. First of all, NATO workshops generally have been very closed, small events where a few experts would get together in order to discuss issues. The fact that we have been able to convince a NATO Science Program, and here I would like to thank Bryan Heet, to widen the scope of the workshops, so as to let us be able to include a much broader range of actors, which include government, academia as well as the other sectors. I think, it says something about the importance of all three sectors for looking at this particular topic.

Secondly, I think, we should be impressed about the fact that we have managed to have a topic such as this as part of NATO workshop which is composed of both the Russian Federation as well as its partners in the west. Information security is a very very difficult topic. Information security when combined with building transnational threats is even more difficult. It is difficult because the threat of terrorism gets at the very basis of physical security. It is difficult because the information security gets at the very basis of what is considered to be national prerogative and national security. It is also very difficult because the flip side of information security is information warfare and these issues touch upon what is probably the closest held and least best kept secret of the state which includes the capacity for strategic signals intelligence as well as offensive information operations. For that again I would like to thank Rector Sadovnichy, professor Sherstuk, Streltsov, Sokolov for their hard work in making this event possible.

Returining to my talk. I chose the title "Unconventional Information Warfare" very deliberately. Although most people turn their attention to the "National Strategy for the Protection of Critical Information Infrastructure" as being the guiding document for information security. This is not necessarily the case. This document has its antecedents in the previous exercise which was held back as far as 1991. I refer to 2000 Defense Science Board Report, recently made public, which in fact looked at the nature of unconventional threats to the national security of the United States in this case. This report identified two specific threats. First was the use of nuclear weapons by non-state actors in undeclared fashion. Second was the use of cyber warfare effective of either destruction or denial of the national information infrastructure also by non-state actors. Both of these substudies effectively framed and were responded through both the "National Strategy for Critical Information Infrastructure Protection" as well as the "National Security Strategy of the United States". What is interesting about this 2000 Document is that it held an equivalent between the destructive and disruptive power, between weapons of mass destruction and cyber weapons, considering these to be of equal importance and equal danger. So why is this important? This is important because since the 2000 Document there've been a major undertaking in terms of building the capacity to deal with the threat of what was known both as conventional information operation as well as defense information security. This effort has largely occurred quietly and under the level of hearing and was led by the Department of Defense rather than the Department of Homeland Security. It has looked at establishing both the capacity for defending against catastrophic cyber attack as well as creating a capacity to proactively and preemptively take out the possible causes and sources of these attacks whether these are nation-states or, as it was conceptualized in the 2000 Document, non-state actors. So this capacity for creation of the conventional information security information operation capacity has increased. And if we look at the differential results between two scenarios that were around: the first, a series of no warning exercises, run in 1997–1998, known as Eligible Receiver, in which a red team of intruders attempted to take control of a number of critical information systems including telephone exchanges and power lines, and the results of the simulation exercise run in 2005 Solid Horiser, two things are clear. One, the nature of the actor which was hypothesized as being a threat was the same, it didn't change. Despite the emergence of new threats in 2001 still the primarily threat was non-state based hackers looking to a political end, taking down the national information infrastructure. But secondly, the other point, was

that the capacity to deter, identify and deal with the consequences of such an attack has exponentially improved between 1997 and 2005.

However, despite the existence of both elaborated defensive and offensive information operations strategy developed capacity, not just stated in the document. Still, the ability to be able to apply successfully this capacity to the new forms of security actors, which were primarily hypothesized as being motivated non-state actors, has not changed. Moreover, there was also recognition that the form of information warfare used by these actors was somehow different, unconventional. And yet the existing strategy mechanisms for dealing with these actors, for reasons we shall describe in this talk, seemed actually to be decreasing. In other words the increasing capacity seemed be leading to digressive effect. Moreover, some of the defensive measures that were being adopted were having a corrosive effect both in terms of the social benefits however effectively the networks were tried to be defended as well as coming at large economic cost.

So what are the key questions that this raises? One: if the main concern, the main focus of these strategies are the new security actors and yet we are not successful with it, are we actually preparing for the right threat? This cyberterrorism which was the linchpin of the strategy from the 2000 is the one we have been focused on, are we actually focusing on the right threat from these actors? Secondly, and more defiantly to the policy of the US (by the way, the reason I am focusing on the US here is because they have the most actively developed and declared strategies, so in the forum like this we can openly address it, but it doesn't mean that such a capacity doesn't exist within the other states as well), why despite the technological superiority and the ability to impose full spectrum information dominance, how the tactical solutions to deal with unconventional actors seem to be so inadequate and many cases so counter productive? So this presentation covers the following. One, review a little bit of contemporary information operations and new security actors. Basically what I want to focus on are two key assumptions that underpin these strategies. Secondly, I want to talk a little bit on the sort of ground the evidence that is available verify or doesn't verify that these assumptions that we before had are actually correct. Thirdly, I would like to present some conclusions and observations.

To understand how information operations as a doctrine emerged there is actually indebtedness here to this region. Concept of the revolution in military affairs holds at least some of its heritage back to marshal Nikolai Ogarkov, who postulated that at some time information becomes the key determinant of the ability to pursue conflict, that that

force better able to achieve information dominance, while denying it to that of the opponent, will all in time prevail. This concept basically was based upon three factors. One, the ability to achieve "God sight" while denying it to the opponent. This simply means the ability to have full pervasive and accurate picture of the informational space in which battle occurs while denying that ability to others. It also means depriving the enemy the situational awareness by either distorting or deceiving his understanding or her understanding of the battle space itself. Thirdly, the delivery of highly accurate lethal force, precisely when and where it is necessary.

Information operations as a doctrine has emerged within a larger doctrine of what is known as effects based warfare. What that means is that no longer is information simply an input to words "the achieving of military ends" but in effect the information dominance in of itself is used to compel or convince an opponent to comply or act in a desired fashion. Therefore military force is only one part of a much broader spectrum of potential options, the larger part of it being the ability to manipulate and shape the environment through information. An information operation as it is understood in the west, and this is a combination of doctrines not just one, is made up of two components. The first one is what you might call a psycho-social which deals with deception and psychological operations. It includes such things as psychological operations, the ability to instill either fear or other form of actions in the opponent prior to these in military force. I can give two very good examples of this. They were just prior to the initiation of the operation "Iraqi Freedom". US Computer Network Operations Taskforce identified and sent personal e-mail messages to every single member at this regime that had an e-mail account effectively telling them that if you resist, go to work or continue to operate you will be killed. A more recent example which happened just last week in the West Bank, the defense force called 9,000 numbers in the Northern Gaza Strip, basically private numbers, informing every individual that if you are seen to harm the member of Islamic Jihad or Hamas in their pursuing launching rockets against Israel, your house will be destroyed and the panic that it caused was phenomenal. So these are psychological operations.

Secondly, there is a kind of propaganda. Kind of propaganda is usually the public diplomacy, for example the use of public radio stations and others in order to create an alternative ideological model. There is operation security which goes without saying and which is simply securing the way that information is used with your particular whipping. And here is military deception. During the first Gulf War the fake

coastal landing in Kuwait is opposed the White Sweeper Forces in the west is a good example of military deception. There is also civic action programming, which basically means winning hearts and minds.

The second component is called material-technical. Material-technical refers to active technical measures which are used to either destroy commandier or disturb the content of information systems belonging to an opponent. It includes such things as broad category computer network operations which has now become a core competency within the US strategic command as well as diversify throughout branches beyond forces. It includes computer network attack using logic and algorithm based attacks in order to take over or deny the use of information systems to an enemy. Computer network exploitation is basically hacking but the idea in the difference of the computer network attack is that the computer network exploitation is reconnaissance activity as well as an activity that is designed to create an active intelligence gathering capacity on networks that are targeted. Electronic warfare includes the physical destruction of radio and electronic means. And of course the Holy Grail — signals intelligence, which has now evolved, so now it is no longer a question of gathering information in motion, which is traditional for the role of signals intelligence, but now also extends to attacking information at rest. The capacities are being developed in both offensive and defensive and have actually been institutionalized within structures and it also includes overt means. Some of these are technical, some of these are interesting enough and are also based on application of tactical segments but specifically on targeting networks

The dilemma with this rather complex organizational and institutional structure is that both effects-based operations and information operations are underpins of the assumptions which are grounded in, on one hand a conventional threat, i.e., something that can be attacked because it is an institution or an organization, or secondly based upon the imagined threat. In other words giving the same characteristics as previously existed in a state to an actor which is a non-state. So what are these assumptions?

First assumption is that the threat that exists is from asymmetric actors willing to use weapons of mass destruction. This is a very important idea. The idea is that it assumes that these actors are predisposed to create the largest possible effect which in the realms of imagination within the Defense Department is the use of weapons of mass disruption, which is what the information weapons are done for. The way the new security actors have hypothesized are basically any actor that is able to leverage an increasingly globalized technology dependent world in three

particular ways. One is that they have agency that simply needs to have a group which is motivated. Secondly, that that agency's willingness to act is multiplied and amplified through the use of technology. And thirdly, that technology is applied to the largest and most symbolically important mass effect.

Second assumption is that the structure, that I showed you before, Information Operations Doctrine, assumes that this non-state actor, new security actor, can be disrupted, affected or shaped through the application of Information Operations Doctrine as it exists. Problems are: does the evidence actually vary these assumptions out because they are so core to the task itself? Well, let's take a look at it.

First of all let's look at cyber terrorism. The most important thing to know is that there are no recorded instances of an extremists group causing significant damage through the use of cyber attacks. Although there have been many cases of the symbolic uses of cyber intrusions — another forms of defacement. There is no documented case where such an actor as this — a new security actor — has been able either to take over or significantly damage. I think it is an important thing to know because this is not a new question and seven years on it is important to see that it hasn't happened. So how have they used them? A lot of existing theory that has been written in independent studies has been drawn some clear examples. First of which was the Zapatista's use of networks in being able of both bring attention to their struggle and also the use of a very simple denial of service attack as a way to effect raise the specter of a nongovernmental politically motivated group using a network attack as a part of its political tactic. Problem with this Zapatistas, who have been eloquently written about by Arcylla and Rundfeld and whose case study has underpinned much of the thinking about how non-state actors use information systems, is that it kind of doesn't hold truth. I mean Zapatistas themselves were largely ignorant of technology or they actually used the technology in their struggle Chiapas. Rather there were nongovernmental organizations primarily from Europe which networked within themselves and on behalf of Zapatistas in fact successfully were able to raise the level of their cause itself. So it seems if look at the retro spectrum of evidence that it was western nongovernmental organizations that found the Zapatistas and ascribed them networking function rather then the Zapatistas organically themselves were using this technology.

Second example, which is being used as a case study within the US Military, has been the campaign of network attack which was waged between proPalestinian and proIsraeli hackers during 2000. Now for those of you who don't know this event: in 2000 a group of proIsraeli hackers

managed to take down the main website of Hizbollah, the Sheriat based group in South of Lebanon. As a result to that a group of proPalestinian hackers systematically attacked the Daedal domain, they managed to bring down a largest ISP for a day, they managed to bring down the websites of Prime Minister's office, of the Ministry of Foreign Affairs and because of the latency they introduced into the network, said, to have caused a 8% dip within the Israeli stock market on that particular day.

However it is important that first of all the limitations in terms of the attack itself were very small. In all cases the damage that it caused virtually through the denial of service was corrected within 24 hours. More importantly for both Palestinians as actors as well as Israelis leading the state these acts had no symbolic meaning whatsoever. In other words these attacks became the lore of the computer information security community. But their actual effects on the ground, their actual impact both symbolically as well as materially were absolutely negligible.

Most importantly in 2000–2003 the Center for Terrorism in Regular Warfare of the US Navy carried out a very interesting experiment where they brought together a number of former quite senior members of militant groups including ETA, Basque oriented group, the IRA, the PLO as well as the elements from the Chechen opposition and brought them together with a number of computer hackers in order to make a scenario, how militants would conceptualize the possible use of cyber weapons within campaigns that they would structure. What was interesting about this experiment, as limited as it was, that it found out there was very little interface or commonality between those two groups. At most measure members of militant groups were looking for symbolic victories. Symbolic victories, which were directed not just at their opponents but more importantly to reinforce their position within the community that they served. What they needed was acts of symbolic violence which were understandable. The bottom line was that there was no equivalent to suicide bombing in cyber space. Therefore they saw cyber terrorist attacks as being high cost because they required a lot of planning, a lot expertise, where as potential benefit, which was symbolic act of violence which will reinforce their positions within their communities, was very small. As a result the experiment concluded that at least for conceivable generation of militant groups cyber terror attacks were probably not as significant or possible as was hypothesized previously. I think this is a very important symbolic finding, which I can say is also being verified through some of the work that presently our Program is undertaking in a number of key locations of the work.

To the second assumption. Can conventional information operation disrupt terrorist organizations? Well, in order to understand it, we have to understand: what is it that extremist groups do in terms of leveraging information networks if it is not cyber terrorism? First of all it is a fact that extremist groups do use the Internet for command and control, for fund raising, for data mining, for messaging, for networking and for recruitment. These kinds of functions are what Rundfeld and Arcylla have broadly called social network. Secondly, it s also clear that the extremist groups use the Internet as way of amplifying their message, exaggerating their importance and instilling fear. The examples of these uses are wide scale distribution of videos, DVDs, exstreming websites that feature in effect videotaped attacks as well as more recently beheadings and others. It is very interesting that many of the groups we looked at effectively are no larger as a group than between 5 and 20 people. So this ability to amplify their message is in effect the strategic multiplier that they use. That's why this particular use for them is far more important than cyber terror. So this second one involves psychological warfare. So basically if we look at it, the current use of information means by terrorist groups is largely if not most entirely held to psycho-social realm. But the problem facing all of us is that in terms of targeting these groups is that this use almost completely indistinguishable from the normal use of the Internet or the media. In other words, there is nothing specific for it to be effective in terms of targeting it using existing IO means. Moreover in terms of what these groups are after is an amplification of psycho-social phenomenon. There is very little to counter in terms of the application of psychological warfare, deception or any of these other well defined means. Moreover what is also very important if we are talking about effecting those groups is that, for example, based upon groups that we know about right now, 80% of them live in diaspora communities which means there is no geographical center of gravity for many of them.

Secondly the actual cells that they are made up for are extremely difficult to penetrate because they are very closed. 20% of those who are operational and active in Afghanistan were actually made out of people who are related to each other. Further 70% were friends, almost all of whom were related through marriage, which means the ability to penetrate or message effectively within these groups is almost impossible. Moreover, when you look at the difference between cells, there is almost no homogeneity whatsoever, which means effectively that the attempts to profile these particular movements is almost impossible, and at the same time their capacity is to spontaneously self form, self mobilise and remain isolated within other cells is almost infinite.

The paradox that is facing us when we look at this particular power is that cost of the actions that we could think of that would degrade the ability of these actors to act by targeting their psycho-social effects are actually ineffective because, one, they are largely indistinguishable behind the noise of the Internet as it begins with and, secondly, when we target them, when they deliberately say: "You, that group of 20, we are going to target you for your messaging", by virtue of the fact that most of them circumvented, we in effect build their incredibility because we have targeted them. So there is a paradoxical relationship here, unfortunately. It is a very interesting case that demonstrates just how inadequate existing IO doctrine has been facing this threat and which can be found in this case study. For those of you who don't know, the IDF (Israel Defense Forces) enjoys probably the most compliment security environment of any armed forces dealing with terrorist threat. They enjoy full tactical dominance over the Palestinians in all forms of arms; they have a full freedom of movement to operate anywhere within the territories and in a given time. There is a complaint international environment that recognizes Israeli right for self defense and the use of extra legal means such is targeted killings, the building of the security barrier, and the legal basis for arbitrary arrest and detention. Moreover, for those of you who come from the signals intelligence community, Israel also possess the most extensive fixed signals intelligence infrastructure anywhere in the world, which was the part of the US guarantee that they underwrote the Oslo Accord allowing the Palestinians to take some of the responsibility in areas of occupation. They spend more on signals intelligence than any other country in the world and possessed a very pervasive and technologically sophisticated GIS enabled system for surveillance that allows them to keep track of individuals in real time throughout the territories. In addition their physical control of the territory, the ability to limit how people travel through the permit system allows them to use "compromat" as a way of gaining a lot of human intelligence on networks that they deal with and an entire Palestinian telecommunication infrastructure as well as the Internet is rotated through Israeli infrastructure. That means they have a fixed collection point on any external access. And yet despite the existence of both tactical dominance as well as full spectrum information dominance, they have been unable to prevail over determined actors such as Palestinians Islamic Jihad and Hamas who have become an expert in unconventional information warfare, focusing entirely on effects rather than on conventional military units. Without getting into details because I am already running out of time, the fact is this that both of these groups are highly technologically sophisticated in

the way they communicate. For example, Hamas holds mass meetings using ever-changing ARC's streams to which literally Hamas people subscribe. They have completely separated their military structure which is based on atomized cells from political structure which is seen as bringing a great benefit to the community it serves by being uncorrupt and serving social needs. Which means that any time when IDF acts either through a targeting killing or through some other form of collective military reprisal against the military army of Hamas the political strength of Hamas as a resistance actually rises. It acts as a recruiting mechanism to these groups rather than not.

So some conclusions and observations. First of all I would suggest that if we look at the structures, strategies and needs that contemporary information operations information security at the state level targets, it actually mistargets new security actors. It plays to conventional strengths, which is attacks and defense within the state doctrines but it is not really real in terms of these particular actors themselves. The result is that the enormous investment of both capacity, institutions and strategy, I would suggest, resembles very much of Maginot line, meaning that it was a defense strategy built around the threat that no longer exists. Main problem is that, at least for strategic thinkers in the west, they have contemplated two different kind of threats: one is non-state actors, which they don't understand and can't get a purchase from with potential strategic competitors such as China, India and others, who have developed and who are developing actively information operations as a way of asymmetrically gaining an advantage over these state based actors.

So main dangers that we are facing are: one — over investment into IO against the unconventional actors such as Israel, which will untimately not bring any kind of strategic return and, in effect, diminish the degree of security we hold against them. Secondly, a danger is of negative collaterals and these includes things such as trying to create barriers and gateways on the Internet as a way of dealing with these actors, vulcanizing perhaps and splitting them off in international segments, using them or creating them through filtering protected zone, effectively eroding freedoms, and probably most importantly the danger of muling corporate interests which are designed to address security with security itself, which means all of a sudden we have unaccountable practices being put into place to address a threat, which actually doesn't even exist, which all the time erodes freedoms and benefits that we get from these networks. The reality is that the sophistication of conventional and unconventional actors in this particular field will definitely

grow and there is some transparent evidence which I will discuss at the round table, which is interesting to see in this area or in others.

However, the problem is there are no simple solutions. Unconventional actors thrive because addressing them requires a degree of international coordination and police work, information "sharing" and willingness to bring these actors and the communities that they support back into the political mainstream. For both, reason of national sovereignty, which deals with signal intelligence, as well as politics, i.e. saying that we are negotiating with terrorism, both of these things are very difficult. And yet at the same time they really run at the core of being able successfully to address this.

Problems in the Fight Against Computer Crime

B. N. Miroshnikov

Good afternoon, Dear Participants of the conference,

It is my pleasure to welcome you to this, I am not afraid of exaggeration, outstanding educational institution which has recently celebrated the 250th anniversary, and is as always, a flagship of our scientific school. This is an educational institution that has become famous in the whole world for its graduates and their discoveries. Therefore it is a great pleasure to be here in spite of the fact that the conference is being held in an absolutely new building. All the same we feel the weight of two centuries of heroic scientific activity to the glory of science. One of the remarkable traditions of this institution is to respond to needs of today, the country, the people, mankind. And one of the problems that unites us today in this hall, which involves more and more minds all over the planet, is the topic of ensuring information security in our information world, in our information century.

This problem is extremely pressing and it is wonderful, that the Russian scientific school participates most actively in its solution. Today in our country, a school has already been formed that allows us to efficiently resist crimes which we call crimes in the sphere of information technology. All over the world, as we know, the term cyber crime is accepted. The important thing is not the name of these phenomena, but the way we understand them. Also, I can say with pleasure that all of us agree in general definitions of these concepts, though some details still need to be specified.

In this sense we rely on sciences both in the fields of mathematics and high technology, and in the field of law, which perfectly coexist under the roof of this educational institution. We rely on their assistance, because even today, in different countries, different groups of experts in different branches treat the theme of terrorism differently. This problem, which is on everyone's tongues, is extremely pressing and is everybody's concern, but a precise definition to terrorism has still to be found. Therefore

referring to the same concept of terrorism, we, nevertheless, mean different things. There are divisions by typology of conducting a crime; by the objectives and goals of the crime; its consequences; its realization; and, eventually, by public reaction. These divisions need to be judged, analyzed, and exact definitions ought to be finally work out.

Why we all require accurate definitions is absolutely clear. We very much count on the aid of our science, on the aid of lawyers-scholars who would help to put in order this conceptual apparatus. Besides, the existing regulatory legal base in this area certainly requires perfection. I would like to express gratitude on behalf of the law enforcement structures to those who started to work on this issue almost ten years ago. At that time, these crimes had no concrete realization in life. Then many crimes and their consequences were represented speculatively.

Nevertheless, our legislators had enough wisdom and managed to predict the situation, to arm the law enforcement bodies with system of protection by a corresponding legal regulatory base, including it in the Criminal code chapter 28.

In my opinion, that was a very progressive phenomenon for that time, it anticipated the succession of events and armed those who were practically engaged in these matters with the necessary legal tools. But years pass, technologies develop, and unfortunately the underworld develops too.

We can see today how it changes its face, and therefore, the legal base being conservative and inertial in its nature, needs updating and improvement in order to meet modern requirements. We expect science, our advanced science, to assist the lawyers — those who put these ideas and wishes into a concrete legal act, and help us perfect our work. It can be done if we work in close contact, for we deal every day with troubles, problems, crimes. This is what I call "feedback put into practice", which should find place in our research, proceedings, and concrete serious decisions at all levels of the government, the State Duma, and in general, all authorities. It is extremely valuable, and, in this area we particularly appreciate interaction with science.

Cyber crimes' main difference from all other crimes, is that besides all the other properties, they are obliged, simply doomed to rely on scientific research, to be constantly connected with science. The very structure of these crimes demands constant examination. Do we have today a school of experts in the field of cyber crimes? Unfortunately, I do not think so. Though today many groups have already developed techniques, instituted methodology, and prepared corresponding experts who have the qualifications and legitimate right to exhibit their exper-

tise in this area. No investigation occurs without such experts. This is a very important circumstance. And as long as one of the characteristics of computer crime is instant coverage of huge territories, including territories divided by administrative borders, ones with different legal regimes — it requires identical readiness of those who conduct the investigations, their identical armament with both techniques and intellectual power.

Today we need a computerized detective, a detective who masters these technologies, who is armed by legal base and can conduct investigations at a high technological level, similar to the unfortunately high level of computer crimes. This is the dialectic of today, and we have to correspond to it. And again I address science, because to develop techniques, to prepare the base for exhibiting expertise, a base for preparation of the first response structure, capable of meeting modern requirements — is possible only in cooperation with science.

Scientific Problems in Counteracting Cyberterrorism

V. A. Vasenin

Terrorism is a cruel fact of today's life. Citizens and whole countries are facing this threat together today. With the growth of technology, new opportunities for terrorist activity are emerging. Cyberterrorism is one of those new opportunities. This form of terrorism is a match for all other high-tech forms in damage caused. Moreover, the possibility of initiating destructive actions from any point of the globe and avoiding detection, along with other factors, makes cyberterrorism a major threat to humanity. Fortunately, as of today, there are no facts of cyberterrorism. The reason for this is the technological difficulty of terrorism at the current level of network infrastructure development. But the pace of change in worldwide telecommunications and IT suggest that we should start to prepare for the threat today. The first step must be the study of cyberterrorism as a phenomenon, a systematization of subjects, objects and the environment. We need a formalized domain that allows efficient reasoning. We will briefly characterize one of the possible approaches to the study of cyberterrorism as a phenomenon, to the development of means to counter it, as a development of the ideas from [1].

1. The basic premises of cyberterrorism as a phenomenon. National interests in information security

Practice shows that in the study of any phenomenon, natural, technological or sociological, the end result is largely determined by the choice and formalization of the basic concepts. The goals of the research that went into this work are:

- the formulation of basic theses that would allow the systematization of approaches to studying computer terrorism;
- the development of a concept of defense against the cyberterrorist threat;
- the practical implementation of a system of models, mechanisms and tools to counteract cyberterrorism.

The starting point for studying any phenomenon is its definition. It should accumulate the basic behavioral attributes in accordance with the researchers' experience, goals and ideas. Terrorism is a complex, multifaceted phenomenon that does not yet have a unified, rigorous definition. The study of cyberterrorism today is interdisciplinary, spanning mathematics, physics, IT, psychology, political science, law and economics. Each of those directions of study has its own peculiarities. We shall use the following definition that unifies the basic characteristics of terrorism for the purposes of this study:

Terrorism is a manifestation of extremism in actions based on disagreements (national and transnational) between government interests/institutions and certain groups of people (in the political, social, religious and criminal spheres), aimed at creating an atmosphere of fear and tension within the society and at destabilizing national security, in order to put forward demands to the government that cannot at present be legally satisfied.

The underlying political motivation is based on disagreements between certain groups of people and government interests and institutions that embody those interests.

Cyberterrorism is a direction of terrorism that:

- is aimed at information systems, network segments and security-critical national information support systems;
- is using computing machinery and software as the means of destructive influence.

So the primary target of a cyberterrorist attack is a critical object (CO) that is influenced through its computer-based control system.

Due to the terrorists' desire to create fear and tension, COs are the potential objects of destructive influence.

An object is security-critical if its degradation or loss of functionality may shortly and directly affect certain aspects of national security: energy resource management (nuclear and hydro), transport flows (railroads and air), defense, critical manufacturing industries, etc.

It is natural to base the methodology of counteracting cyberterrorism on the methods and approaches of traditional information security (IS), or, more rigorously, information technology security (ITS).

The overall goal of a traditional information security system is the creation of a system of measures:

- preventive — at the legal, administrative and operational levels;
- dynamically monitoring the security of objects belonging to the national information and telecommunication infrastructure, and adequately reacting to the threats.

On the other hand, it should be noted that the security state of each object is determined by its needs to be protected from potential threats. Those needs are different between operating systems, databases and complex distributed structures that manage whole sectors of the national economy. The problems in achieving the global goal of ITS are aggravated by the fact that the legal field is in a formative stage. With the huge diversity of objects, the demand for defense grows much quicker than the technological capabilities. For example, the capabilities of dynamic monitoring depend on mathematical, algorithmic, technological and technical means. It is a hard problem to develop such means for lots of diverse objects, when the domain hasn't been formalized and unified. We can conclude that, as of today, society does not have a system that would fully solve the main problem of information security.

Due to the above, the large field of ITS should be subdivided into smaller, more concrete and feasible parts, whose formalization would allow to solve practical problems efficiently. One of those problems is ITS of security-critical objects. In this context, it makes sense to single out a set of national interests that can be briefly formulated as follows:

- Protecting the basic elements of the National Information and Telecommunication Infrastructure(NITI) that directly affect national security.
- Ensuring the stable functioning of the national backbone network.
- Constant development of an integrated security system for nationally important computer and network structures.
- Creation and support of a national system for training and retraining IT security staff.

The two latter positions are less time-critical, but also influence national security.

As the above shows, there is a conflict between national IT security interests and the goals of cyberterrorism.

From here on, we will talk about security-critical objects (COs) and information systems that control them (COIs).

Cyberterrorism, and approaches to counteracting it, can be studied. We need to create:

- a well-formalized theoretical base of the subject area: objects, subjects, the environment, means of counteraction etc.;
- recommendations for creating national and international structures to counter the threat adequately and on time;
- adequate tools of defense for COIs.

The Russian government and scientific community are aware of this fact, as evidenced by the introduction of a topic, "Methods and means of

counteracting computer terrorism", under a federal program "Research and development in priority areas of science and technology" for years 2002 through 2006.

Summing up, the directions for the basic problems in studying cyberterrorism are as follows:

- The basic concepts that identify cyberterrorism as a social phenomenon (objects, subjects, the environment).
- A comprehensive set of threats, models and scenarios of computer-aided attacks at critical objects.
- A system of measures at the legal, administrative and operational levels of implementing information security.
- A hardware and software system that would support a representative set of mechanisms, models and scenarios to counteract the threat of cyberterrorism.

2. Objects, subjects, the environment. Systematization, categorization, requirements

A formal diagram of subject-object interaction in a cyberterrorist attack can be proposed (see Fig. 1):

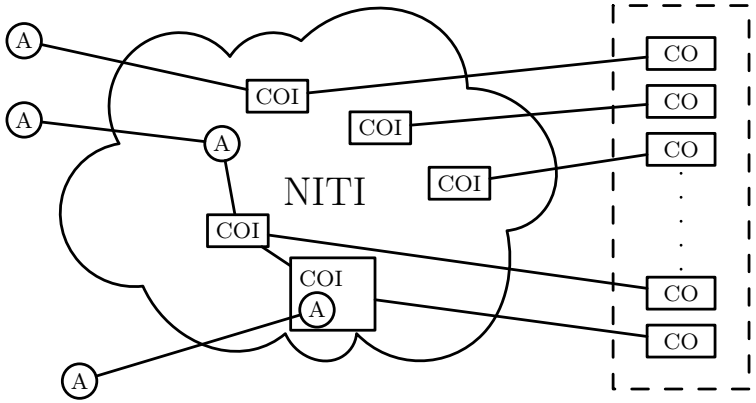

Figure 1.

As noted above, the end goal of a cyberterrorist attack is COs, whose functioning directly infuences various aspects of national security. COIs that control those objects, as a rule, use the national information and telecommunications infrastructure. Agents with varying levels of intelligence, separately or together, using open or protected channes, try to

influence COIs to degrade or stop the functioning of COs. The key directions of research for developing efficient measures and instruments to counteract the threat of cyberterrorism are as follows:

- identifying and grouping critical segments and NITI objects;
- systematizing CT threats, a taxonomy of attacks and their implementation methods;
- the development of scenarios and models that would allow dynamic definition of interactions between COI, their separate elements, state analysis processes and active reaction to anomalous situations;
- the categorization of critical segments and COs based on the estimated chances of successful attack and the projected effect on national security.

Today, the basic element of active research on this topic worldwide is the critical segment. A classification and grouping of critical segments with respect to cyberterrorist threats has made it possible to build a defense control structure for them in the higher levels of government. However, this level of detail does not allow us to work out requirements and means of defense for specific objects. This fact greatly diminishes the importance of the described research.

The problem formulation chosen in this work emphasizes atomic COs and COIs. At the first stage of research, we choose two groups of identifiers. Macroidentifiers reflect general properties, and microidentifiers specify local properties.

Macroidentifiers represent the social relations sphere, economic segments, significance in the territorial manufacturing hierarchy, architectural and topological properties.

Microidentifiers, on the other hand, characterize flow types, protection of communication channels and software features.

An important direction here is the classification of threats and attacks. The priority tasks in this direction are:

- The development of a COI threat model: a set of threats to its critical elements (operational environment, communication facilities, information and other application-level resources).
- A taxonomy of cyberterrorist attacks and their implementations.
- An assessment of risks of (successful) attacks on COIs and COs.

A key problem on this direction is the development of a formal model of dependencies between threats, attacks and their implementations with risk assessment for individual classes of COI + CO.

Another direction is gathering requirements for methods and tools that would ensure the security of different classes of CT-vulnerable objects (COI + CO). Those include requirements:

- to estimated trust levels for the technical means used in CIOI, including the information security subsystem;
- to architectural and technical solutions, security mechanisms and services;
- to methods and mechanisms for defining security policies and continuously enforcing them;
- other requirements.

3. General principles of building a system of defense against cyberterrorism. Models, criteria, tests

Traditional approaches to implementing IT security are based on the following methods of assessing protection levels of objects:

- mathematical methods that adequately (within accepted restrictions) describe the objects under consideration;
- architectural and technological security solutions, mechanisms and services;
- criteria-based approaches (expert estimates) at all stages of the object's lifetime;
- physical and imitational testing of objects.

The COI is an architecturally and technologically complex entity, which raises the question: can a COI be assessed? The answer is positive, but the assessment will require a decomposition into separate elements and the use of models that account for subtle aspects of their interactions within the larger system.

The mathematical models of objects that are assessed from the IT security point of view, as a rule, belong to one of the two categories:

- models that describe (in the form of security policies) the vulnerable properties of the assessed objects;
- models that verify the correspondence between mathematical models and actual objects.

The first category includes non-interference methods and their development. This direction of research was founded by Goguen and Meseguer [2, 3]. Automata models [4] gave a major impulse to its development. Works by Russian researchers [5, 6] are also worthy of mention.

From the mathematical modeling point of view, there is promising research that concerns security predicates based on messages [7, 8] and process algebras [9]. If we continue to apply this work to COIs and their elements, we'll be able to more accurately estimate their vulnerabilities and attack counteraction capabilities, and, most importantly, develop more efficient mechanisms, models and tools of defense.

An important niche in this direction is occupied by analytical models, logico-lingual means of describing security policies based on discretionary, multi-tiered (mandatory), role-based and mixed models of logical access control. Such models are currently being developed in a number of countries — notably, in conjunction with the development of new security mechanisms in OS kernels (SELinux, RSBAC, grsecurity and several others). The results of research in this area will be presented at the conference by K. A. Shapchenko and O. O. Andreev (IISI MSU) from Moscow and I. V. Kotenko and A. V. Tishkov (IIAS RAS) from St. Petersburg.

A traditional COI is a set of interacting and changing AOs, each with a separate security policy. Therefore, the approach of multi-level decomposition into objects and their trust relationships is interesting and promising. This approach will be presented by V. B. Savkin and A. A. Itkes (IISI MSU).

Of course, not all mathematical models from the first group have been mentioned. However, their efficient implementation requires lots of computational resources. Methods of efficient usage of large computational resources pose another interesting problem.

Mathematical models of the second category verify the correspondence between actual objects and their theoretical models. For example, when creating defense software with high trust levels, Open Source software is generally used. We need to analyze the source code of this software to detect vulnerabilities (buffer overruns, memory leaks etc.) that threaten security. Such models are researched at the RAS Institute for Systems Programming and the MSU Institute for Information Security Issues. Some of them are featured in the proceedings of last year's MaBIT'04.

A separate and important type of mathematical models is aimed at verifying program models. This verification is needed not only for whole software systems, but also its constitutive parts, up to the implementation of security mechanisms in the OS kernel. This is an important direction of research. It will be partially demonstrated at this conference in the section reports of K. A. Shapchenko, I. V. Kotenko and A. V. Tishkov.

A novel and promising approach based on mathematical modeling is proof-carrying code.

Solutions to problems enumerated above can significantly raise the estimated trust level of many components within the complex object which is COI.

Another direction is estimating the security of COIs based on testing. The tests can be physical (real-world) or imitational. The first approach is more promising, but difficult and very expensive from the resource point of view.

The second direction is less accurate, less resource-intensive, but requires intellectual work to build adequate models. This direction is being actively pursued today. Packages for imitational modeling of network segments (NS2, OMNET++, INET and others) are in active use and constantly being updated for new IT security problems. This line of research is bringing its first fruits in Russia as well [14]. Those results will be presented at the plenary session tomorrow by I. V. Kotenko and A. V. Ulanov (IIAS RAS), and at the section reports of I. V. Batov and M. V. Bolshakov (IISI MSU).

As for criteria-based approaches to evaluating the security of a COI at all stages of its lifecycle, this approach also has problems. The first problem is that these objects are unique: assessed objects (AOs) aren't manufactured goods. Another difficulty is connected with the geographical distribution of AOs and separate evolution of their different parts during the AO's lifecycle. The third set of questions is due to the broadness of tasks and the diversity of software and hardware used in different components of the large system. Therefore, we need requirements that would guarantee correct integration of the security mechanisms of different COI components into a unified system. Some mechanisms for this unification will be discusses in the section reports of O. O. Andreev, V. B. Savkin and A. A. Itkes.

We need to develop functional and trust requirements and defense profiles for COI components at the following levels:

- the operations environment;
- the communications environment;
- the application environment (DBMS, other application services).

This is a very important line of work. No noteworthy practical results have been achieved to date, but we have reason to believe they will appear shortly, due to theoretical developments.

The analysis of existing COs shows that they have three typical topologies:

- "star" topology;
- complexes with a strong subobject hierarchy based on the territorial principle (corporate or department "trees");
- complexes with lots of horizontal relationships (systems of inter-department, inter-corporate interaction).

The most complex and interesting COIs have lots of horizontal relationships. This is due to the fact that unified admistrative measures and operational procedures are hard to introduce, as well as standardization of technonogical solutions and technical means. Isolating and classifying threats and attacks on segments of COIs and communication channels poses some separate questions. A potential vulnerability is inconsistency of security policies in segment-to-segment interaction. For example, the "Electronic Russia" system of interaction between government institutions that is currently under development is an example of such system. The difficulties of building such systems are well-known. But some approaches can and should be employed even today. Some guiding principles are:

- a schema for internal subject/object interaction in each segment is defined and strictly regulated;
- external interactions are based on trust relations.

Based on these principles, we can create "islands" of mutual trust, as preconditions for exchanging data about security policies, with an eye towards aggregating them into a united "big system" policy.

4. Approaches to implementing a system of models, mechanisms and tools to counteract cyberterrorism

The conceptual framework of CO security should be used as a base to build a system of interconnected models, mechanisms and tools to counteract cyberterrorism. In this regard, the main directions of introduction of those principles into systems of counteracting cyberterrorism at all levels of implementation of the integrated approach to IT security are of major interest. Let us briefly characterize each of those directions, keeping in mind that their role has already been justified.

At the administrative level, the main directions of interest in developing security systems for COIs within an integrated approach to IT security are as follows:

- Development and implementation of efficient COI models and security policies, tools for their description (specification) and permanent control of their execution.
- Analysis of mathematical, algorithmic, and technical aspects of promising operating systems, to find efficient techniques for resisting cyberterrorist attacks.

- Development of new access control models and their implementation in present and new operating systems.
- Development and deployment of new logico-linguistic means of formal and effective description (specification) of security policies for distinct parts of a COI, which take in consideration interactions between parts with respect to global policy.

The operational level presumes the execution of security measures under the control of staff. The first practical steps here are as follows:

- Development and implementation of staff-related COI security measures, including application of tools for typical business-process automation.
- Statement of problems for the safe maintenance of complex objects with a high confidence level, profiles of defense against cyberterrorist attacks.
- Development of processes for control of security policy compiance at distinct COI objects, their interaction with respect to the global model, prompt response to exceptional situations.
- Creation of (heterogeneous, distributed) systems that would ensure continuous functional monitoring of distinct COI elements and the system as a whole. Analysis and prompt response to exceptional situations.
- Source and executable code analysis to identify and remove vulnerabilities.

At the software and technical level, we need to implement mechanisms and services that ensure security policies without staff interference. Keeping that in mind, the main measures to efficiently implement systems of COI defense from cyberterrorist attack today are as follows:

1. Development and implementation of COI specific services which support a multi-stage security architecture including the following:

- Intrusion detection systems at the first stage (system call trace analysis, traffic analysis, firewalls and filters, etc).
- Efficient means of identification and authentication, authorization and access control at the second stage.
- Integrity control, active security monitoring, analysis and prompt reaction to attacks at the third stage.

2. The development and implementation of traditional COI specific software and technical services as high estimated confidence systems, supporting a multi-tiered security architecture, should be carried out on the following two directions:

- Building services based on existing tools.
- Building services based on models, mechanisms and tools that support new functionality and new confidence levels.

Conclusion

If we solve those basic problems and a set of related problems, it will lead to:

- unification of concepts and creation of legislation that would regulate activity in the field of cyberterrorism counteraction;
- development of formal models, including mathematical models, of cyberterrorist activity, security systems and systems to defend;
- creation of a permanent monitoring system that would reveal the state of cyberterrorist activity and analyze potential vulnerabilities and means of defense.

References

[1] Vasenin V. A, Galatenko A. V. Computer Terrorism and Information Security in the Internet. Proceedings of Russian-American Workshop on Computer Terrorism, June 4–6, 2001, p. 211–224.

[2] Goguen J. A., Meseguer J. Security Policies and Security Models. Proceeding of the IEEE Symposium on Security and Privacy, Oakland, CA, 1982.

[3] Goguen J. A., Meseguer J. Inference Control and Unwinding. Proceeding of the IEEE Symposium on Security and Privacy, Oakland, CA, 1984.

[4] Moskowitz I. S., Costich O. L. A Classical Automata Approach to Noninterference Type Problems. Proceedings of the Computer Security Foundations Workshop 5, Franconi, NH: IEEE Press., 1992.

[5] Grusho A. A., Timonina E. E. Non-interference Model of a Network. Review of Applied and Industrial Mathematics, vol. 7, 2000 (in Russian).

[6] Galatenko A. V. Probabilistic Models of Systems with Assured Security. Proceedings of the Mathematical and Program Computer Security Ensuring conference. MSU, October 23–24, 2003, p. 234–237 (in Russian).

[7] Mantel H. Possibilistic Definitions of Security — An Assembly Kit. Proceedings of the 13th IEEE Computer Security Foundations Workshop, Cambridge, United Kingdom, July 3–5, 2000, p. 185–199.

[8] Mantel H., David S. Controlled Declassifications based on Intransitive Noninterference. Proceedings of the 2th ASIAN Symposium on Programming Languages and Systems, APLAS 2004, Taipei, Taiwan, LNCS 3302, November 4–6, 2004, p. 129–145.

[9] Ryan P., Sneider S. Process Algebra and Non-interference. In IEEE Security Foundation Workshop, 1999, p. 214–227.

[10] Gaisaryan S. S., Chernov A. V., Beleventsev A. A., Malikov O. R., Melnik D. M., Menshikov A. V. On Problems of Program Analysis and Transformation. Proceedings of Institute of System Programming, vol. 5, 2004, pp. 7–41.

[11] Puchkov F. M., Shapchenko K. A. Static Analysis Method for Detecting Buffer Overflow Vulnerabilities. Programming and Computer Software, vol. 31, no. 4, 2005, p. 179–189.

[12] Puchkov F. M., Shapchenko K. A. On Buffer Overflow Detection via Static Analysis of Source Code. Proceedings of the Mathematical and Program Computer Security Ensuring conference, MSU, October 28–29, 2004, p. 347–360 (in Russian).

[13] Appel A. Foundational Proof-Carring Code. In 16th Annual IEEE Symposium on Logic in Computer Science (LICS 01), June, 2001.

[14] Kotenko I. V. Multiagent Models of Defense and Offense Agents Opposition in the Internet. Proceedings of the Mathematical and Program Computer Security Ensuring conference, MSU, October 28–29, 2004, p. 257–266 (in Russian).

Important Branches of Discrete Mathematics Connected with Applications in Cryptography

M. M. Glukhov, A. M. Zubkov

Information technology security problems are practically inexhaustible. Methods for the information protection during its storage and transmission constitute an essential part of their solution. A considerable part of such methods are developed, investigated and realized by the cryptography which is a science on the processes of information transformations aiming to exclude the possibility of uncontrolled access to this information.

The cryptography, in its turn, has to use the results of different branches of mathematics, in particular, results of discrete mathematics (which should be understood in a wide sense). Good examples of practical realization of deep mathematical ideas are public key cryptography (based on the hardness of some number-theoretic problems) and modern block cipher construction (using transforms over complex algebraic structures as in AES).

There are several reasons for intensive and continuous investigations of the security of new and existing ciphering methods.

Really, because of the value of secret information, the reliability of cryptographical information protection should be out of doubt for a long time from the moment of ciphering. But the security of almost all cryptographical methods is based in essence on the assurance that eavesdroppers could not find the secret key or protected information. Such assurance should stem from the knowledge of possibilities of contemporary mathematical methods, computational algorithms and technical devices. These possibilities are constantly growing in time. So, the cryptographer should be able to improve the cypher many years before the growth of these possibilities will become dangerous really. Let's mention two examples:

- DES (it has became unreliable due to the development of computer power),

- security of public-key algorithms of Diffie–Hellman and ElGamal is based on the hardness of number-theoretic problems, but the set of solvable cases of such problems is growing slowly (e.g., in [3], [4] some cases of low-exponent RSA are considered).

Secondly, cryptographical methods of information security are elaborated and investigated (as a rule) as formal mathematical constructions, but its applications take place in a real world. Practical realization of ciphers reflects the contemporary level of technique and possesses properties which were not present in theoretical construction. These unprovided properties create additional possibilities for eavesdroppers to obtain information. Two examples of such type are:

- the differential power analysis ([5]) based on the data of energy consumption by the smart-card processor during the public key protocol,
- differential fault analysis (see e.g. [2],[6]), based on some specific faults appearing in the process of computation.

So (in view of possible detecting and using the deficiencies of information protection systems by eavesdroppers) it is necessary to support intensive and wide-spread investigations in different branches of "pure" science (not only those which are closely cryptography-related).

In the process of design and investigation of cryptographic transformations many methods of number theory, algebra, complexity theory, probability theory are applied. From the other hand, needs of cryptography (in fact — needs of secret holders) are valuable sources of new mathematical problems and theories. For example, computational number theory, computational algebra, finite algebraic structures (in particular, elliptic curves over finite fields), algorithmic complexity, pseudorandomness, zero-knowledge proofs, protocol testing etc. are developing largely under the influence of concrete cryptographical problems.

A major part of mathematical background of cryptography may be divided into three classes: Discrete algebra problems, Number-theoretical problems and Probabilistic problems. Each class contains several branches. Some of these branches are listed below, and for each branch we review results of different authors published in the annual volumes "Trudy po discretnoi matematike" ("Proceedings in Discrete Mathematics") [7, 8, 9, 10, 11, 12, 13, 14] published by the Russian Academy of Sciences in association with Academy of Cryptography of Russian Federation.

1. Discrete algebra problems

Permutation groups. A review of recent results was made by B. A. Pogorelov ("Permutation groups. Part I (Review for 1981–95 years)", [8], pp. 237–281). Main themes of this review are: O'Nan — Scott theorem, Maximal subgroups, Primitive permutation subgroups, Uniprimitive subgroups, Multiple-transitive groups, Actions of groups on k-orbits, Solvable and nilpotent groups, Operations over permutation groups.

Some papers were connected with the generation of given subgroups of the permutation group $S(\Omega)$ over the set Ω by different sets of permutations.

For example, let $V_n = \mathrm{GF}(2)^n$ is identified with Z_{2^n}, and S_{2^n} is the permutation group over V_n. Let $g = (0, 1, \ldots, 2^n - 1) \in S_{2^n}$ be a cyclic permutation, and the permutation $D \in S_{2^n}$ be defined by elements $\alpha_0, \alpha_1 \in V_n$ and by a function $f\colon V_n \to \{0, 1\}$:

$$Dx = x \oplus \alpha_{f(x)},$$

where $\oplus$ denotes addition in $\mathrm{GF}(2)^n$. Let $G = \{g^k D, k = 0, 1, \ldots, 2^n - 1\} \subset S_{2^n}$. It was proved by M. M. Gluhov ("On numerical parameters connected with the generation of finite groups by systems of generating elements", [7], pp. 43–66) that if

$$f(0, x_{n-2}, \ldots, x_0) + f(1, x_{n-2}, \ldots, x_0) = 1 \quad \text{for all} x_{n-2}, \ldots, x_0 \in \mathrm{GF}(2)$$

then there exists $k \geq 5$ such that the set G^k is 2-transitive.

Further, let S_N be a permutation group on $\{0, 1, \ldots, N - 1\}$, $g = (0, 1, \ldots, N - 1)$ be a cyclic permutation, and $h = (0, 1) \in S_N$ is a transposition of elements 0 and 1. Denote by D the diameter of S_N with respect to the generator system $\{g, h\}$, i.e. the minimal number d such that every permutation $c \in S_N$ may be represented as a product of no more than d multipliers from $\{g, h\}$. A. Ju. Zubov ("On a diameter of group S_N with respect to the generating system consisting of a one-cycle permutation and a transposition", [8], pp. 112–150) proves asymptotically equivalent upper and lower bounds for the diameter:

$$D \leq \left[\frac{N-1}{2}\right]\left(\left[\frac{N}{2}\right] + N - 1\right) + 2N - 1,$$

$$D \geq \frac{3N^2}{4} - 2N, \text{ if } N \text{ is even},$$

$$D \geq 3\left[\frac{N}{2}\right]^2 - N + 3, \text{ if } N \text{ is odd}.$$

F. M. Malyshev ("Inheritance of some properties of generating families by a substitution group", [14], pp. 155–175) considers permutation group G on the finite set Z possessing a family of representations in the form of direct product of two subsets; moreover, there exists generating system of group G such that each element of this system doesn't change one of two coordinates in some of representations. There are found conditions on the families of representations and on systems of generating elements sufficient for (respectively) transitivity, primitivity and 2-transitivity of group G and the inclusion $A \subset G$ where A is the alternating group on Z.

Group-theoretic classification of functions and automata. This branch of research includes classification of functions (Boolean functions, in particular) with respect to different groups of transformations. For example, A. V. Cheremuskin ("Methods of affine and linear classification of binary functions", [10], pp. 273–314) describes methods of construction the tables of representatives of equivalence classes of Boolean functions of n variables with respect to generalized linear and affine groups. He finds some new classifications for the cases $6 \le n \le 8$.

Now let us consider non-autonomous binary shift register (NABSR) of length n, i.e. an automaton with the input alphabet GF(2), state space $\mathrm{GF}(2)^n$ and the transition function

$$\delta(x, (a_1, \ldots, a_n)) = (a_2, \ldots, a_n, x + f(a_1, \ldots, a_n)),$$

where Boolean function $f(x_1, \ldots, x_n)$ is linearly dependent on x_1. NABSR is called linear if $f(x_1, \ldots, x_n) = c_0 x_1 + \ldots + c_{n-1} x_n$, and in this case the polynomial $\chi(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1} + x^n$ over GF(2) is called a characteristic polynomial of NABSR. V. A. Bashev ("Group-theoretic characterization of nonautonomous linear shift registers", [14], pp. 52–68) proves that NABSR is linear iff its group is an extension of an elementary Abelian group by means of a cyclic group. Classes of linear NABSR having irreducible (over GF(2)) primitive characteristic polynomial without multiple roots are characterized too.

Investigation and construction of mappings with given properties. Functions and mappings to be used in cryptography should satisfy many different conditions. So many papers are devoted to the investigations of properties of mappings and to the construction of mappings with given properties. One of the main cryptographic conditions for discrete functions over the field or the ring is the absence of the typical properties of linear functions.

There are different approaches to compare an arbitrary function with a linear one (see, for example, [1]). Consider some related papers from [7, 8, 9, 10, 11, 12, 13, 14].

The deficit $d(s)$ of a substitution s on a finite group G of order n is defined as the difference $n - r(s)$, where $r(s)$ is a minimal number of group G translations sufficient to realize all transitions of the substitution s. In other words, $d(s)$ equals to the number of translations of G having no common transitions with s. V. N. Sachkov ("Deficits of finite group permutations", [13], pp. 156–175) investigates different properties of deficit of uniform random substitution ς; in particular, he shows that mean random variable $d(\varsigma)$ depend on the order n of group G only and obtains the following formulas for the mean and variance:

$$\mathbf{E}d(\varsigma) = n \sum_{k=0}^{n} \frac{(-1)^k}{k!},$$

$$\mathbf{D}d(\varsigma) = \frac{n}{e}\left(1 - \frac{2}{e}\right) + \frac{n}{n-2}\left(\frac{1}{2e^2} + \theta\frac{2}{n-3}\right), \quad 0 < \theta \le 1.$$

Let V be a n-dimensional space over $\mathrm{GF}(q)$. A mapping $f\colon V \to V$ is called k-piecewise-linear if k is the minimal number for which there exist linear maps $L_1, \ldots, L_k\colon V \to V$ such that for every $x \in V$ the set of values of these maps contains the value $f(x)$:

$$f(x) \in \{L_1(x), \ldots, L_k(x)\} \quad \text{for all} \quad x \in V.$$

Quasiderivative of a bijective mapping $f\colon V \to V$ along $a \in V$ is defined as $f_a(x) = f^{-1}(f(x + a) - f(x))$. N. D. Podufalov ("On some characterizations of exponential functions on linear spaces", [14], pp. 216–239) proved that the set of bijections $f\colon Z_p \to Z_p$ such that each its quasiderivative is k-piecewise-linear for some $k \le 3$ coincides with the set of exponential functions $g(x) = \theta^x, x \in Z_p \backslash \{0\}, g(0) = 0$, where θ is a primitive element of Z_p.

The minimal value k for k-piecewise-linear function s is analogous to the characteristic $r(s)$ from the paper by V. N. Sachkov because bijective linear transforms of a space $\mathbf{Z}_p$ are translations of a multiplicative group $(\mathbf{Z}_p)^*$.

Let A be a finite alphabet, A^n is a set of all words of length n over alphabet A. In 1956 A. A. Markov have proved that every bijective mapping $A^n \to A^n$ free of reproduction of the substitution errors is a superposition of substitution and permutation cyphers. M. M. Gluhov

("Injective mappings free of error reproduction", [10], pp. 17–32) generalizes this theorem to injective mappings free of reproduction of more types of edit errors: substitution, deletion, insertion of letters.

Let $(G, *)$ be a quasigroup. The mapping $f\colon G \to G$ is strongly bijective if mappings f and h are bijective, where $h(g) = g * f(g), g \in G$. If $(G, +)$ is an Abelian group then the mapping $f\colon G \to G$ is fully strongly bijective if all mappings f and h_k are bijective, where $h_k(g) = kg + f(g)$, $k = 0, 1, \ldots, g \in G$. These notions are connected with the construction of transversales in quasigroups. M. B. Fedyukin ("On some classes of strongly bijective and completely strongly bijective transforms", [12], pp. 226–238) describes a class of strongly bijective mappings and finds a criterion of fully strongly bijectivity for the elementary Abelian p-group.

Linear recurrent sequences (LRS). The study of recurrent sequences has a long history. Large number of papers were published in the second half of XX century due to the applications of LRS over the finite fields and rings in cryptography. In the USSR and Russia many interesting results in the investigations of LRS were obtained by A. A. Nechaev and his school: A. S. Kuz'min, V. L. Kurakin and others. Along with LRS properties of linear and polylinear recurrent sequences over quasi-Frobenius modules and Galois rings were studied. In particular, they consider:

- conditions ensuring the maximality of period,
- ranks of coordinate sequences,
- distributions of elements on cycles,
- representations of sequences.

A number of their results may be found in "Trudy po discretnoi matematike": *Kuzmin A. S., Kurakin V. L., Nechaev A. A.* "Pseudo-random and polylinear sequences", [7], pp. 139–202; "Properties of linear and polylinear recurrent sequences over the Galois rings (I)", [8], pp. 191–222; "Structural, analytical and statistical properties of linear and polylinear recurrent sequences", [9], pp. 155–194; "Structural properties of linear recurrent sequences over Galois rings and quasi-Frobenius modules of charasteristic 4", [10], pp. 91–128; "Almost uniform linear recurrent sequences over Galois rings and QF-modules of characteristic 4", [11], pp. 103–158; *Nechaev A. A.* "Multidimensional shift registers and multisequence complexity", [12], pp. 150–164; "Finite Frobenius bimodules in a linear codes theory", [14], pp. 187–215; *Kurakin V. L.* "Binomial linear complexity of polylinear sequences", [12], pp. 82–138; "Polylinear transforms of linear recurrent sequences over modules", [13], pp. 89–113.

2. Number-theoretic problems

Number-theoretic papers are connected mainly with the analysis and synthesis of public-key or key distribution systems. In particular, problems of factorization of numbers and polynomials, discrete logarithms and algebraic analogues of RSA or ElGamal systems are discussed.

If a, m are coprime integers then their Fermat quotient is defined as $Q(a, m) = (a^{\lambda(m)} - 1)m^{-1} \pmod m$, where $\lambda(m)$ is the exponent of group $(\mathbf{Z}/m\mathbf{Z})^*$. Ju. V. Nesterenko ("Fermat quotients and p-adic logarithms", [11], pp. 173–188) had apply this notion to the discrete logarithm problem. He constructs a set of triples (g, m, r) such that m is a period of the function $Q(x, r)$ and $x \equiv Q(a, r)/Q(g, r) \pmod r$ is a solution of the congruence $g^x \equiv a \pmod m$. For such triples this congruence is not hard to solve.

M. I. Anohin ("On the reducibility of the integer factorization problem to the Diffie–Hellman problem", [9], pp. 7–20) shows that if probabilistic algorithm A does solve the Diffie–Hellman problem for a set N of modules with probability $p \geq \varepsilon(N)$ then there exists a probabilistic algorithm B which finds some divisors of numbers from N with probability $k(N)\varepsilon(N)$, $0 < k(N) < 1$. Moreover, if A is a polynomial algorithm, then B is a polynomial algorithm too.

O. N. Vasilenko ("Some identities for trigonometrical Gauss sums and their applications", [14], pp. 69–78) considers the ring $Z_K[1/q]$, where Z_K is the ring of algebraic integers in the circular field K which is the extension of the field $\mathbf{Q}$ by means of a primitive $p^k q$-degree root of 1. He proves the identity for the Gauss sum and suggests to use this identity in the RSA scheme instead of the usual identity $(a^\alpha)^\beta \equiv a \pmod n$.

M. M. Gluhov ("Investigation of residue rings of biquadratic extensions of an integer number ring and public key schemes", [13], pp. 31–55) suggests to use in the RSA scheme the residue ring mod $n = pq$ (p, q — prime numbers) of the biquadratic extension of the field $\mathbf{Q}$) and describes the structure of this ring.

V. E. Tarakanov ("On the set of values of a cubic polynomial over a simple finite field", [9], pp. 283–294; "Divisibility properties of points on elliptic curves over finite field", [10], pp. 243–258) investigates elliptic curves $y^2 = x^3 + Ax + B$ over the field $\mathbf{Z}_p$, where $p \neq 2, 3$ and $4A^3 + 27B^2 \neq 0$. For the mapping $\psi(x) = x^3 + Ax + B$ the numbers of elements with $k \in \{0, 1, 2, 3\}$ preimages were found and elements of the elliptic curve group having orders 3 and 4 are described. A criterion for a point of the elliptic curve to be of the order 2 is constructed.

3. Probabilistic and statistical problems

Probabilistic models and methods are widely used in cryptography. The diversity of cryptographic approaches to the information security generates a wide spectrum of probabilistic and statistical problems.

Systems of random equations. Secret key finding problems may be reduced to the solution of some systems of equations over finite algebraical structures. Due to the stochastic nature of the data it is natural to consider these equations as random ones.

G. V. Balakin investigates different methods of solving some classes of systems of equations having the type

$$\phi_i(x_{i,1}, x_{i,2}, \ldots, x_{i,k}) = b_i, \quad i = 1, \ldots, T,$$

with unknowns $x_1, \ldots, x_n$ from a finite field, where ϕ_i are known functions,

$$b_i = \phi(x^*_{i,1}, \ldots, x^*_{i,k}) + \varepsilon_i, \quad i = 1, \ldots, T,$$

and $\varepsilon_1, \ldots, \varepsilon_T$ are independent unknown errors ("Introduction to the theory of system of random equations", [7], pp. 1–18; "Systems of random equations over finite field", [8], pp. 21–37; "Systems of random Boolean equations with a random choice of unknowns in each equation", [9], pp. 21–28; "Criteria for a selection of satisfiable system of equations with corruptions in a right hand side", [10], pp. 7–16; "Sequential criterion of extracting a system of linear equations with corruptions in a right hand side", [11], pp. 21–28; "An estimate for parameters of sequential selection of unknowns", [12], pp. 7–13; "Algorithm of searching a minimal set which contains a true solution with a given probability", [13], pp. 7–21; "On some criterion of extraction a system of linear equations with corruptions in a right side", [14], pp. 25–33).

V. F. Kolchin describes the threshold effect for systems of random linear equations and investigates methods of solution systems of equation arising from pair-comparison-based classification problems ("On a threshold effect for systems of random equations", [8], pp. 183–190; "Satisfiability probability for systems of random equations", [9], pp. 130–146; "A problem of classification by means of pair comparisons", [10], pp. 83–90).

Probabilistic models of finite automata. As a rule finite automata modelling the cryptographic devices are very complex. To investigate typical properties of finite automata from concrete classes different probabilistic models of finite automata are considered (usually such models have the form of Markov chains).

Ju. I. Maksimov ("On Markov chains connected with binary shift registers with random elements", [7], pp. 203–220) investigates some analytical properties of Markov chains corresponding to the binary shift registers with noise: spectra of transition matrices, rates of convergence to a uniform distribution. For example, he shows that if

$$y_{t+n} = a_{n-1}y_{t+n-1} + \ldots + a_0 y_t + z_t, \quad z_t = z_{t-1} + \xi_t,$$
$$\mathbf{P}\xi_t = 1 = (1+\Delta)/2,\ P\xi_t = 0 = (1-\Delta)/2,\ \Delta > 0, \quad t = 0, 1, \ldots,$$

and p_t is a distribution of $(y_{t+n-1}, \ldots, y_t, z_t)$, ω is a uniform distribution on $\mathrm{GF}(2)^{n+1}$ then

$$\|p_t - \omega\|^2 \le \Delta^{[(t-1)/(n+1)]}.$$

V. N. Sachkov ("Probabilistic transformers and regular multigraphs. I", [7], pp. 227–250; "Markov chains of iterative transformations systems", [12], pp. 165–183; "Probabilistic transformers and sums of elementary matrices. II", [14], pp. 240–252) considers Markov chains with finite state space S defined by recurrent equation

$$y_{t+1} = f(y_t, x_t), \quad t \ge 0,$$

where x_t is iid sequence with values in $\{1, \ldots, k\}$ and for each x the function $f(\cdot, x)$ is a bijection of S. Ergodicity conditions for the chain y_t are formulated in combinatorial terms.

Ju. N. Gorchinskii derives estimates of mean-square convergence rate for random walks on the set of permutations ("On the improved estimates of mean-square deviations of transition matrices for products of independent random variables on finite permutation groups", [9], pp. 53–72; "On the mean-square deviations of transition matrices on finite permutation groups of even order", [9], pp. 73–94).

V. G. Mihailov ("Investigation of the number of cyclic points in a controlled shift registers automaton", [11], pp. 167–172; "Investigation of combinatorial-probabilistic model of a controlled shift registers automata", [12], pp. 139–149) compares transition graphs generated by a system of irregularly clocked shift registers with a graph generated by a random mapping of finite vector space of their states Q (multidimensional discrete torus). In the first case graph is defined by a system of random transitions from points $x \in Q$ to neighbor points. It is proved that the mean number of cyclic points for a graph corresponding to a system of registers is greater than that for a grapf of a uniform random map $Q \to Q$.

V. A. Ivanov ("Automata transforms of random sequences", [8], pp. 151–168; "On the influence of outer noises on the finite automata performance", [9], pp. 95–110) consider the influence of outer and inner noises on the functioning of finite non-autonomous automata from some classes. Formulas for the probability of the change of output symbol as a result of noises in input and control sequences are obtained.

M. I. Rojkov ("On the summation of Markov chains on a finite group", [9], pp. 195–214) finds conditions under which the sum of Markov chains on a finite group appears to be a Markov chain too.

S. Ju. Mel'nikov ("Polyhedra characterizing the statistical properties of finite automata", [13], pp. 126–137) considers finite non-autonomous automata. He shows that the set of all possible distributions of frequencies of words from a given finite set in the input and output sequences constitutes a convex polyhedron.

Probabilistic-combinatorial problems. Probabilistic-combinatorial problems arise in different braches of cryptography; many of them are interesting and non-trivial from the viewpoint of probability theory.

G. I. Ivchenko, Ju. I. Medvedev apply methods of decomposable statistics theory to problems of random allocations of particles, random polynomials, random permutations, generalized Polya urns ("Mixtures of probabilistic distributions and random allocations", [8], pp. 169–182; "On the structure of random polynomials over finite fields", [9], pp. 111–129; "Extremal characteristics of random polynomial over a finite field", [10], pp. 71–82; "On random permutations", [11], pp. 73–92; "Investigation of urn schemes with changing parameters", [12], pp. 64–81; "On a class of nonuniform permutations of a random order", [13], pp. 75–88; "Statistics of a parametric model of random substitutions", [14], pp. 116–127).

B. A. Sevastyanov ("Probability distribution of permanents of random matrices with independent elements from a field GF(p)", [9], pp. 235–248) describes limit distributions of permanents of random $m \times n$-matrices over the finite field GF(p). In ("Structural characteristics of some nonuniform mappings of finite sets", [12], pp. 184–193) he considers bipartite random injective mappings $f\colon X \to X$ of finite set $X = X_1 \cup X_2$ having the uniform distribution on the set of all mappings satisfying conditions $f(X_1) \subseteq X_2$, $f(X_2) \subseteq X_1$. It is proved that if $|X_2| \to \infty$ and $|X_1|^2/|X_2| \to 0$ then for every fixed $k \leq |X_1|$, for every set of different elements $\{x_1, \ldots, x_k\} \subset X_1$ and for every set $\{y_1, \ldots, y_k\} \subset X_1$

$$\mathbf{P} f(f(x_j)) = y_j, j = 1, \ldots, k = |X_1|^{-k} \left(1 + O\left(\frac{|X_1|^2}{|X_2|}\right)\right).$$

A number of limit theorems for distributions on finite groups and for products of random group elements were proved by Ju. N. Gorchinskii, I. A. Kruglov, V. M. Kapitonov, F. K. Aliev (*Gorchinskii Ju. N., Kruglov I. A., Kapitonov V. M.* "Problems of the theory of distributions on finite groups", [7], pp. 85–112; *Gorchinskii Ju. N., Kapitonov V. M.* "On the mean-square deviations in the rows of transition matrices on finite permutation groups", [8], pp. 88–100; *Aliev F. K.* "Products of independent identically distributed random variables with values in a finite simple semigroup", [8], pp. 1–20; TDM-2,3). Ju. N. Gorchinskii ("On π-automorphisms of finite groups", [10], pp. 33–50) began the study of finite group mappings which coincide with group automorphisms only on some part of the group.

V. I. Sherstnev ("Resolution of a uniform distribution on a finite Abelian group", [10], pp. 315–318) considers pairs of independent random variables with values in a finite Abelian group such that their sum has a uniform distribution on this group. He gives a geometric description of the set of pairs of distributions of such random variables and show that this set constitutes a convex polyhedron.

A sequence of independent identically distributed trials are an ideal random sequence. Investigation of its properties is necessary, for example, for the construction of statistical tests detecting differences between the properties of observed sequence and of ideal random sequence.

V. G. Mihailov, A. M. Shoitov consider sequences $\xi_1, \ldots, \xi_n$ of iid discrete random variables and prove several limit theorems for the number of pairs (i, j) such that s-tuples $(\xi_{i+1}, \ldots, \xi_{i+s})$ and $(\xi_{j+1}, \ldots, \xi_{j+s})$ are in some sense similar, for example, differ by some permutation or substitution of elements only (*Mihailov V. G.* "Inequalities for the mean number of m-tuples repetitions and for the mean number of nonappeared m-tuples from a given class", [9], pp. 147–154; "Poissonian limit theorems for the number of H-connected tuples", [13], pp. 138–155; "On the features of asymptotic behavior for the number of structurally similar tuples pairs", [14], pp. 176–185; *Shoitov A. M.* "On a feature of asymptotic distributions of the number of H-equivalent n-tuples sets in the nonequiprobable polynomial scheme", [13], pp. 227–238; "Limit distributions of random variables characterizing the connection between tuples in a polynomial scheme by means of structural equivalence", [14], pp. 312–326).

A. M. Zubkov ("Computational algorithms for distributions of sums of random variables", [11], pp. 51–60) suggests an effective method of exact computation of distributions of sums dependent random variables based on introduction of special time-inhomogeneous Markov chains.

Statistical problems. A cycle of papers by M. I. Tihomirova, V. P. Chistyakov is devoted to the investigations of statistical tests (mainly, modifications of Pearson's test) aimed to the hypotheses testing on the structure of discrete random sequence and based on the frequences of tuples composed of the elements of this sequence ("On the missing s-tuples statistical criteria", [7], pp. 265–278; "On the chi-square statistics for the output of a finite automata", [8], pp. 305–314; "Statistical tests based on s-tuples from some set", [9], pp. 295–302; "Normal approximation of multidimensional χ^2 distribution", [10], pp. 259–272; "On some characteristic of two-stage procedure for multiple hypotheses selection", [11], pp. 241–246; "Approximate computation of limit distributions of some statistics functionals", [12], pp. 213–225; "Limit distributions of some statistics connected with recurrent events", [13], pp. 201–212; "Multidimensional χ^2-statistics in disorder problems", [14], pp. 281–298).

Asymptotic efficiency of decomposable statistics was studied by G. I. Ivchenko and Ju. I. Medvedev ("On the asymptotic efficiency of decomposable statistics in a polynomial scheme", [7], pp. 121–138) and S. V. Polin ("Design of the Pitman's most efficient decomposable statistics for testing hypotheses on a superposition of random mappings", [11], pp. 189–204).

A. V. Lapshin considers sums of independent random variables taking values in a finite group and suggests some statistical estimates for the degree of nonuniformity of summand distribution based on observations of sums only ("Statistical estimation of addend distribution based on a series of observations of a sum of independent random variables on a finite Abelian group", [10], pp. 129–148; "Estimation of a distributional parameter of random variable on a finite Abelian group by the sum of its realizations with elements of a random permutation", [14], pp. 139–147).

Papers reviewed here constitute approximately half of contents of [7]–[14].

References

[1] O. A. Logachev, A. A. Sal'nikov, V. V. Jaschenko Boolean functions in coding theory and in cryptography. — MCCME, Moscow, 2004 (in Russian).

[2] D. Boneh, R. A. DeMillo, R. J. Lipton. On the importance of checking cryptographic protocols for fault. — EUROCRYPT'97, Lect. Notes Comp. Sci., 1997, v. 1233, pp. 37–51.

[3] D. Coppersmith. Small solutions to polynomial equations, and low exponent RSA vulnerabilities. — J. Cryptology, 1997, v. 10, No. 4, pp. 233–260.

[4] C. Coupé, P. Nguyen, J. Stern. The effectiveness of lattice attacks against low-exponent RSA. — PKC'99, Lect. Notes Comp. Sci., 1999, v. 1560, pp. 204–218.

[5] P. Kocher, J. Jaffe, B. Jun. Differential Power Analysis. — CRYPTO'99, Lect. Notes Comp. Sci., 1999, v. 1666, pp. 388-397.

[6] D. Wagner. Cryptanalysis of a provably secure CRT-RSA Algorithm. — CCS'04, October 25-29, 2004, Washington, DC, USA.

[7] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 1, Moscow, TVP, 1997.

[8] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 2, Moscow, TVP, 1998.

[9] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 3, Moscow, FIZMATLIT, 2000.

[10] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 4, Moscow, FIZMATLIT, 2001.

[11] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 5, Moscow, FIZMATLIT, 2002.

[12] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 6, Moscow, FIZMATLIT, 2002.

[13] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 7, Moscow, FIZMATLIT, 2003.

[14] Trudy po discretnoi matematike (V. Ja. Kozlov, ed.), in Russian. Vol. 8, Moscow, FIZMATLIT, 2004.

Software Testbed and Experiments for Exploring Counteraction of Attack and Defense Agents in the Internet

I. V. Kotenko, A. V. Ulanov

1. Introduction

At present the formalization of processes that occur in Internet is an important direction of scientific research in computer network security domain. The goal is to provide the appropriate defense mechanisms against present and emerging threats.

In the given context this problem can be considered as the problem of formalization of organizational and technical counteraction between information defense and offense systems. The solution of this problem can be based on the investigative modeling and simulation of the mentioned counteraction processes using the family of various models (from analytical to scaled-down (emulational) and full-scale) (Fig. 1).

In this paper we are developing an agent-oriented approach to the modeling and simulation of offense and defense systems' counteraction. This counteraction is represented as an antagonistic interaction between teams of software agents. The approach is stated in [1, 2, 3].

The main accent in the paper is given to two main aspects:

(1) the presentation of developed software environment (testbed) for multi-agent modeling and simulation of mentioned counteraction based on the principles of packet-level simulation (Fig. 1) and

(2) the description of the experiments on imitation of distributed denial of service attacks (DDoS) (targeted to the violation of information resources availability) and defense mechanisms realizing attack detection, prevention and pro-active reaction.

This research is being supported by grant of Russian Foundation of Basic Research (No. 04-01-00167), grant of the Department for Informational Technologies and Computation Systems of the Russian Academy of Sciences (contract No. 3.2/03) and partly funded by the EC as part of the POSITIF project (contract IST-2002-002314).

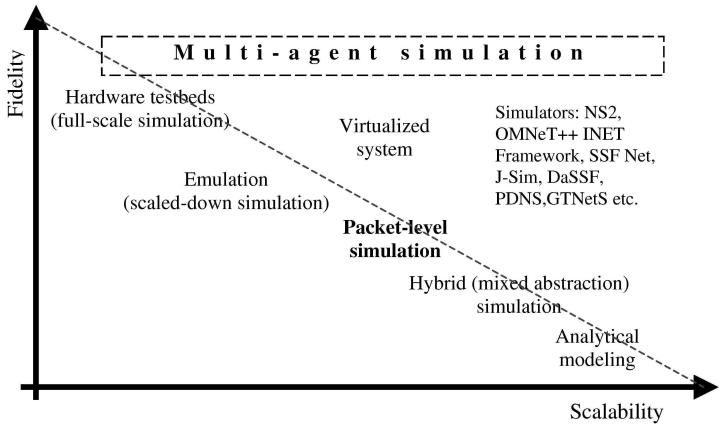

Figure 1. Family of models that are used for investigative modeling and simulation of computer network counteraction

2. The approach to modeling and simulation

The multi-agent approach for modeling and simulation of defense processes in the Internet supposes that the cybernetic counteraction is represented as the interaction of various teams of software agents [1, 2]. At least two agent teams are distinguished: the team of agents-malefactors and the defense team. They act upon computer network and each other. The agents from different teams compete to achieve contrary intentions. The agents of the same team collaborate to achieve a joint intention.

The global goal for each team is achieved by the joint efforts of many components. The components of each team have the following features: autonomy; the presence of knowledge about itself, interacting entities and the environment; the presence of knowledge or the hard-coded algorithm allowing to get and process the external data from the environment; the presence of the goal and the list of actions to achieve this goal; the fulfillment of communications for achieving the common goal.

There are a number of approaches for organizing agent teamwork. The basic approaches are as follows: joint intensions theory [4], shared plans theory [5] and combined approach [6].

In the joint intentions theory the agent team has joint commitments and intentions. Agents have individual commitments that are their per-

manent goals. The individual intention of each agent is to achieve the goal.

The team plan is the basis of shared plan theory. This plan assigns the joint fulfillment of some set of team actions. The agent team has to reach the agreement on team action fulfillment.

The combined theory unites two first approaches.

A lot of teamwork approaches are implemented in different multi-agent systems. GRATE* [7] is the implementation of joint responsibilities teamwork. The following notions are at the heart of OAA [8] framework: "blackboard" for agent communications and "facilitator" that manages it. The main idea of CAST [9] is to use the shared mental model of agents for pro-active information exchange to achieve an effective teamwork. It is supposed in RETSINA-MAS [10] that every agent has the personal copy of partial plan. This copy lets them to estimate their abilities and to choose the corresponding roles. In "Robocup Soccer" [11] agents have the joint rules and knowledge and also the individual world models. These features manage their cooperative behavior. COGNET/BATON [12] is the system for modeling the teamwork of people using intelligent agents.

The proposed approach for teamwork is based on the joint use of the elements of the joint intentions theory, shared plans theory and combined approach. It also takes into account the experience on realization of multi-agent systems.

The structure of agent team is described in terms of hierarchy of group and individual roles [1]. The leaves of the hierarchy correspond to the roles of particular agents and the intermediate nodes — to the group roles. The mechanisms of agent interaction and coordination are based on the following three groups of procedures: (1) action coordination; (2) monitoring and recovering of agent functionality; (3) communication selectivity (for the choice of the most "useful" communication acts).

The specification of action plan hierarchy is made for each of the roles. For every plan the following elements are described: the initial conditions, when the plan is proposed for execution; the conditions under which the plan is ended; the actions that are executed on the team level as a part of shared plan. The group plan has joint actions.

3. DDoS attacks and defense mechanisms

The proposed approach to the multi-agent modeling and simulation of computer network counteraction was proved on the basis of DDoS attack and defense mechanisms against them.

The main idea of DDoS attack is that the global goal — "the denial of service" of some resources — is accomplished by the joint operations of many components acting on the attack side. Thus the original task on DDoS is divided into simple subtasks that are ordered to particular specialized components. On the top level the goal remains the same for all components. On the lower level the local goals are formed. Their achievement is needed to solve the joint goal. The components are interacting with each other to coordinate the local solutions. This is needed to achieve the required quality of joint goal solution.

There are several kinds of DDoS attacks. They can be divided into two categories: exhaustion of network resources and exhaustion of host resources. The attacks are fulfilled by sending to the victim the large amount of packets (for example, UDP flood, ICMP flood, and also Smurf, Fraggle — via intermediate hosts), too long packets (Ping of Death), incorrect packets (Land), the large amount of laborious requests (TCP SYN), etc.

Building of effective defense system against DDoS is a very complex task. The usual measure to defense the subnet (not only from the DDoS attacks) is to apply the filtering rules for the packets from reserved IP addresses, protocols and ports (for example, for the incoming packets with the addresses from the internal pool, for the outgoing packets with the addresses not from the internal pool, for the packets to/from the unused ports, for the packets using unused protocols, etc.). Furthermore, the limitation on traffic for every protocol and for input/output streams can be applied.

Knowing this measures the malefactor can use such parameters of DDoS attack that it will be impossible to distinguish the attack from, e.g., the users requests caused by an increased interest to the given server. This complicates defense mechanisms.

The common approach to defense against DDoS is as follows. The information about the normal traffic for this network is collected by sensors. Then the component-analyzer compares in real-time the current traffic with the model of normal traffic. The system tries to trace back the source of abnormalities (with the help of "traceback" mechanisms) and shows the recommendations of how to sever or to lower them. The system applies the countermeasures the system administrator (or the system user) chooses.

There can be distinguished two main tasks of defense systems: attack detection and attack counteraction.

The mechanisms of attack detection can be classified by the place of deployment and by the method of detection. The components of

detection can be deployed in the attacked, the source or the intermediate sub-networks. The attack detection occurs due to the comparison of the current and model traffic. The model of normal network traffic is created using the available traffic data: either evidently, or after processing by some method. As a rule, this model is based on the load [13, 14, 15, 16, 17], on the signature [18, 19, 20], on the statistics [21, 22, 23, 24, 25, 26, 16, 17, 27], with the use both standard statistical methods and other methods (e.g., due to hierarchical system of various classifiers which can learn [30]).

The mechanisms of DDoS attack counteraction can be classified as detection mechanisms taking into account the place of deployment and the defense method used. The place of deployment is determined by the defense target. This can be the attacked, the source or the intermediate sub-networks. Besides own protection, the system of effective counteraction influences also positively on the remaining network as a whole, e.g., by blocking the attack packets within itself. The defense methods may be as follows: packet filtering (it is used in the most cases), flow filtering [26], changing the amount of resources [32, 33, 34, 27], authentication [13, 31, 35], etc.

Additionally three variants of applying the packet filtering can be distinguished. The first (traditional) variant is a standard filtering preformed on one host. The second variant is with "pushback" [14, 26, 15, 16, 17] when the filter is applied on every iteration nearer to the attack source. The third — is with "traceback" [36, 37, 38, 22, 39, 23, 24] when the source of attack is traced and the filter is applied on the nearest host (on the router).

4. Attack agent team

Attack agents are divided, at least, into two classes. They are "daemons" that realize the attack directly and "master" that coordinates the actions of other system components.

On the preliminary stage the master and daemons are deployed on available (compromised) hosts in the Internet. The important parameters on this stage are agents' amount and the degree of their distribution. Then the attack team is established: daemons send to master the messages saying they are alive and ready to work. Master stores the information about team members and their state.

The malefactor sets the common goal of the team — to perform DDoS attack with some parameters. Master receives attack parameters. Its goal is to distribute these parameters among all available daemons.

Then daemons act. Their local goal is to execute the master command. To fulfill attack they send the attack packets to the given host with the intensity (attack rate) appointed by master. After this it is believed that the goal on this stage of attack is reached.

Master asks daemons periodically to find out that they are alive and ready to work. Receiving the messages from daemons the master manages the given rate of attack. If there is no any message from one of the daemons the master makes the decision to change the attack parameters. For example, it can send to some or all daemons the commands to change the attack rate.

Daemons can execute the attack in various modes. This feature affects on the potentialities of defense team on attack detection, blocking, traceback and attack agents defeating. Daemons can send the attack packets with various rate, spoof source IP address and do it with various intensity.

Malefactor can stop the attack by sending to master the command "stop the attack". Then master distributes this command among all daemons. When they receive this command they stop the attack.

5. Defense agent team

Corresponding to the general approach there are distinguished the following defense agent classes [3]: initial data processor ("sensor"); attack detection agent ("detector"); filtering agent ("filter"); investigation agent ("investigator").

Let us describe the main functionality of these agents in one of the experiments described in the paper. In other experiments their functionality can be extended, and additional classes of agents can be deployed.

In the initial moment of time the defense agents are deployed on hosts according to their roles:

- sensor is deployed on the way of traffic to defended host;
- detector — on any host in defended subnet;
- filter — on the entrance to defended subnet;
- investigator — on any available host beyond the subnet.

The joint goal of defense team is to protect against DDoS attack. Detector watches on its accomplishing.

Sensor processes the information about network packets and collects statistic data on traffic for defended host. Sensor determines the size of overall traffic (*BPS — bit per seconds*) and the addresses of n hosts that make the greatest traffic (in developed prototype — all hosts). Its local goal is to give these parameters to detector every k seconds.

The local goal of detector is to make the decision that the attack happens. In experiments described in the paper the following method is realized. If detector determines that BPS is more than given rate (that is determined on the basis of amount of typical traffic for this subnet) then it decides that there is the DDoS attack. It sends its decision and the addresses of n hosts that make the greatest traffic to filter and investigator.

The local goal of filter is to filter the traffic on the basis of data from detector. If it was determined that the network is under attack, then filter begins to block the packets from the given hosts.

The goal of investigator is to identify and defeat attack agents. When investigator receives the message from detector it examines the given addresses on the presence of attack agents and tries to defeat identified agents. To simplify the model the admission is made that the defeating rate is 30%.

When detector determines (using data from sensors) that the attack is stopped, it believes that the joint goal of agent team is achieved at the given time interval.

6. Simulation environment

To choose the simulation tool the comprehensive analysis of the following software simulators was made: NS2 [40], OMNeT++ INET Framework [41], SSF Net [42], J-Sim [43] and some others. We used the following main requirements to the simulation environment: the detailed implementation of the protocols (from the network layer and higher) that are used in DDoS attacks (to simulate the main classes of DDoS attacks); the availability of writing and plugging in the new modules to implement the agent approach; free for use in research and educational purposes; advanced graphical user interface, etc. We discovered that the OMNET++ INET Framework satisfies these requirements best of all.

OMNET++ is a discrete event simulator [41]. The events occur inside simple modules. The exchange of messages between modules happens due to channels (modules are connected with them by the gates) or directly by gates.

We are developing now the environment for multi-agent simulation of DDoS defense and attack mechanisms on the basis of OMNeT++ INET Framework. For example, the following new modules have been created: the filtering table for network layer (for defense actions modeling); the "sniffer" that allows scanning of all traffic for the given host (to collect

the statistics for simulation the defense side actions and also for attack actions simulation). The modules that provide "sockets" were changed to accurately simulate the attack mechanisms. The agent kernels were made as co-routines, as it is convenient for implementing the interaction protocols (on which the agent teamwork is based). The other modules were made as the handlers of events from the kernel and external environment.

The example of user interface of the simulation environment is represented in Fig. 2.

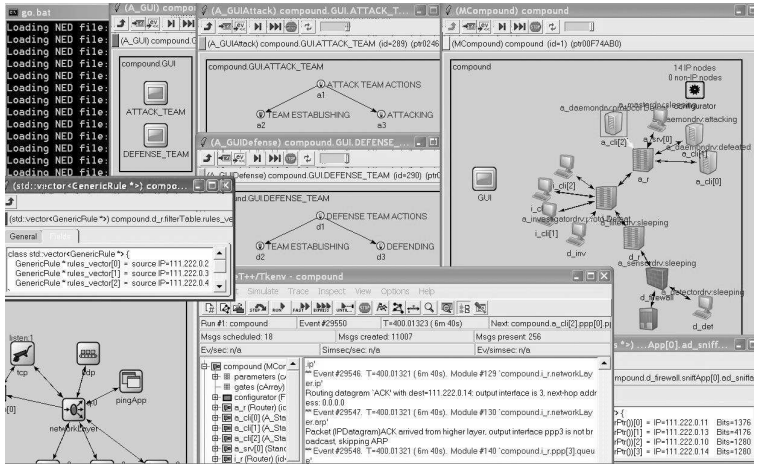

Figure 2. Example of user interface for simulation

At the basic window of visualization (Fig. 2, at upper right), a simulated computer network is displayed. The network represents a set of the hosts connected by data channels. Hosts can fulfill different functionality depending on their parameters or a set of internal modules. Internal modules provide the corresponding protocols and applications at various levels of the OSI model. Hosts are connected by channels which parameters can be changed. Applications (including agents) are deployed on hosts by connecting to corresponding protocol modules.

The window for simulation management (at the bottom of Fig. 2, in the middle) allows looking through and changing simulation parameters. There are corresponding state windows that represent the current state of agent teams (at the top of Fig. 2, in the middle). There are available several information windows that depict the functioning (or statistics

data) of particular hosts, protocols and agents. For example, the window of one of the hosts is represented in Fig. 2.

Each network for simulation consists of three sub-networks: (1) the subnet of defense where the defense team consisting from K hosts (including the defended hosts) is deployed; (2) the intermediate subnet where N hosts with generic clients are deployed; (3) the subnet of attack where the attack team is deployed, including M hosts with daemons and one host with master. The sizes of subnets may be set by the corresponding simulation parameters.

7. Experiments

Several experiments were made using the models of DDoS defense and attack processes.

Let us examine one of simple simulation scenarios to demonstrate possibilities of the software environment developed. The network for this simulation scenario is represented in Fig. 2 (at the upper right). The routers in this network are connected with each other by fiberglass channels with bandwidth 512 Mbit. The other hosts are connected by 10 Mbit Ethernet channels.

Some time after the start of simulation, clients begin to send the requests to server and it replies. That is the way generic (normal) network traffic is generated.

The formation of defense team begins some time after the start of simulation. The defense agents (investigator, sensor and filter) connect to detector. They send to detector the messages saying that they are alive and ready to work. Detector stores this information in its knowledge base. The formation of attack team occurs in the same way.

The defense team actions begin after this team formation. Sensor starts to collect the traffic statistics (the amount of transmitted bytes) for every IP-address. Detector requests data from sensor every S seconds (e.g., 60 sec). It gets statistics and detects if there is an attack. Then it connects to filter and investigator and sends them the IP-addresses of suspicious hosts.

When attack actions begin, master requests every daemon if it is alive and ready to work. When all daemons were examined, it was found that they were all in working condition. Master calculates the rate of attack for every daemon. Then master sends the corresponding attack command to every daemon. Daemons start the attack by sending, e.g., the UDP packets to the victim server with the given rate.

Sensors send to detector the list of IP addresses and the amount of bits transmitted for the given time interval. Detector determines which hosts (IP addresses) transmit the traffic that exceeds the maximum allowable size. Detector sends these addresses to filter to apply filtering rules and to investigator to trace and defeat the attack agents. After applying the filtering rules by filter the traffic to the server was lowered. And agent-investigator tries to defeat attack agents. It manages to defeat two of them. The remaining daemon continues the attack. Master redistributed the attack load for it. But the attack packets do not reach the goal and are filtered at the entrance of the defended network.

The dependence between traffic volume transmitted to the server subnet and time is represented in Fig. 3.

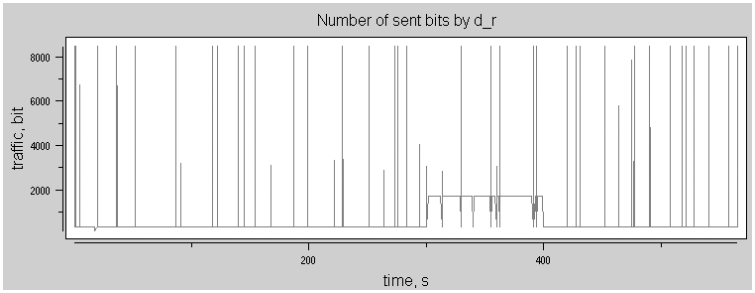

Figure 3. Dependence between the amount traffic and time

In time interval 0–300 seconds the main traffic was generated by the client requests to the server. This process is depicted by the vertical straights with low density. When the attack begins (the label of 300 seconds) the high-density traffic appears — the plateau between 300 and 400 seconds. But approximately at 400 seconds the filtering rules were applied and the attack packets began to being dropped at the entrance to the server subnet. After that the normal state returns.

8. Conclusion

In this paper we described the multi-agent environment for modeling and simulation of counteraction between the teams of malefactor and defense agents in the Internet. The environment developed is written using C++ and OMNeT++. The various classes of attacks and defense mechanisms were implemented. A set of experiments was carried out on an example of modeling and simulation of attacks “Distributed Denial

of Service". The experiments showed the effectiveness of the proposed approach and that it can be successfully used for modeling and simulation of prospective defense mechanisms and for security level analysis on the stage of network design.

Future work is connected with the further development of proposed counteraction models, including design and implementation of formal models of antagonistic interaction between the teams of defense and attack agents; implementation of greater amount of particular defense and attack mechanisms; evaluating effectiveness of implemented defense mechanisms; providing the recommendations on building of prospective defense systems against DDoS; further development of the simulation environment; investigation and improvement of agent teamwork mechanisms; developing the mechanisms of agent teams adaptation and self-learning.

References

[1] I. V. Kotenko. Multiagent models of counteracting malefactors and security systems in Internet. Third All-Russian Conference "Mathematics and security of information technologies". Moscow, Moscow State University, 2004 (in Russian).

[2] I. Kotenko. Agent-Based Modeling and Simulation of Cyber-Warfare between Malefactors and Security Agents in Internet. 19th European Simulation Multiconference "Simulation in wider Europe". ESM'05. 2005.

[3] I. Kotenko, A. Ulanov. Multiagent modeling and simulation of agents' competition for network resources availability. Second International Workshop on Safety and Security in Multiagent Systems. SASEMAS'05. 2005.

[4] P. Cohen, H. J. Levesque. Teamwork. Nous, No. 35, 1991.

[5] B. Grosz, S. Kraus. Collaborative Plans for Complex Group Actions. Artificial Intelligence, Vol.86, 1996.

[6] M. Tambe. Towards flexible teamwork. Journal of AI Research, Vol.7, 1997.

[7] N. R. Jennings. Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. Artificial Intelligence, Vol.75, No.2, 1995.

[8] D. Martin, A. Cheyer, D. Moran. The open agent architecture: A framework for building distributed software systems. Applied Artificial Intelligence, Vol.13, No.1–2, 1999.

[9] J. Yen, X. Fan, S. Sun, R. Wang, C. Chen, K. Kamali, M. Miller, R. Volz. On Modeling and Simulating Agent Teamwork in CAST. Proceedings of the Second International Conference on Active Media Technology, 2003.

[10] J. A. Giampapa, K. Sycara. Team-Oriented Agent Coordination in the RETSINA Multi-Agent System. Tech. report CMU-RI-TR-02-34, Robotics Institute, Carnegie Mellon University, December, 2002.

[11] L. A. Stankevich. A cognitive agent for soccer game. Proceeding of First Workshop of Central and Eastern Europe on Multi-agent Systems (CEEMAS'99). 1999.

[12] X. Fan, J. Yen. Modeling and Simulating Human Teamwork Behaviors Using Intelligent Agents. Journal of Physics of Life Reviews, Vol.1, No.3, 2004.

[13] C. Sangpachatanaruk, S. M. Khattab, T. Znati, R. Melhem, D. Mosse. Design and Analysis of a Replicated Elusive Server Scheme for Mitigating Denial of Service Attacks. Journal of Systems and Software, Vol.73(1), 2004.

[14] A. Keromytis, V. Misra, D. Rubenstein. SOS: Secure Overlay Services. Proceedings of ACM SIGCOMM'02, Pittsburgh, PA, 2002.

[15] T. Peng, C. Leckie, R. Kotagiri. Defending Against Distributed Denial of Service Attacks Using Selective Pushback. 9th IEEE International Conference on Telecommunications, Beijing, China, 2002.

[16] J. Ioannidis, S. M. Bellovin. Implementing Pushback: Router]-Based Defense Against DDoS Attacks. Proceedings of Symposium of Network and Distributed Systems Security (NDSS), San Diego, California, 2002.

[17] R. Manajan, S. M. Bellovin, S. Floyd, J. Ioannidis, V. Paxson, S. Shenker. Controlling High Bandwidth Aggregates in the Network. ICSI Technical Report, July 2001.

[18] Peakflow Platform. Arbor Networks. `http://www.arbornetworks.com`.

[19] DDoS-Guard. Green Gate Labs. `http://www.ddos-guard.com`.

[20] Prolexic Solutions. Prolexic. `http://www.prolexic.com`.

[21] C. Jin, H. Wang, K. G. Shin. Hop-count filtering: An effective defense against spoofed DDoS traffic. Proceedings of the 10th ACM Conference on Computer and Communications Security, 2003.

[22] K. T. Law, J. C. S. Lui, D. K. Y. Yau. You Can Run, But You Can't Hide: An Effective Methodology to Traceback DDoS Attackers. Proceedings of the 10th IEEE International Symposium on Modeling, Analysis, & Simulation of Computer & Telecommunications Systems. MASCOTS'02. 2002.

[23] A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, B. Schwartz, S. T. Kent, W. T. Strayer. Single-Packet IP Traceback. IEEE/ACM Transactions on Networking, Vol.10, No.6, 2002.

[24] J. Li, M. Sung, J. Xu, L. Li. Large-scale IP traceback in high-speed Internet: Practical Techniques and theoretical foundation. Proceedings of the IEEE Symposium on Security and Privacy. S&P'04. 2004.

[25] J. B. D. Cabrera, L. Lewis, X. Qin, W. Lee, R. K. Prasanth, B. Ravichandran, R. K. Mehra. Proactive detection of distributed denial of service attacks using mib traffic variables — a feasibility study. Proceedings of International Symposium on Integrated Network Management, 2001.

[26] D. Xuan, R. Bettati, W. Zhao. A Gateway-Based Defense System for Distributed DoS Attacks in High Speed Networks. Proceedings of the 2nd IEEE SMC Information Assurance Workshop, West Point, NY, June, 2001.

[27] J. Mirkovic, G. Prier, P. Reiher. Attacking DDoS at the Source. Proceedings of ICNP 2002, Paris, France, 2002.

[28] J. Kang, Z. Zhang, J. Ju. Protect E-Commerce against DDoS Attacks with Improved D-WARD Detection System. Proceedings of 2005 IEEE International Conference on e-Technology, 2005.

[29] Y. Xiang, W. Zhou. An Active Distributed Defense System to Protect Web Applications from DDOS Attacks. Proceedings of the Sixth International Conference on Information Integration and Web-based Applications Services, iiWAS'2004. Jakarta, Indonesia, 2004.

[30] V. Gorodetsky, O. Karsaev, V. Samoilov, A. Ulanov. Asynchronous alert correlation in multi-agent intrusion detection systems. Lecture Notes in Computer Science, Vol.3685, 2005.

[31] X. Wang, M. K. Reiter. Mitigating bandwidth-exhaustion attacks using congestion puzzles. Proceedings of the 11th ACM Conference on Computer and Communications Security, 2004.

[32] D. Mankins, R. Krishnan, C. Boyd, J. Zao, M. Frentz. Mitigating Distributed Denial of Service Attacks with Dynamic Resource Pricing. Proceedings of the 17th Annual Computer Security Applications Conference. ACSAC'01. 2001.

[33] Y. Bernet, J. Binder, S. Blake, M. Carlson, B. Carpenter, S. Keshav, E. Davies, B. Ohman, D. Verma, Z. Wang, W. Weiss. A Framework for Differentiated Services. IETF Internet Draft, 1999.

[34] H. Wang, S. G. Shin. Transport-aware IP Routers: A Built-in Protection Mechanism to Counter DDoS Attacks. IEEE Transactions on Parallel and Distributed Systems, Vol.14, No.9, 2003.

[35] H. Wang, A. Bose, M. El-Gendy, K. G. Shin. IP Easy-pass: Edge Resource Access Control. Proceedings of IEEE INFOCOM'04, Hong Kong, 2004.

[36] B. W. Gemberling, C. L. Morrow, B. R. Greene. ISP Security — Real World Techniques. Presentation, NANOG, October 2001.

[37] Y. A. Perrig, D. P. Song. A path identification mechanism to defend against DDoS attacks. Proceedings of the 2003 IEEE Symposium on Security and Privacy, 2003.

[38] S. Bellovin, M. Leech, T. Taylor. ICMP Traceback Messages. Internet-Draft draft-ietf-itrace-01.txt, October 2001.

[39] S. Savage, D. Wetherall, A. Karlin, T. Anderson. Practical network support for ip traceback. Proceedings of the 2000 ACMSIGCOMM Conference, Stockholm, Sweden, August 2000.

[40] NS-2 homepage. `http://www.isi.edu/nsnam/ns/`.

[41] OMNeT++ homepage. `http://www.omnetpp.org`.

[42] SSFNet homepage. `http://www.ssfnet.org`.

[43] J-Sim homepage. `http://www.j-sim.org`.

Public Key Infrastructure Protection of Facilities and Networks

L. Eilebrecht

I would like to start with a quick introduction on what system actually does and how it works. I am going to talk about the motivation and the threat model. I am going to explain the concept and what the system is supposed to achieve followed by some implementation details and attack scenarios of how the system works.

As you all probably already know, most solutions use public key or most solutions for encryption and signature use actually some form of public key cryptography but it can only make sense if you trust the public key you are using to encrypt some piece of data. What really matters is the integrity and atomicity of the public key. These systems usually use so-called trusted third party you may know as a certification authority (CA) actually issues a public key. So you actually have an issued signature and clients verify the issued signature and the authority it indicates in the certificate. The problem with that is clients need to trust the trusted party (that is where the name comes from). The problem with it is that such trust relationship is not desirable or not even possible in some environments. So what the CA3 fingerprint system, as it is called, provides a mechanism for clients to indicate the original certificate without relying on direct trust relationship with central authority, with a CA, who issues certificates. So author identification can be performed by client without using or relying on the issued signature of the certificate. So clients are actually enabled to detect malicious changes to certificates.

Before I continue explaining how the system actually works, a brief note about related work. We are not aware of any work in direct relation to the system we have developed, of course there are plenty of solutions and have been plenty of research, systems and solutions that try to avoid trust relationship with central authority. Some examples include system policy maker, key note or ASPKI system. The example you may very well know is open PGP which uses the web of trust to indicate public keys. The problem with work of trust is that very often people have

to revert to performing a manual verification of a public key. So they have manually verified the fingerprint, the hash of the public key. The fingerprint system is mainly based on one way hash function, so we are using hash training techniques. They are known from time stamping services. But before I explain the details of hash training and the system itself, let's look at the motivation and what the threat actually is.

Certificates signed by a trusted third party are very helpful in limiting threats like a creation of fake certificates because clients can rely on issuer signature. However the problem is the insider who has access to the CA, to the certification authority and especially to the certification authority's private keys and to the related PKI systems, can basically modify a certificate, issue new certificate or revoke a certificate. Such individual can already have an access to the system because he/she may be the system administrator taking care of the service. There is also a threat that somebody has actually gained an access to the authorized service and modified some certificates. So in one way or another the trusted third party becomes an adversary either deliberately or unintentionally. But changing the certificate doesn't do much harm itself, this is usually using the combination of so-called man in the middle attack, meaning if someone changes a certificate, especially the public key, included in the certificate, and then will be able to trick users into using this fake or misfortunate certificate, would be able with combination of man in middle attack to read victim's encrypted communication because he can re-encrypt the data sent to the wrong public key and no one would notice.

So what do we need in order to prevent this or to avoid such a trust relationship? Basically what this system does is not the preventing the initial attack because that is not possible; someone with access to database or certificates storage or issue can always do something manually. What the system does, it makes it impossible for an attacker, or insider to hide such an attack. So we enable the clients to detect such malicious changes and to use the system to indicate certificates without relying on trusted relationship, without relying on the issuer signature of the certificate. In order for a PKI client to do this, he needs to be able to perform certain verifications. This list of verifications requires actually doing that. First of all the client needs to prove the integrity of the certificate to make sure that it is still the original certificate, issued in the very beginning, that has not been modified. Apart from that a client needs to be able to verify that the certificate has not been revoked without consent of the user. In addition the client needs to be able to verify if the certificate exists or not, that the CA can not deny the issue of the

certificate. If the PKI operated in a closed environment, where we expect or require every client, every user to have a certificate, we don't need it; otherwise it will also be helpful for client to be able to obtain proof of the non-existence of the certificate. So if somebody simply doesn't have a certificate, it will be helpful for a client to actually obtain the proof that really no certificate exists versus that it has just been removed from the database. In addition the client needs to be able to verify that no duplicate certificates for given entity, for given user exists. And in order to perform these verifications, the clients need specific data to do that, I'll explain later how exactly that works. And of course clients being able to verify and validate the data, so that they are sure that they are using the correct data for verifications. What the system is intended to be is to serve as an add on to any PKI technology that it being used. Of course there are some requirements. First of all we need support on the server and client side. Especially on the client side we have to look at the usability. So we don't want any manual verifications, so the user should not and must not be required to perform manual verifications as a fingerprint check as you might know from PGP when you verify a public key. So the system must provide procedures and mechanisms to automate all the verifications. In addition the client needs to download the data for doing the verifications. So the amount of data that is required for each client needs to be reasonable. We can not require a client to download megabytes or gigabytes of data just to verify this single certificate. And of course the process itself needs to be reasonable as well in order to ensure usability. In addition the system tries to be usable for the PKI arbitrary number of users. Basically, let's say, several millions of users in PKI. But of course the system is not just an add-on or like a plain solution that you just add to existent PKI, there are some constraints and requirement for PKI. One is that the PKI environment has to ensure that no duplicate certificate for an entity allowed, so there must be only one active certificate at the given time. Of course user can have multiple certificates but they need to have different information, different user ID, different e-mail address or different information about the use of the certificate. In addition, especially if the PKI domain includes multiple CA's, we need a unique identifier for each certificate. In this case it can be just a serial ID of the certificate. And of course the whole system of protecting against the inside attacks only makes sense if the private keys are activated by the client himself. Because otherwise he doesn't need to worry about the inside attacks because private key are created and stored centrally.

I would like to focus on the concept and explain the concept with a simple solution first and later mention some of the implementations details. From the abstract point of view we have client and server side procedures that are required for the system. So basically on the CA, on the server side we have a creation of the identification information for each certification actions performed by the certification authority. This information we will later call the certificate fingerprints. The CA needs the procedure to publish this information. Consequently the client needs this procedure to download identification information, to validate it and, of course, to validate, to indentify the certificate using this information he has downloaded. So what exactly happens? As I have already mentioned the system is based on one way hash functions. Basically you can use any hash functions as SHA or Whirlpool. So whenever the CA issues a new certificate, renews a certificate or revokes a certificate, performs a certification action, a new fingerprint consisting of multiple hash values and some miter data has to be created by the CA. The values include a hash of the complete certificate, a hash of the certificate's unique identifier, as a SID (serial identifier), for example, hash of the certificate's subject data that can be user's ID, an e-mail address, a real name or a combination of this information. The certificate is actually revoked, we need the information of entry, the fingerprint entry of the relocation or the initial creation of the certificate, and we have in fact to indicate the revocation. Indicate the revocation depending on what kind of PKI technology, what kind of public key system you are using, you might not have a certificate that is issued for the revocation, in that case the hash of the certificate, won't be a hash of the certificate and will just be, for example, in case of X509 systems, a hash of the certificate revocation list entry. In addition to that we have a time stamp, which defines the exact time when the certification action took place, like when the certificate was issued by the CA. With every new fingerprint a new summary hashes is created, calculated over the previous hashes, previous fingerprints and a new fingerprint, a new hash. Together with the summary hash and a fingerprint, that is stored in so-called fingerprint list. This is a hash changing part. Hash changing ensures that you always include previous entries in your new hash calculations, that once you have edited more entries you can not change more entries in the list because otherwise you will invalidate new entries because the hash will no longer match. This data has to make available to the clients, it has to be published, it could be published in a form for intervals or some other form some kind of directory service like an ALDAP service that mostly duplicates just like

a flat file. The similar system will be a certificate divulgation list, which just includes entries for certificates.

So once the clients have this information, have this indication information, they can use it to indicate the certificate because if they have a certificate they can calculate various hashes, calculate the fingerprint and compare it to the entry on the fingerprint list. The look-up is based on the creation time. Of course that only makes sense if they are sure that they are using the correct fingerprint data. So the fingerprint data, if provided by the central authority, the clients can not trust the fingerprints, they have to validate the fingerprint data otherwise it doesn't make sense to use fingerprint data for verifications. First of all when they download data they have to do the general verification of the integrity of the list, including the general verifications of the hashes, the data, and the structure of the list. Every client has one or more certificates, so the client is the most authoritative source actually to confirm that the entries actually corresponding to their own certificates are correct because the private key and the public key can calculate the fingerprint. In addition to that it becomes the most important system which is P to P element of the system; it is called cross-client verification. Every client includes a summary hash, as I have just mentioned, which is created for each entry, includes summary hashes into communication to other clients. So we assume here that the public key infrastructure is used here to enable clients to communicate securely, it could be an e-mail, instant messaging, voice of IP. So it is every regular message to exchange a P with another user, with another entity of the PKI. A copy of the most count summary hash is included automatically by the client and includes the corresponding time. So the receiving client verifies the summary hash against its local copy of the fingerprint data. This doesn't fully indicate the fingerprint data but as it is done over and over again with every new message, the trust to fingerprint system, the integrity and atomicity of the fingerprint data is increased. And this ensures it becomes very difficult to make modifications to the fingerprint system without causing security warnings for the clients because new entries would no longer match. For example, if the client has been compromised and has committed a wrong fingerprint data, the summary hash will no longer match. So the client will detect that something is wrong with his copy of the fingerprint data.

Let me give you a quick summary. Basically at the top we have the certification authority that creates certificates, issues certificates, revokes certificates and then in addition creates a fingerprint. We can have a dedicated fingerprint authority, which just takes care of the fingerprint

list; it depends on how it is implemented. It could be the same system actually. So the fingerprint authority creates a list and makes sure that they are available to the clients like a fingerprint directory or some other form that clients are able to download the information. The PKI clients of the users download regularly or on demand the fingerprint data from the central directory and while they communicate they always include summary hash in the e-mail or in instant messaging communication and verify that automatically. So over time we basically have every client verifying that the fingerprint data is actually the same fingerprint data used by every client in a PKI system. Most importantly every client checks his own certificates and thus ensures that these entries are correct in the system. Of course a single fingerprint list and creating summary hash with every certification action would not work in large PKI if you have more than just few users. That will simply not work because client will be required to download a lot of data.

I only briefly talked about the implementation details because covering each part will take too much time but basically what we did is not a single list. We partition the list into smaller lists, basically using interval (it could be 10 minutes, 1 hour, a day) and just for each interval, which includes an arbitrary number of certification actions performed by the CA, we create multiple lists in a tree like structure for each of these intervals with leaves of the tree basically including the fingerprint, the actual fingerprint entries. At the root we have a summary hash, which is then added to a single list, which goes over time and just includes the summary hash, the corresponding time stamp. So we don't have a summary hash, a new entry for each certification actions but just a defined number based on the interval length we are using in our list. So clients can of course still choose to download all data but they can also choose to limit the data they are going to download just to the main list or just to certain information, to specific interval, to verify a particular certificate but it could be implemented on a demand way like client who is about to use the certificate can just download the data, verify the data and indicate the certificate. An extension to that is so-called fingerprint tree. The normal fingerprint list system, which I have just explained, requires the client to have a certificate and a creation time to look at the entry. If we are operating the PKI in the closed environment, where every user is required or supposed to have a certificate, we don't need a fingerprint tree. If the client has only user ID, e-mail address or domain name to perform a look up in order to obtain the certificate, it will be helpful if the client is able to prove the response from the certificate directory. For example, if the central server tells the clients there is no

certificate, the fingerprint tree can be used by the client to verify that really no certificate exists. Basically what we do is we use the hash of the user ID and use the hash value to create a hash ID, define depth of the tree and the leaves will contain the actual fingerprints. The whole fingerprint can be included or could be limited to adjust the hash of the user ID and the corresponding time stamp. So these also serve the purpose if have a list with fingerprints that the client can verify, if it contains the duplicate entry for the same user ID, which is not supposed to be allowed, of course, it has to be rebuilt with every new interval. It can be different for every fingerprint, this I have explained earlier, but basically time intervals have to be recreated. And again we have a single list just containing the summary hashes of the tree that basically will be a summary hash of the summary of the each leave lists. This again will be used for cross-client verification in order to make sure every client has the same fingerprint tree information.

Let's look at the detection of the inside attacks on some example scenarios how it actually works. The general scenario, as follows, we basically have Allis and Bob, our favorite colleges, who are the users of the same PKI, which uses the CA3 fingerprint system. So Allis and Bob communicate by e-mail and, of course, they would like to do it securely. Eve tries to compromise their communication either by attacking Allis or Bob or actually both. Eve is supposed to be an insider with an access to the CA private keys and access to the central PKI system central service.

So to the first scenario. Allis is about to send an e-mail to Bob for the first time. So she has to obtain the certificate. Eve replaces Bob's certificate in the central directory, which stores all the certificates, using a fake one, like a certificate with modified or changed the public key, so Allis's client actually obtained a fake certificate from Bob. Allis can use the creation time of the certificate to locate the corresponding entry in the fingerprint list. Allis can cable the fingerprint hashes and compass them with entry in the fingerprint list. In this case there will be no match because the entry is simply not in the fingerprint data. So Allis's client is supposed to display a warning or actually prevent Allis from sending an e-mail to Bob. Of course Eve is very clever and tries to modify the fingerprint data. What happens? We have the same scenario basically, as I have explained earlier, if the Allis's client has a copy of the fingerprint data, so later the modification will make no change because each client is supposed to download and verify fingerprint data once and then keep it. If Eve changes fingerprint data or try to fix the fingerprint list by fixing all the succeeding hashes in order to make sure

the integrity of the fingerprint data is till there. Actually all clients will notice immediately that they have some new data compared to some data that have just been modified. So many clients will show a warning that something with the fingerprint data is wrong because when downloading a new part form the fingerprint data will not match the fingerprint data they have downloaded earlier. This is considered the very end that Eve is able to feed Allis with specific data, modified data. May be if she is controlling Allis's communication, for example. So Allis's client has wrong data, Allis might end up using the fake certificate because the entry will be in the fingerprint system. However, with every message received from another user, the client will do the cross client verification, verification of the summary hash. So Allis's client will show the warning that something with a fingerprint data is not correct, basically saying if such a warning appears. There are two possibilities. Either the client has seen the warning with corrupted data or compromised data or the other sender of the message has corrupted data. The client, who has corrupted data, will receive a warning for every message that is being received while the other one will receive one just for e-mail from only a specific user. So both clients know that something is wrong with a fingerprint data.

Another variant would be if Bob is just creating the certificate. Eve would be able before all the fingerprint data is published in the certification interval to actually change the entry in the fingerprint list. In that case the fingerprint would include fingerprint entry for the fake certificate. However the clients are supposed to self verify. So Bob who actually creates the certificate will verify that actually the correct entry appears in the fingerprint list and would also notice in that case that something is wrong, probably before Allis even considers sending an e-mail to Bob.

Let me give you a quick summary of the system. Basically the system ensures that for each certification action information is being made available to the clients. Somehow the clients are enabled to audit the certification actions of the central authority. They are enabled to detect if the certificates have been modified without the consent of the owner of certificate and if something is wrong with a fingerprint data itself as a system has chosen that every client has the same fingerprint data. Thus, the system allows a secure key exchange and allows this to be implemented in automotive way. So we don't have to rely on human doing the verification, doing a manual fingerprint check. So the system can also be used for autonomy systems, devices, where no user is involved, but still needs to ensure that the certificates, the public keys that they

are using for encryption and securing communication are actually the correct system.

I would also like to note that the system is not meant to replace the conventional security properties of the certificate. So we still need a signature issuer, we still need a central certification authority. What the system actually does is remove the authority factor from central systems. So we actually have a combination of hierarchical trust model with a distributed trust model as we are exchanging some of the hashes between the clients without any central system being involved. So end up with hybrid trust model for the fingerprint system.

Question:

There is multilevel system of key distribution and a tree-like system on one certification server. How are the hashes distributed from one leave to another?

Answer:

In every tree-like structure we have a master list, which contains summary hashes of the list below that. So every note is actually a list. At the lowest level the leaves contain the fingerprints. We have a summary hash over these fingerprints which were created during that interval. The list is again based on the user ID, so they distributed over the branches of the tree. I am talking about the fingerprint list system now like the tree we create for a specific interval. So we have a summary hash for each of these leave lists, this is added to a hire list, which ends up in a master note, which contains summary hashes. A master of the summary hash is the one, which is used in the single list, for keeping summary hashes used for cross-client verification.

Question:

What kind of data is used for a single fingerprint?

Answer:

The fingerprint basically consists of the multiple hashes. Most likely we a have a hash of the complete certificate, which allows the client to verify the integrity because if the client has a certificate, he can calculate a hash and compare it to that value of the fingerprint on the fingerprint list. In addition used for looks up basically we have additional hashes with a serial ID, serial number. We have a hash of the user ID, a hash of an e-mail address. This also enables clients to look up entry in fingerprint tree system because it is based on the hash of the user ID. So when the user has an e-mail address and we would like to fetch a certificate for your e-mail address, I can calculate the hash and look the specific entry.

In addition we have some middle data like a time stamp for the creation of the fingerprint.

Question:

When are the new fingerprints added? Is this related to sending a receiving message?

Answer:

For every certification action performed by the certification authority like creation or relocation of the certificate one fingerprint has to be created. At some point, if you do it in a specific interval and to publish data, the client can download this information. The downloading of the fingerprint information and using the fingerprint data is completely unrelated to issuing the certificate or to actually using the certificate. In the very moment when the certificate is issued, clients don't have the fingerprints because it takes 10 minutes, an hour or a day before it is actually included in one of the summary hashes and in the fingerprint system. The idea is that these entries are at some point in the fingerprint system. So the client can actually not use the fingerprint system at all and still rely on the normal issuer signature but the client has at some point the data in order to be able to perform an additional identification compared to normal verification of the issuer signature and can, thus, use the fingerprint system data to indicate the certificate to make sure that it is an original certificate that was initially issued by the CA.

Question:

Does the system require a lot of resources?

Answer:

It depends, of course, on the number of the certification action you have in the system, on the umber of clients. If you have, let's say, 1000 users of PKI, especially in the closed system, we don't have much data because we don't usually regenerate the certificates every day. Even for the large public certification authority with several million users, the system is designed to require a client, a user to only download, let's say, several kilobytes of data. It will be indicating a particular certificate on average something between 20 or 30 or 40 kilobyte in order to verify the certificate, depending on how much data the client downloads in order to verify the certificate. Of course, if the client downloads all fingerprint list data which he is not meant to do, then we have data in a large PKI. But especially in closed environment, where we actually require every user to have a certificate and we don't have that many users, let's say, a few thousand, several ten thousand users may be, there is actually

not that much data because it can be downloaded as a background, it doesn't have to happen when the user is sending a message, it can be done regularly or in the background.

Question:
What kind of hash functions can we use within the system?

Answer:
Basically the system is unrelated to what kind of hash functions you would like to use. You can use SHA1, SHA2, and SHA256, basically any function you would like to use. Basically you can use any hash function or any algorithm you would like to use. The system is solely based on hashes; the implementation also includes the signature from authority to make sure it so not too easy to create wrong fingerprint data.

Question:
Are there any special requirements to the communication channels?

Answer:
Basically it depends on client application, client's use of the communication and client's use for encryption. Basically any custom encryption solutions can be included. Basically what you need is some kind of miter data in a message to include in summary. It can even be some plain text information, en e-mail. If you are using a real time communication as a voice of IP, usually they have a settled base for communication where you can exchange the summary hashes, in some cases it might even be possible to exchange the hashes first to be sure the other person has a right setting before actually sending the data.

Question:
How often has been security violated? If there were cases of security violation, did they mostly occur in state certification centers or only in commercial centers?

Answer:
Most of the attacks are actually performed by the insiders. Some statistics claim it to be 70% or more. As I mentioned in the beginning the system was designed and focuses on environments, where we need a highest possible security, where we simply can not trust the insiders like a system administrator or may be have a fear that somebody with an inside access even an outsider who is obtaining an authorized access to the system. As far as if we have ever seen such examples of attacks. I don't know any examples of the public certification authorities, any specific cases, where certificates have been modified. But there are cases, let's say, when authorities try to use SSL, to actually use the certification

authority, which is known to web-browsers and create a fake certificate. So the browser will just verify that based on the trust for the issuer signature and use the certificate. The fingerprint would not be able to detect that it is actually an incorrect fingerprint system. And, of course, in various closed environment there have been cases where people just modify certificates in order to trick user into using wrong certificates.

An Affine Equivalence and Its Application for Studying Discrete Function Properties

A. V. Cheremushkin

The report contains an overview of some tasks concerning discrete function properties related to an affine equivalence and classification results.

Introduction

There are many tasks concerning discrete function properties that are closely related to an affine or linear equivalence. This relation may appear in different ways. It may be an invariance of some properties under affine transformations, an existing of a more simpler definition or a realization, an approximation by affine functions, and so on. An affine equivalence is effectively used in classification of Boolean functions of more than five variables. Let's consider some examples of tasks of this type.

Notations

For any $n \geq 1$ let $V_n(2) = \mathrm{GF}(2)^n$ and $\mathcal{F}_n$ be a set of n variable Boolean functions. $\mathrm{GL}(n,2)$ is a linear transformations group over an n-dimensional vector space $V_n(2)$ under $\mathrm{GF}(2)$, $\mathrm{AGL}(n,2) = \mathrm{GL}(n,2)\,\mathrm{H}_n$ is an affine group, H_n is a translation group.

For any $s \geq 0$ denote by

$$\mathcal{U}_s = \mathrm{RM}(s,n) = \{f : \deg f \leq s\} \leq \mathcal{F}_n \tag{1}$$

a Reed-Muller code of order s. For $s < 0$ define $\mathcal{U}_s = \{0\}$. For any s let

$$\mathcal{U}_s^{(0)} = \mathrm{RM}_0(s,n) = \{f \in \mathcal{U}_s : f(0) = 0\}.$$

An action of a group G, $G \leq \mathrm{AGL}(n,2)$ on a set $\mathcal{U}_k/\mathcal{U}_s = \{f \oplus \mathcal{U}_s\}$, $-1 \leq s < k \leq n$, is defined by: $(f \oplus \mathcal{U}_s)^\alpha = f^\alpha \oplus \mathcal{U}_s$, $f \in \mathcal{U}_k$, $\alpha \in G$. We define a group

$$G\mathcal{U}_s = \{(\alpha, h) : \alpha \in G,\ h \in \mathcal{U}_s\} \tag{2}$$

The work was supported by the grant NSh No. 2358.2003.9.

with an operation $(\alpha, h)\cdot(\beta, f) = (\alpha\beta, h^\beta \oplus f)$, where (α, h), $(\beta, f) \in G\mathcal{U}_s$. The action on the set $\mathcal{F}_n$ is defined as $f^{(\alpha,h)} = f^\alpha \oplus h$, where $f \in \mathcal{F}_n$ and $(\alpha, h) \in G\mathcal{U}_s$. (If $G \leq \mathrm{GL}(n,2)$, then it is more convinient to use groups $G\mathcal{U}_s^{(0)}$.) Let $(G\mathcal{U}_s)_f$ be a stabilizer group of function f in a group $G\mathcal{U}_s$

$$(G\mathcal{U}_s)_f = \{(\alpha, h) \in G\mathcal{U}_s : f^{(\alpha,h)} = f\}.$$

Denote by $G_f^{(s)}$ a group $\{\alpha \in G : \exists h, (\alpha, h) \in (G\mathcal{U}_s)_f\}$. We obviously have

$$G_f^{(s)} \cong G_{\{f \oplus \mathcal{U}_s\}} \cong (G\mathcal{U}_s)_f, \tag{3}$$

$$G_f = G_f^{(-1)} \leq G_f^{(0)} \leq G_f^{(1)} \leq \ldots \leq G_f^{(\deg f)} = G.$$

Asymptotic results

C. Sannon has established triviality of stabilizer group for almost all n-variable functions as $n \to \infty$. So a number N of equivalence classes under the group G, $G \leq \mathrm{AGL}(n,2)$ may be averaged as:

$$\frac{2^{2^n}}{|G|} < N \leq \frac{2^{2^n}}{|G|}\big(1 + o(1)\big).$$

In [5] Shannon's effect is established for a wide class of groups. In [1], it is proved an asymptotic expansion for this value. For example, the expansion for the number of equivalence classes under $\mathrm{GL}(n,2)$ is

$$N \sim \frac{2^{2^n}}{|\,\mathrm{GL}(n,2)|}\left(1 + \sum_{t=1}^{\infty} N_t 2^{-2^{n-1}(1-2^{-t})}\right),$$

where

$$N_t = \frac{|\,\mathrm{GL}(n,2)|}{2^{t(2n-3t)}|\,\mathrm{GL}(t,2)| \cdot |\,\mathrm{GL}(n-2t,2)|}.$$

For $\mathrm{AGL}(n,2)$ we should replace N_t by $N_t' = 2^t N_t$.

In [38], this property is generalized for $\mathrm{AGL}(n,2)\mathcal{U}_s$ (the case $s = 1$ is studied in [53]). By the method from [38], we may prove more general result.

Theorem 1 ([22]). *Let $s = s(n) \leq \frac{n}{2}(1-\delta)$, $k = k(n) \geq \frac{n}{2}(1+\epsilon)$, $0 < \delta \leq 1$ and $0 < \epsilon \leq 1$. Then almost all forms contained in factorspace $\mathcal{U}_k/\mathcal{U}_s$ has a trivial stabilizer group in $\mathrm{AGL}(n,2)$ as $n \to \infty$.*

An interesting problem is to find the values and the upper bounds for the numbers $n_0(s)$ and $n_1(s)$, which are defined as minimal n such that there exist a n-variable function f with a trivial stabilizer group $\mathrm{GL}(n,2)_f^{(s)}$ and $\mathrm{AGL}(n,2)_f^{(s)}$, respectively. From ([12, 56, 21]) we have

$$n_0(-1) = n_1(-1) = n_0(0) = n_1(0) = 5, \quad n_0(1) = n_1(1) = 6.$$

Applying the theorem, we obtain the following.

Corollary 1. *As $s \to \infty$ we have*

$$n_0(s) \le n_1(s) \le 2s(1 + o(1)).$$

It is proved in [21] *that such functions may be constructed for $s \ge 2$ and*

$$n_0(s) \le n_1(s) \le (s+2)^2 + 1.$$

A value $A_G(f)$, which is defined as a minimal Hamming distance from a function f to a function h with a nontrivial stabilizer group G_h, is studied in [2].

Theorem 2 ([3])**.** *Let G is one of* $\mathrm{GL}(n,2)$ *and* $\mathrm{AGL}(n,2)$*. As $n \to \infty$*

$$P\left(|A_G(f) - 2^{n-3}| \le n2^{\frac{n}{2}-1}\right) \ge 1 - o(1).$$

Enumeration

This is a problem of counting a number of equivalence classes of functions under a group action. Table 1 contains the numbers of equivalence classes for $n \le 7$ and $s \le 1$ ([46, 53, 56, 22]).

Table 1.

n	1	2	3	4	5	6	7
$\mathrm{GL}(n,2)$	4	8	20	92	2 744	950 998 216	$> 2 \cdot 10^{24}$
$\mathrm{AGL}(n,2)$	3	5	10	32	382	15 768 919	$> 10^{22}$
$\mathrm{GL}(n,2)\mathcal{U}_0$	2	4	10	46	1 372	475 499 108	$> 10^{24}$
$\mathrm{AGL}(n,2)\mathcal{U}_0$	2	3	6	18	206	7 888 299	$> 8 \cdot 10^{21}$
$\mathrm{GL}(n,2)\mathcal{U}_1$	1	2	3	14	176	7 880 620	$> 8 \cdot 10^{21}$
$\mathrm{AGL}(n,2)\mathcal{U}_1$	1	2	3	8	48	150 357	$> 6 \cdot 10^{19}$

Note that linear and affine groups are acceptable only for $n \le 5$. As $n \ge 6$ we need to enlarge a transformation group or to bound a set of

functions. $\mathrm{AGL}(n,2)$ ($\mathrm{GL}(n,2)$) is maximal in symmetric (alternative) permutation group on $V_n(2)$ ($V_n(2) \setminus \{0\}$) ([9, 52, 55]), respectively. So we may to consider the set $\mathcal{U}_k/\mathcal{U}_s$, $-1 \le s < k \le n-1$. The case $s = 0$ was studied in [45]. In the case $s = 1$, we may use Polia-deBruin's theory for a linear representation of a group $\mathrm{AGL}(n,2)\mathcal{U}_1$ ([53, 54, 22]).

Hou X.-D. [49, 50] prove a general case for enumeration equivalence classes in $\mathcal{U}_k/\mathcal{U}_s$, $-1 \le s \le n-1$, under a groups $\mathrm{AGL}(n,2)$ and $\mathrm{GL}(n,2)$. Let $m_1(n,s,t)$ ($m_0(n,s,t)$) is a number of affine (linear) classes in $\mathcal{U}_t/\mathcal{U}_s$, $-1 \le s < t \le n$. The values $m_1(n,s,t)$ for $n = 6, 7$ (the symmetry relation will be formulated below) are presented in Tables 2, 3 ([49]). A bold number indicates a known classification.

Table 2.

$s \setminus t$	0	1	2	3	4	5	6
-1	**2**	**3**	**11**	205	150 357	7 888 299	15 468 919
0		**2**	**7**	120	75 761	3 947 989	
1			**4**	**34**	**2 499**		**150 357**
2				**6**	**34**		

Table 3.

$s \setminus t$	0	1	2	3	4	5	6	7
-1	**2**	**3**	**12**	3 486	30 230 045	63 379 147	8 112 499	16 244 999
					341 814	320 777 408	583 888 855	167 506 438
						548	378 066	730 294
0		**2**	**8**	1 890	15 115 039	31 689 573	4 056 249	
					412 866	670 826 699	792 080 063	
						852	701 952	
1			**4**	**179**	118 140 881	247 576 791		
					980	326 613 080		
2				**12**	68 433			

The values $m_0(n,s,t)$ for $n = 6, 7$ are presented in Tables 4, 5. Computation was produced by Lakaev K. S. using Hou's method.

Table 4.

$s \setminus t$	0	1	2	3	4	5	6
-1	**2**	**4**	**20**	1 534	7 880 620	475 499 108	950 998 216
0		**2**	**10**	767	3 940 310	237 749 554	
1			**4**	**85**	74 596		
2				**6**	**85**		

Table 5.

$s \setminus t$	0	1	2	3	4	5	6	7
-1	**2**	**4**	**22**	161 652	3 868 829	8 112 499	1 038 397	2 076 795
					382 074 516	583 617 583	981 840 994	963 681 989
						352 228	509 577 948	019 155 896
0		**2**	**11**	80 826	1 934 414	4 056 249	519 198 990	
					691 037 258	791 808 791	920 497 254	
						676 114	788 974	
1			**4**	1 596	15 115 005	31 689 573		
					928 948	649 950 738		
						696		
2				**12**	7 384 214			

Table 6 presents results [50] for the number equivalence classes of homogeneous form in $\mathcal{U}_k/\mathcal{U}_{k-1}$ under $\mathrm{GL}(n,2)$ as $6 \le n \le 11$.

Table 6.

(k,n)	
$(3,6)$	**6**
$(3,7)$	**12**
$(3,8)$	**32**
$(3,9)$	**349**
$(3,10)$	3 691 561
$(3,11)$	60 889 759 853 600
$(4,8)$	999
$(4,9)$	121 597 673 132 830
$(4,10)$	4 490 513 974 418 226 922 710 218 421 015 600
$(4,11)$	2 847 591 793 161 852 775 156 648 952 439 351 349 039 810 229
	699 431 354 841 358 028
$(5,10)$	19 749 489 318 110 485 970 697 971 583 208 968 127 316 501 515
$(5,11)$	15 503 764 406 428 075 099 751 345 714 321 442 971 845 134 712
	815 092 309 403 084 719 632 923 886 700 698 844 470 235 742
	196 625 592 840

Duality

It was pointed out in [25] that

$$\mathcal{U}_k/\mathcal{U}_{k-1} \cong \mathcal{U}_{n-k}/\mathcal{U}_{n-k-1}$$

as $1 \le k < n$. The correspondence between forms of degree k and $n-k$ is established in the following way. For $X = x_{i_1} \dots x_{i_k}$ denote

$X^o = x_{j_1} \dots x_{j_{n-k}}$, $\{j_1, \dots, j_{n-k}\} = \{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. Put into correspondence with a homogeneous form $f = \sum_s X_s$ a complement homogeneous form $f^o = \sum_s X_s^o$. For any $\beta \in \mathrm{GL}(n,2)$ denote $\beta^* = (\beta^{-1})^t$, where t denotes a transpose matrix. For any $G \le \mathrm{GL}(n,2)$ denote by G^* a dual group

$$G^* = \{\beta^* : \beta \in G\} \subseteq \mathrm{GL}(n,2).$$

Theorem 3 ([50]). *Let $1 \le k < n$. For any $f \in \mathcal{U}_k$ and any $\beta \in \mathrm{GL}(n,2)$ we have*

$$(f^\beta)^o \equiv (f^o)^{\beta^*} \pmod{\mathcal{U}_{n-k-1}},$$

where $\beta^* = (\beta^{-1})^t$. In particular,

$$\mathrm{GL}(n,2)_{f^o}^{(n-k-1)} = (\mathrm{GL}(n,2)_f^{(k-1)})^*.$$

If $t - s \ge 2$ the duality don't take place. Nevertheless there is a symmetry relation for the number $m_1(n,s,t)$ of orbits under $\mathrm{AGL}(n,2)$ (and for the number $m_0(n,s,t)$ of orbits under $\mathrm{GL}(n,2)$, analogously).

Theorem 4 ([49]). *For any $-1 \le s < t \le n$ and $i = 1, 2$ we have*

$$m_i(n,s,t) = m_i(n, n-t-1, n-s-1), \quad i = 1, 2.$$

Classification

Classification of functions under a group action is a description of all orbits (equivalence classes). In particular, we must obtain a *complete enumeration* that is to find a complete list of orbit representatives and to compute a number of functions in it.

Let us consider known results (see also [22, 23] for details). The case $n = 4$ was completely studied in [60]. A first classification for $n = 5$ was obtained for a group $\mathrm{AGL}(5,2)\mathcal{U}_1$ in [25]. In [12] it was constructed a classification for the group $\mathrm{AGL}(5,2)$. For $n = 6$ linear and affine classifications of $\mathcal{U}_3/\mathcal{U}_1$ is contained in [11] (an affine classification also is announced in [27]). In [56] (see also [43]) it was constructed a classification for the group $\mathrm{AGL}(6,2)\mathcal{U}_1$. In [16], it was obtained a linear and affine classifications of $\mathcal{F}_6/\mathcal{U}_3$ and $\mathcal{U}_4/\mathcal{U}_2$, the affine classification was announced also in [59]. For $n = 7$, the affine classification of $\mathcal{U}_3/\mathcal{U}_2$ was obtained in [15], and for $\mathcal{U}_3/\mathcal{U}_1$ — in [59].

The classification of quadratic Boolean functions is well known ([39]). The classification of cubic forms for $n = 6$ is refered in [61], and for $n = 7$

it is contained in [15, 50]. The complete list of cubic forms for $n = 8$ is presented in appendix to [50]. In [17, 18], there are constructed stabilizer groups for all orbit representatives for this case. Earlier in [63], there were computed the stabilizer group orders of all orbit representatives. A recent paper [30] claim the classification of cubic forms for $n = 9$.

Affine equivalence of vectorial functions

Consider the set of all permutations $V_n(2)$. Permutations F, G be affine (linear) equivalent iff there is a pair of affine (linear) prermutations $\alpha, \beta \in \mathrm{AGL}(n, 2)$ ($\mathrm{GL}(n, 2)$) such that

$$F(\alpha(x)) = \beta(G(x)). \tag{4}$$

The number of equivalence classes is presented in Table 7 ([46, 26]).

Table 7.

Group $G \times H$ \ n	1	2	3	4	5
$\mathrm{GL}(n,2) \times \mathrm{GL}(n,2)$	2	2	10	52 246	2 631 645 209 645 100 680 644
$\mathrm{AGL}(n,2) \times \mathrm{AGL}(n,2)$	1	1	4	302	2 569 966 041 123 963 092

In [26] it was constrained an algorithm for testing a linear (affine) equivalence (4). The basic idea of the algorithm is to test a value $F(e^i)$ for basis vectors e^i of the space $V_n(2)$, $1 \le i \le n$, and construct the substantial matrices.

Subspace of essential variables

Let $V = V_n(2)$ and V^* be a dual space containing linear functions on V. For $x \in V$ and $a^* \in V^*$ denote by (x, a^*) a value of linear function a^* on the vector x. Any function $f \in \mathcal{F}_n$ may be considered as one of representations of function from vector space V associated with a fixed basis $e^1, e^2, \ldots, e^n$. Denote by $x_e = (x_1, x_2, \ldots, x_n)$ a coordinate representation of an element $x = \sum_{i=1}^{n} x_i e^i \in V$. We have

$$f(x) = f_e(x_e) = f_e(x_1, x_2, \ldots, x_n)$$

with $x_i = (x, e^{*i})$, $i = \overline{1, n}$.

Let $0 \le t \le n-1$, $1 \le k \le n$. The variables $x_{k+1}, \ldots, x_n$ for function $f(x_1, \ldots, x_n)$ is said to be *unessential modulo* $\mathcal{U}_t$, iff there exist a function $h_e(x_1, \ldots, x_k)$ such that $f \oplus h \in \mathcal{U}_t$. It's easy to see a variable x_n is

unessential modulo $\mathcal{U}_t$ for a function f iff $D_{e_n} f \in \mathcal{U}_{t-1}$ or, in equivalent form, iff

$$\begin{pmatrix} x \\ x \oplus e_n \end{pmatrix} \in (\mathrm{H}_n)_f^{(t-1)},$$

with $e_n = (0, \ldots, 0, 1)$. We say a function f essentially depends exactly on k variables modulo $\mathcal{U}_t$, $1 \le k < n$, if

$$f(x) = f_e(x_e) \equiv h(x_1, \ldots, x_k) \pmod{\mathcal{U}_t},$$

and k is minimal with this property for any linear transformations (or equivalently for all basis). To each function f we assign two subspaces: *the subspace of essential variables modulo* $\mathcal{U}_t$ $V_1^* = \langle e^{*1}, \ldots, e^{*k} \rangle \subseteq V^*$ (any nonzero vector e^* belonging to this subspace may be added by some vectors to form a basis such that $x_1 = (x, e^*)$ be essential), and dual subspace $(V_1^*)^\perp = \{x : (x, e^*) = 0, e^* \in V_1^*\} \subseteq V$ containing vectors a with the property $\binom{x}{x+a} \in (\mathrm{H}_n)_f^{(t-1)}$. So we may use a notation $f = f(V_1^*)$.

Note that the set

$$\left\{ a \in V_n(2) : \begin{pmatrix} x \\ x \oplus a \end{pmatrix} \in (\mathrm{H}_n)_f^{(0)} \setminus (\mathrm{H}_n)_f \right\}$$

is called a *linear structure* of function f. If the functions f has a nontrivial linear structure, then there exists a basis $e^1, e^2, \ldots, e^n$ such that f_e has a variable which is essential modulo $\mathcal{U}_0$ but unessential modulo $\mathcal{U}_1$.

Linear disjunctive decomposition

Consider the function f from vector space V. If there exists a basis $e^1, e^2, \ldots, e^n$ such that f_e has a nontrivial disjunctive decomposition, then f is said to be linear decomposable. This fact may be used for investigating a structure of stabilizer group of f. Let us consider the simplest forms of disjunctive decomposition when the function f_e is a nontrivial sum or a product of some functions. For details see [22].

Theorem 5. *For any $t \ge 0$ and any $f = f(x_1, \ldots, x_n)$ the following statements are equivalent:*

(a) $|(\mathrm{H}_n^{(t-1)})_f| = 1$;

(b) $\deg D_a f \ge t$ *for any* $0 \ne a \in V_n(2)$;

(c) *the subspace of essential variables of the function f modulo $\mathcal{U}_t$ coincides with the dual space V^*.*

If there exists a basis such that a function f has a nontrivial disjunctive decomposition of f into sum of items modulo $\mathcal{U}_s$, then $f = f(V^*)$ is said to be linear decomposable into a disjunctive sum modulo $\mathcal{U}_s$. In this case there exists a direct expansion $V^* = V_1^* + V_2^*$ such that

$$f \equiv f_1(V_1^*) \oplus f_2(V_2^*) \pmod{\mathcal{U}_s}.$$

If one of the summand is a function of one variable, then investigation of initial function reduced to a dimension $n-1$.

Theorem 6. *If there exists a basis $e^1, e^2, \ldots, e^n$ such that*

$$f_e(x_1, \ldots, x_n) = h_e(x_1, \ldots, x_{n-1}) \oplus x_n$$

and $|(\mathrm{H}_n)_f| = 1$, then $|(\mathrm{H}_n)_h^{(0)}| = 1$ and the following isomorphisms hold:

$$\begin{aligned} \mathrm{GL}(n,2)_f &\cong \mathrm{GL}(n-1,2)_h^{(1)}; \\ \mathrm{AGL}(n,2)_f &\cong \mathrm{AGL}(n-1,2)_h^{(1)}; \\ \mathrm{AGL}(n,2)_f^{(0)} &\cong \mathrm{AGL}(n-1,2)_h^{(1)} \times \mathrm{H}_1\,. \end{aligned}$$

If the summand has degree two, then no reduction took place. For instance, any quadratic form is linear decomposable to disjunctive sum modulo $\mathcal{U}_1$, but a stabilizer group may be irreducible. The same time if $s \geq 2$, then the reduction is usually presents.

Lemma 1. *Let there are two expansions of V^* into direct sum*

$$V^* = V_1^* + V_2^* = U_1^* + U_2^*.$$

If for some $s \geq 2$ the function $f = f(x_1, \ldots, x_n) = f(V^*)$ has a trivial stabilizer group $(\mathrm{H}_n^{(s-1)})_f$ and

$$f \equiv f_1(V_1^*) \oplus f_2(V_2^*) \equiv h_1(U_1^*) \oplus h_2(U_2^*) \pmod{\mathcal{U}_s},$$

then the the function f is linear decomposable into sum

$$f \equiv d_1(W_{11}^*) \oplus d_2(W_{12}^*) \oplus d_3(W_{21}^*) \oplus d_4(W_{22}^*) \pmod{\mathcal{U}_s},$$

where $W_{ij}^* = V_i^* \bigcap U_j^*$, $i, j = 1, 2$.

Theorem 7. *If for some $s \geq 2$ the function $f = f(x_1, \ldots, x_n)$ has a trivial stabilizer group $(\mathrm{H}_n)_f^{(s-1)}$, and f is linear decomposable into disjunctive sum modulo $\mathcal{U}_s$, then there exists a uniquely determined linear decomposition of f into disjunctive sum modulo $\mathcal{U}_s$. The uniqueness means that any other linear decomposition of f to disjunctive sum modulo $\mathcal{U}_s$ has the same expansion of space V^*, and the summands corresponding to the same subspaces are congruent modulo $\mathcal{U}_s$.*

Corollary 2. *Let $s \geq 2$ and $K = \{f_1, \ldots, f_m\}$ is a set of linear indecomposable to disjunctive sum modulo $\mathcal{U}_s$ functions. Let for each function the dimension of a subspace of essential variables modulo $\mathcal{U}_s$ equals to the number of variables. If K is divided into affine equivalence classes modulo $\mathcal{U}_s$: $\{f_{\mu_1}, \ldots, f_{\mu_p}\}, \ldots, \{f_{\nu_1}, \ldots, f_{\nu_q}\}$, then*

$$\mathrm{AGL}(n,2)_{f_1 \oplus \cdots \oplus f_m}^{(s)} \cong [\mathrm{AGL}(n_{\mu_1},2)_{f_{\mu_1}}^{(s)}]\mathrm{S}_p \times \cdots \times [\mathrm{AGL}(n_{\nu_1},2)_{f_{\nu_1}}^{(s)}]\mathrm{S}_q.$$

The notation $[G]\mathrm{S}_p$ denotes an exponentation of group g by symmetric group S_p.

An analogous statement is true for the group $\mathrm{GL}(n,2)$.

Let us now consider the linear decomposition of a function into disjunctive product. Let $-1 \leq s \leq n-1$. We say *f has a linear (inverce linear) factor modulo $\mathcal{U}_s$*, if there exist $l = (x, a^*)$ ($l = (x, a^*) \oplus 1$), $x \in V_n(2)$, $0 \neq a^* \in V_n^*$, and a function h such that $f \overset{s}{=} l \cdot h$. We also say that f has an *affine factor modulo $\mathcal{U}_s$*, if f has a linear or inverse linear factor modulo $\mathcal{U}_s$. If f has $k \geq 1$ affine factors modulo $\mathcal{U}_s$ $l_i(x) = (x, a^{*i}) \oplus b_i$, where $a^{*i} \in V_n^*$, $x \in V_n(2)$, $b_i \in \{0,1\}$, $i \in \{\overline{1,k}\}$, such that vectors a^{*i}, $i \in \{\overline{1,k}\}$, are linearly independent, but f has not $k+1$ such factors, then we say that *f has exactly k affine factors modulo $\mathcal{U}_s$*. Note, that a subspace $\langle a^{*1}, a^{*2}, \ldots, a^{*k}\rangle$ is uniquely determined. It is easy to prove the following sentences ([18]).

Lemma 2. *For any function $l(x) = (x, a^*) \oplus b$ the following statements are equivalent:*

(a) *f has an affine factor l modulo $\mathcal{U}_s$;*

(b) *$l \cdot f \equiv f$;*

(c) *$\overline{l} \cdot f \in \mathcal{U}_{s+1}$.*

Theorem 8. *Let $k \geq 1$ and $s \leq \deg f - 1$. Then the following statements are equivalent:*

(a) *f has exactly k affine factors $\mathcal{U}_s$;*

(b) *there exist a linear transformation A such that*

$$f(xA) \equiv x_1^{b_1} \cdot \ldots \cdot x_k^{b_k} h(x_{k+1}, \ldots, x_n) \pmod{\mathcal{U}_s},$$

where $b_i \in \{0,1\}$, $i \in \{\overline{1,k}\}$ and function h has no linear factors modulo $\mathcal{U}_{s-k}$.

Let $k \in \{0, \ldots, n\}$. Denote by $\mathcal{F}_n(k)$ a set of all n-variable Boolean functions, which have exactly k affine factors. Let $f \equiv 0$ be not included in $\mathcal{F}_n(k)$, $k = \overline{1,n}$. It is easily shown, that

$$\mathcal{F}_n = \bigcup_{k=0}^{n} \mathcal{F}_n(k) \cup \{0\}.$$

As $n \geq 1$ the numbers

$$\left[\begin{array}{c} n \\ k \end{array} \right]_2 = \begin{cases} \prod\limits_{i=0}^{k-1} \frac{2^n - 2^i}{2^k - 2^i}, & \text{if } k = \overline{1,n}, \\ 1, & \text{if } k = 0, \end{cases}$$

is called Gauss coefficients.

Theorem 9 ([22])**.** *For any $1 \leq k \leq n$ we have*

$$|\mathcal{F}_n(k)| = 2^k \cdot \left[\begin{array}{c} n \\ k \end{array} \right]_2 \cdot |\mathcal{F}_{n-k}(0)|,$$

$$|\mathcal{F}_n(0)| = \sum_{k=0}^{n} (-1)^k 2^{\frac{k(k+1)}{2}} \left[\begin{array}{c} n \\ k \end{array} \right]_2 (2^{2^{n-k}} - 1) \cdot 2^k.$$

Consider a general case of a disjunctive product (see [19] for details).

Theorem 10. *Let the function $f = f(x_1, \ldots, x_n)$ has a trivial stabilizer group $(\mathrm{H}_n)_f$, has no affine factors, and is linear decomposable into disjunctive product*

$$f(V^*) = f_1(V_1^*) \cdot f_2(V_2^*) \cdot \ldots \cdot f_m(V_m^*), \quad m \geq 1.$$

Then there is a uniquely linear decomposition of f into a disjunctive product of factors, which are not linear decomposable to disjunctive product. Any other such decomposition has the same expansion of the dual space to direct sum, and the factors corresponding to the same subspaces are equals to each other.

Theorem 11. *If the function $f = f(x_1, \ldots, x_n)$ has trivial stabilizer group $(\mathrm{H}_n)_f$ and is linear decomposable into disjunctive product, then it is not linear decomposable into disjunctive sum.*

Corollary 3. *If the function $f = f(x_1, \ldots, x_n)$ has trivial stabilizer group $(\mathrm{H}_n)_f$, then it may be linear decomposable only with one type of operations $* \in \{\&, \vee, \oplus\}$.*

Fourier transform

If f is a Boolean function then Fourier transform $\widehat{f}$ is defined by equations

$$f(x) = \frac{1}{2^n} \sum_{a^* \in V_n^*(2)} \widehat{f}(a^*) \cdot (-1)^{(x,a^*)}. \tag{5}$$

$$\widehat{f}(a^*) = \sum_{x \in V_n(2)} f(x) \cdot (-1)^{(x,a^*)}. \tag{6}$$

Note, that $\widehat{f}$ is a real-valued function over the dual space $V_n^*(2)$.

Proposition 1. *If f is an invariant for G, $G \subseteq \mathrm{GL}(n,2)$, then $\widehat{f}$ is an invariant for the group*

$$G^* = \{\beta^* : \beta \in G\} \subseteq \mathrm{GL}(n,2).$$

In particular, a group G coinsides with a stabilizer group of function f in $\mathrm{GL}(n,2)$ iff the group G^* coinsides with a stabilizer group of function $\widehat{f}$ in $\mathrm{GL}(n,2)$.

Note that it is more convenient to consider a Fourier transform of a function $\chi_f = \chi(f) = (-1)^f$ which is called Walsh transform: $W_f(a^*) = \widehat{\chi_f}(a^*)$, $a^* \in V_n^*(2)$.

An approximation by affine functions

For any Boolean function f nonlinearity is defined as minimal Hamming distance to a set of all affine functions

$$\mathrm{NL}_f = \min_{g \in \mathcal{U}_1} d_H(f,g) = d_H(f, \mathcal{U}_1).$$

For vectorial function $F\colon V_n(2) \to V_m(2)$

$$\mathrm{NL}_F = \min_{0 \neq u \in V_m^*(2)} \left(\min_{g \in \mathcal{U}_1} d_H((F,u^*),g) \right).$$

A nonlinearity admits an exrpession by Walsh transform coefficients

$$\mathrm{NL}_f = 2^{n-1} - \max_{a\in V_n^*(2)} |W_f(a^*)|.$$

$$\mathrm{NL}_F = 2^{n-1} - \max_{0\neq u\in V_m^*(2)} \max_{a^*\in V_n^*(2)} |W_{(F,u^*)}(a^*)|.$$

The universal bounds for NL_f are:

- from Parseval's relation

$$\mathrm{NL}_F \le 2^{n-1} - 2^{\frac{n}{2}-1};$$

- Sidelnikov–Chabaud–Vaudenay's bound

$$\mathrm{NL}_F \le 2^{n-1} - \frac{1}{2}\sqrt{3\times 2^n - 2 - 2\frac{(2^n-1)(2^{n-1}-1)}{2^m-1}}. \quad (7)$$

The last unequality improves upon the previous bound only for $m \ge n$, and it is tight for $n = m$ odd, only. A function F is called a *Bent* function iff equality $\mathrm{NL}_F = 2^{n-1} - 2^{\frac{n}{2}-1}$ holds. The function F is called *Almost Bent, AB* ([35]) if it achieves the bound (7): $\mathrm{NL}_F = 2^{n-1} - 2^{\frac{n-1}{2}}$.

Compare the mean value for approximation by affine and quadratic functions. Let $f \in \mathcal{F}_n$. Consider a random variable

$$\varrho_n(f) = \min_{g\in\mathcal{U}_2} d_H(f,g).$$

Theorem 12 ([10])**.** $\forall x > 0$

$$\lim_{n\to\infty} P\left(\frac{\varrho_n(f) - a_n}{b_n} \le x\right) = 1 - e^{-e^x},$$

where

$$a_n = 2^{n-1} - 2^{n/2-1}\sqrt{\ln 2}\left\{n - \frac{1}{2} - \frac{\ln n}{n\ln 2} - \frac{4\ln\ln 2 + 4\ln\pi - 3\ln 2}{8n\ln 2}\right\},$$

$$b_n = \frac{2^{n/2-1}}{n\sqrt{\ln 2}}.$$

As a consequence, we obtain an estimate for the mean value for the approximation by quadratic forms:

$$\mathrm{M}\, d_H(f,\mathcal{U}_2) = 2^{n-1} - 2^{n/2-1}\sqrt{2\ln|\mathcal{U}_2|}(1+o(1)).$$

The mean value for an approximation by affine functions

$$d_H(f,\mathcal{U}_1) = \min_{g\in\mathcal{U}_1} d_H(f,g).$$

can be estimate as ([62])

$$\mathrm{M}\, d_H(f,\mathcal{U}_1) = 2^{n-1} - 2^{n/2-1}\sqrt{2\ln|\mathcal{U}_1|}(1+o(1)).$$

A "proximity" between the function and a set of linear functions may be characterized in another way. According to standard definition, the function f is linear iff for any two vectors x, y the equality $f(x) + f(y) = f(x+y)$ holds. To characterize the degree of holding this identity, we define an *additivity probability of order k* (this term was suggested by M. M. Gluhov to distinguish it from nonlinearity) by equality

$$p_k(f) = P\left(\sum_{i=1}^k f(x^{(i)}) = f(\sum_{i=1}^k x^{(i)})\right), \quad k = 2,3,\ldots.$$

We have

$$p_k(f) = \frac{1}{2}\left(1 - \frac{1}{2^{(k+1)n}} \sum_{a^*\in V_n^*(2)} W_f^{k+1}(a^*)\right).$$

The nearest value of this probabilities to 1/2 leads to most nonlinear functions.

For vectorial function $F : V_n(2) \to V_m(2)$ we have

$$p_k(F) = P\left(\sum_{i=1}^k F(x^{(i)}) = F(\sum_{i=1}^k x^{(i)})\right) =$$

$$\frac{1}{2^{(k+1)n+m}} \sum_{b^*\in V_m^*(2)} \sum_{a^*\in V_n^*(2)} W_{(F,b^*)}^{k+1}(a^*),$$

as

$$\sum_{b^*\in V_m^*(2)} \sum_{a^*\in V_n^*(2)} W_{(F,b^*)}^{k+1}(a^*) =$$

$$\sum_{b^*\in V_m^*(2)} \sum_{a^*\in V_n^*(2)} \left(\sum_{x\in V_n(2)} (-1)^{(F(x),b^*)+(x,a^*)}\right)^{k+1} =$$

$$\sum_{x^{(1)},\dots,x^{(k+1)}\in V_n(2)} \left(\sum_{b^*\in V_m^*(2)} (-1)^{(\sum_{i=1}^{k+1} F(x^{(i)}),b^*)}\right)\left(\sum_{a\in V_n^*(2)} (-1)^{(\sum_{i=1}^{k+1} x^{(i)},a^*)}\right)=$$

$$2^{n+m}\left|\left\{(x^{(1)},\dots,x^{(k)}) : \sum_{i=1}^{k} f(x^{(i)}) = f(\sum_{i=1}^{k} x^{(i)})\right\}\right|.$$

The notions of nonlinearity and 3-additivity are relative to each other in the next situation. The proof of bound (7) is based on unequality:

$$\max_{0\neq b^*\in V_m^*(2)} \max_{a\in V_n^*(2)} |W^2_{(F,b^*)}(a^*)| \geq \frac{\sum\limits_{0\neq b^*\in V_m^*(2)} \sum\limits_{a^*\in V_n^*(2)} W^4_{(F,b^*)}(a^*)}{\sum\limits_{0\neq b^*\in V_m^*(2)} \sum\limits_{a^*\in V_n^*(2)} W^2_{(F,b^*)}(a^*)}.$$

According to Parseval's relation the denominator is a constant not dependable on F. Thanks to 3-additivity relation we have

$$\sum_{b^*\in V_m^*(2)} \sum_{a^*\in V_n^*(2)} W^4_{(F,b^*)}(a^*) =$$

$$2^{n+m}|\{(x,y,z) : f(x)+f(y)+f(z)=f(x+u+z)\}| \geq$$

$$|\{(x,y,z) : x=y \vee x=z \vee y=z\}|.$$

Function is called *Almost Perfect Nonlinear, APN* iff

$$F(x)+F(y)+F(z)+F(x+y+z)=0 \iff (x=y \vee x=z \vee y=z).$$

Every AB function is APN. The property of Bent, AB and APN functions are presented in the table 8 ([35]):

Table 8.

	Bent functions	AB functions	APN functions
n	even	odd	? odd
m	$m \leq n/2$	$m=n$	$m=n$
$\forall\, a\neq 0\ D_af(x)=b$	has 1 solution	has 1 solution	≤ 2 solutions
$\forall\, u^*\neq 0\ (F,u^*)$	Bent	plateaued	plateaued

Following by [35], let us consider known classes of such functions.

Known classes of vectorial Bent functions $F : V_n \to V_m$, $n=2k$, $x,y\in \mathrm{GF}(2^k)$:

(i) $m\leq k$, $F(x,y)=L(x\times\pi(y)+g(y))$, π — permutation, g — any function, $L\colon V_k\to V_m$ — linear transform (Nyberg);

(ii) $F(x,y)=G(\frac{x}{y})$ ($\frac{x}{y}=0$ if $y=0$), G — balanced $(n/2,m)$-function.

One of the most convenient method for constructing the AB and APN functions is using a class of power functions $\tau(x)=x^d$ over the field $\mathrm{GF}(2^n)$. By this method we may attain a high nonlinearity of coordinate function linear combinations. For instance, if we use a function $f(x)=x^d$ from $V_n(2)$ with $(d,2^n-1)=1$, then coordinate function linear combinations are linear equivalent to each other. The proof of this fact is based on the following presentation of $(f(x),u^*)$, $u^*\in V_n^*(2)$, by the trace function $\mathrm{tr}:\mathrm{GF}(2^n)\to\mathrm{GF}(2)$. Let $l(x)=(x,u^*)$ is a linear function and $l(x)=\mathrm{tr}(ax)$, $a\in\mathrm{GF}(2^n)$. We have

$$(f(x),u^*)=l(f(x))=\mathrm{tr}(af(x))=\mathrm{tr}(ax^d)=\mathrm{tr}((a^{1/d}x)^d),$$

where $a^{1/d}x=Ax$ is a liner transformation. So $l(f(x))=\mathrm{tr}(f(Ax))$ with some $A\in\mathrm{GL}(n,2)$. So all coordinate function linear combinations are linear equivalent to each other and have the same nonlinearity.

Known classes of AB functions $F:V_n\to V_n, F(x)=x^d$:

(i) $d=2^k+1$, $\gcd(n,k)=1$ (Gold, 1968);

(ii) $d=2^{2k}-2^k+1$, $\gcd(n,k)=1$ (Kasami, 1971) ;

(iii) $d=2^k+3$, $n=2k+1$ (Canteaut, Charpin, Dobbertin, 2000) ;

(iv) $d=2^k+2^{k/2}-1$ if k is even, $d=2^k+2^{(3k+1)/2}-1$ if k is odd, where $n=2k+1$ (Hollman, Xiang, 2001).

Known classes of APN functions $F:V_n\to V_n$, $F(x)=x^d$:

(i) $d=2^k+1$, $\gcd(n,k)=1$ (Gold, 1968);

(ii) $d=2^{2k}-2^k+1$, $\gcd(n,k)=1$ (Kasami, 1971 ; Janwa, Wilson, 1993) ;

(iii) $d=2^k+3$, $n=2k+1$ (Dobbertin, 1999);

(iv) $d=2^k+2^{k/2}-1$ if k is even; $d=2^k+2^{(3k+1)/2}-1$ if k is odd, where $n=2k+1$ (Dobbertin, 1999);

(v) $d=2^{2k}-1$, $n=2k+1$ (Beth, Ding, 1994);

(vi) $d = 2^{4k} + 2^{3k} + 2^{2k} + 2^k - 1$, $n = 5k$ (Dobbertin, 2000).

The APN power functions listed above are not permutations when n is even. The question of knowing whether there exist APN permutations when n is even is open.

Examples of APN and AB functions which are non-equivalent to power functions are in [35].

Nonlineality degree

Consider one more parameter, which can be used to characterize "nonlineality" of function over $\mathrm{GF}(p^m)$. The *nonlineality degree* of p^m-valued function F (denote $\mathrm{dl}\, F$) is defined as a minimal m such that

$$D_{a_1} \dots D_{a_{m+1}} F(x) = 0$$

for all $a_1, \dots, a_{m+1} \in \Omega$. It is easy to see that $\mathrm{dl}\, F$ is maximal m such that for some $a_1, \dots, a_m \in \Omega$ we have

$$D_{a_1} \dots D_{a_m} F(0) \neq 0.$$

The following properties are evident.

1. $\mathrm{dl}\, D_a F \leq \mathrm{dl}\, F - 1$ *for all* $0 \neq a \in \Omega$*; there exist* $0 \neq a \in \Omega$ *such that* $\mathrm{dl}\, D_a F = \mathrm{dl}\, F - 1$.
2. $\mathrm{dl}(F_1 + F_2) \leq \max\{\mathrm{dl}\, F_1, \mathrm{dl}\, F_2\}$.
3. $\mathrm{dl}(F_1 \cdot F_2) \leq \mathrm{dl}\, F_1 + \mathrm{dl}\, F_2$. *If functions* F_1 *and* F_2 *depend on non-intersected variable sets, then* $\mathrm{dl}(F_1 \cdot F_2) = \mathrm{dl}\, F_1 + \mathrm{dl}\, F_2$.

This definition is independent on product operation. There is another standard definition of this parameter: The *nonlineality degree* of algebraic normal form F_p of p^m-valued function F is the maximal value

$$\|b_1\| + \dots + \|b_n\|$$

for all monoms $x_1^{b_1} \cdot \dots \cdot x_n^{b_n}$ in F_p, where $\|b_i\| = \sum_j a_j$ if $b_i = \sum_j a_j p^j$, $1 \leq i \leq n$. This two "additive" and "multiplicative" definitions are equivalent when we consider the function over $\mathrm{GF}(p^m)$ ([22]). Note that for Boolean function, the nonlineality degree is equal to the algebraic degree $\deg F$.

Supports and covering sequences

Support of Boolean function f is a set

$$\mathrm{supp}(f) = \{a \in V_n(2) : f(a) \neq 0\}.$$

Support of Walsh transform $\widehat{\chi_f}$ is denoted by

$$S_f = \mathrm{supp}(\widehat{\chi_f}) = \{b^* \in V_n^*(2) : \widehat{\chi_f}(b^*) \neq 0\}.$$

The following property of supports are obvious.

If $g(x) = f(x) \oplus (x, a^*)$, then $S_f = a^* + S_g$.

If $g(x) = f(xA)$, $A \in \mathrm{GL}(n, 2)$, then $S_g = A^*(S_f)$, $A^* = A^{-t}$.

If $h(x, y) = f(x) \oplus g(y)$, then $S_h = S_f \times S_g$.

If f is an odd function, then $S_f = V_n(2)$.

If $f(x) = (x, a^*) \oplus b$ is affine function, then $S_f = \{a^*\}$.

If f is a quadratic form, then S_f is a flat of even dimension.

If $f(x) = g(xA) \oplus (x, a^*) \oplus b$ is a partial bent function (g is a bent function of $2t$ variables), then S_f a flat of dimension $2t$.

Examples of balanced and bent functions, witch supports is coincides with a vector space or a flat of odd dimension, are in [34]. An effective tool for such investigation is a notion of covering sequence introduced in [33].

A *covering sequence* of switching function f is any real-valued function λ such that

$$\sum_{a \in V_n(2)} \lambda_a D_a f(x) = \rho$$

is a constant function.

Theorem 13 ([33])**.** *f is balanced iff f admits a non-trivial covering sequence. f admits a covering sequence λ if and only if $\widehat{\lambda}$ is constant function on the support S_f.*

If the function is balanced, we may use a covering sequence with $\lambda = 1$. Consider the case when the function λ has a value 0, 1, i.e., a function λ is a support for some subset. Let the support $S = \mathrm{supp}(\lambda)$ of the covering sequence λ is contained in a subspace $E \leq V_n(2)$. In this case, the function

$$\sum_{a \in E} \lambda_a D_a f(x)$$

is constant ρ if and only if its restriction to any coset $a + E$, $a \in V_n(2)$ is a constant ρ. Hence

Proposition 2 ([34]). *Let E be any subspace of $V_n(2)$. Then the function f admits a nontrivial covering sequence λ with support $S \subset E$ if and only if the restrictions of f to any coset $a+E$ (viewed as a function on E) admits the same covering sequence.*

If $D_a f(x) = 1$ for some $a \in V_n(2)$ then f is obviously balanced. So we exclude this case.

Proposition 3 ([34]). *Let E be any subspace of $V_n(2)$ and $a+E$ be any of its cosets. Let f has no derivatives equal to 1. Then f admits the indicator of $a+E$ as a covering sequence if and only if $S_f \bigcap E^{\perp} = 0$,*

$$E^{\perp} = \{b^* \in V_n^*(2) : (a, b^*) = 0, \forall a \in E\}.$$

This is equivalent to the fact that the restrictions of f to any coset $a+E$ of $V_n(2)$ is balanced. More generally, any sequence λ such that for every $a \in E$ and every $u \in V_n(2)$, $\lambda_{a+u} = \lambda_u$ is also a covering sequence of f.

Let us consider the support of plateaued functions. If the Walsh coefficients equal $0, \pm 2^{n-h}$, then by Parseval's relation we have $|S_f| = 4^h$. It may be shown, that $\deg f \le h+1$. Let affine rank of the set S be minimal dimension of a subspace E such that $S \subseteq a+E$ for some $a \in V_n(2)$, i.e., the dimension of the subspace of essential variables of function f modulo $\mathcal{U}_1$.

Theorem 14 ([14]). *If $|S_f| = 4^h$ and k is an affine rank of the support Walsh transform of the plateaued function f, then*

$$2h \le k \le 2^{2h-1} - 2^h + h.$$

Theorem 15 ([14]). *For any natural k, $2h \le k \le 2^{h+1} - 2$, there exist plateaued function with $|S_f| = 4^h$ and affine rank equals k.*

Theorem 16 ([14]). *If $h = 2$, then affine rank of the Walsh transform of the plateaued function f may be equals 4, 5, or 6.*

Algebraic immunity

A function g is called an annigilator for f iff $f \cdot g = 0$ holds. The set of all annigilators for a function f is an ideal $\mathrm{Ann}(f)$ in the ring of all Boolean polinomials generated by a function $f \oplus 1$.

An *algebraic immunity* ($\mathrm{AI}(f)$) of function f is defined as the minimal degree of nonzero annigilator for f or $f \oplus 1$, that is

$$\left(< f >_{\mathcal{F}_n} \bigcup < f \oplus 1 >_{\mathcal{F}_n} \right) \bigcap \mathcal{U}_d \ne (0). \tag{8}$$

Another definitions are ([58]):

Proposition 4. *The following statements are equivalent:*

(i) $\mathrm{AI}(f) = d$;

(ii) *d is a minimal degree of function $(f \oplus a) \cdot g \ne 0$ for all $g \in \mathcal{F}_n$ and $a \in \{0, 1\}$;*

(iii) *d is a minimal number such that there exists functions $g, h \in \mathcal{U}_d$ not equal to null simultaneously with $f \cdot g = h$.*

Proof. Equivalence (i) $\Leftrightarrow$ (ii) followed by (8). (i) $\Rightarrow$ (iii): If $\mathrm{AI}(f) = d$, $0 \ne g \in \mathcal{U}_d$ and $fg = 0$, then $h = 0$. If $(f \oplus 1)g = 0$, then $h = d$. (iii) $\Rightarrow$ (i): If $f \cdot g = h$ for some $g, h \in \mathcal{U}_d$, then $f \cdot h = f \cdot (f \cdot g) = h$ as $h \ne 0$, so $h \in \mathrm{Ann}(f \oplus 1)$. If $h = 0$, then $g \in \mathrm{Ann}(f)$. □

Note, that a class of functions with $\mathrm{AI}(f) = d$ is invariant under affine transformations, and for any $\alpha \in \mathrm{AGL}(n, 2)\mathcal{U}_1$ we have:

$$\mathrm{AI}(f) = d \quad \Rightarrow \quad \mathrm{AI}(f^{\alpha}) \le d + 1.$$

Theorem 17 ([37]). *For any n-variable Boolean functions $\mathrm{AI}(f) \le \lceil \frac{n}{2} \rceil$.*

So the classes $\mathrm{AI}_d = \{f \in \mathcal{F}_n : \mathrm{AI}(f) \le d\}$ forms an increasing sequence:

$$\mathrm{AI}_1 \subset \mathrm{AI}_2 \subset \cdots \subset \mathrm{AI}_d \subset \mathrm{AI}_{d+1} \subset \cdots \subset \mathrm{AI}_{\lceil \frac{n}{2} \rceil} = \mathcal{F}_n.$$

The upper bound for the cardinality of AI_d gives the next results.

Theorem 18 ([58]). *The probability that a random n-variable balanced function f has algebraic immunity at most d is upper bounded by the number:*

$$P(\mathrm{AI}(f) \le d) \le \frac{2\left(2^{1+n+\cdots+\binom{n}{d}} - 1\right)\binom{2^n - 2^{n-d}}{2^{n-1} - 2^{n-d}}}{\binom{2^n}{2^{n-1}}}.$$

Even though this bound is not tight, it helps to determine the asymptotic behavior of the probability.

Theorem 19 ([58]). *Let $d(n)$ be a sequence of positive integers such that $d(n) \le \mu n$ where*

$$\mu = \frac{1}{2}\left(1 + \frac{\ln n}{2} - \sqrt{\left(1 + \frac{\ln n}{2}\right)^2 - 1}\right) \approx 0.22.$$

Then $P(\mathrm{AI}(f) \le d) \to 0$ as $n \to \infty$.

Normal and k-normal functions

A Boolean function is said to be *k-normal* if the function is constant on a flat $a+U$, $a \in V_n(2)$, $U \leq V_n(2)$, of dimension $\dim U = k$. Normality of a Boolean function is the property which determines if the function is constant on a flat of dimension $\lceil \frac{n}{2} \rceil$.

It is clear that a constant function $f(x) = c$ is n-normal, an affine function $f(x) = (x, a^*) \oplus c$, $a^* \neq 0$, is $(n-1)$-normal, because it is normal on the flats $\{x : (x, a^*) = 0\}$ and $\{x : (x, a^*) = 1\}$ of dimension $n-1$.

This concept was introduced by [40], in order to construct highly nonlinear balanced Boolean functions. Later, this property was used to distinguish different classes of bent functions. As the first bent function which is nonnormal occurs for dimension 14 ([32]).

The [28] presents an asymmetric Monte Carlo algorithm to determine whether a given Boolean function is normal. The algorithm is far faster than the best known (deterministic) algorithm of Daum et al. In a first phase, it checks for flats of low dimension whether the given Boolean function is constant on them and combines such flats to flats of higher dimension in a second phase. This way, the algorithm is much faster than exhaustive search. Moreover, the algorithm benefits from randomising the first phase. As an application, in [28] is determined the level of normality for several known, highly nonlinear Boolean power functions.

Constructing the classes of functions with given properties

Note, that the existence of affine classification allows us to obtain a complete characterization of functions with given properties. Many properties are concerned with a linear and affine equivalence. For example, a set of bent, plateaued, and algebraic immune of order t functions are invariant under an affine transformation. So the problem of describing this classes is reduced to investigating the set of representatives. The problem of finding a linear structure is equivalent to a problem of finding a stabilizer group $(H_n)_f^{(0)}$, and so on.

Unfortunately, many discrete function properties are not invariant under affine transformations, such as:

- *correlation-immune of order t* ($CI(t)$);
- *resilient of order t* (balanced $CI(t)$);
- satisfy *strict avalanche criteria* (SAC) ($\|f(x) \oplus f(x \oplus a)\| = 2^{n-1}$ for all $a \in V_n(2)$, $\mathrm{wt}(a) = 1$);

- satisfy *propagation criterion* of degree p ($PC(p)$)

$$(\|f(x) \oplus f(x \oplus a)\| = 2^{n-1} \quad \text{for all } a \in V_n(2), 1 \leq \mathrm{wt}(a) \leq p);$$

and so on.

However, there is a method employs the idea behind the "change of basis" construction, which allows to describe a discrete functions properties such as $CI(1)$ and $SAC = PC(1)$. Recall that a function f is $CI(t)$ if and only if its Walsh transform W_f satisfies $W_f(b^*) = 0$, for $1 \leq \mathrm{wt}(b^*) \leq t$. A function f is $PC(p)$ if and only if its autocorrelation transform r_f

$$r_f(a) = \sum_{x \in V} (-1)^{f(x)+f(x+a)}$$

satisfies $r_f(a) = 0$, for $1 \leq \mathrm{wt}(a) \leq p$.

In order to compute a number of $CI(1)$ and $PC(1)$ functions we need to count a number of all bases in the zero sets $V^* \setminus S_f$ and $V \setminus Supp(r_f)$, respectively. Therefore, the number of functions that are affine equivalent to f and satisfy a certain property can be determined by counting basises in zero sets (see [27] for details). [27] presents an efficient approach for classification of the affine equivalence classes of cosets of the $\mathcal{U}_3/\mathcal{U}_1$ in dimension $n = 6$ with respect to cryptographic properties such as correlation immunity, resiliency and propagation characteristics and its intersections.

In [51] the affine classification of cubic forms in 8 variables is extended to the affine classification of cubic bent functions.

Papers [36, 29] investigate cubic $(n-4)$-resilient functions. It is proved, that the dimension of subspase of essential variables modulo $\mathcal{U}_1$ is not greater than 6 and there are exactly seven equivalence classes of functions under a group $\mathrm{GL}(n, 2))\mathcal{U}_1$, which contain such functions. Note that in [36, 29] it is proved a number of $(n-4)$-resilient n-variable functions for $n \leq 10$.

References

[1] Amrosimov A. S., Sharov N. N. Some asymptotic expansions for a number of functions with a trivial stabilizer group. // Cybernetic Probl. – 1979. – No. 36. – P.65–86. (in Russian)

[2] Denisov O. V. Majority function in Shannon's effect for Boolean functions under a symmetric group.// Discrete Math. – Vol. 5. – No. 3. – 1993. – C. 64–75. (in Russian)

[3] Denisov O. V. Binary codes consisting of functions with nontrivial stabilizer group. // Mathematical cybernetics quiestions / Ed. by O. B. Lupanov. – No. 11. – Moscow: Phyzmathlit, 2002. – P.91–148. (in Russian)

[4] Kirienko D. P. About the number of correlation-immune and recilient functions of order $n - 4$. // Trans. VIII Inernational seminar "Discr. math. and appl." (Moscow, 2 – 6 february 2004). – Moscow: Publ. of mech.-math. dept. MSU. – 2004. – P. 421-424. (in Russian)

[5] Kloss B.M., Nechiporuk E. N. Classification of multivalued functions. // Cybern. Problems. – Moscow: Physmathgiz, 1963. – No. 9. – P. 27–36. (in Russian)

[6] Kuznetsov U.V., Shkarin S. A. Reed-Muller Codes (an overview). // Mathematical cybernetics quiestions – Moscow: Nauka - Physmathlit, 1996. – Vol. No. 6. – P. 5–50. (in Russian)

[7] Logachev O. A., Jashenko V. V., Salnikov A. A. One property of dual representations for the grop $\mathrm{GL}(n, k)$. // Discrete Math. – 2000. – Vol.12. – No.2. – P.154–159. (in Russian)

[8] Logachev O. A., Jashenko V. V., Salnikov A. A. Boolean functions in coding theory and cryptography. – Moscow: MCCME, 2004. – 469 p. (in Russian)

[9] Pogorelov B. A. Maximality of subbgroups of symmetric group on projective space under finite field. // Math. notes. – 1974. – Vol.16. – No. 1. P. 91–100. (in Russian)

[10] Ryazanov B. V., Checheta S. I. An approximation of Boolean function by the set of quadratic forms. // Discrete Math. – 1995. – Vol. 7. – No. 3. – P. 130–145. (in Russian)

[11] Semenov A. S., Cheremushkin A. V. Classification of functions of the thrird degree of six variables. // Radioelectr. Quiest. Ser. EVT. – 1988. – No. 11. – P. 132–140. (in Russian)

[12] Strazdin I. A. Affine classification of functions of five variables. // Automatics and comput. techn. – 1975. – No. 1. – P. 1–9. (in Russian)

[13] Tarannikov U. V. On correlation-immune and recilient Boolean functions. // Mathematical cybernetics quiestions / Ed. O. B. Lupanov. – No. 11. – Moscow: Phyzmathlit, 2002. – P.91–148. (in Russian)

[14] Tarannikov U. V. An affine rank of the support of plaeaued function Walsh transform. // Math. and inform. techn. security: Trans. of Conf. in MSU 28-29 october 2004. — Moscow: MCCME, 2005. - P. 226–231. (in Russian)

[15] Cheremushkin A. V. Cubic forms of seven variables. // Trans. 4 Int. seminar on discr. math. and appl. 2–4 february 1993 / Ed. O. B. Lupanov. – Moscow: Publ. of mech.-math. dept. MSU, 1998. – P. 145. (in Russian)

[16] Cheremushkin A. V. Classification of switching functions of six variables. // Trans. 4 Int. seminar on discr. math. and appl. 2–4 february 1993 / Ed. O. B. Lupanov. – Moscow: Publ. of mech.-math. dept. MSU, 1998. – P. 143–144. (in Russian)

[17] Cheremushkin A. V. Cubic forms of eight variables. // Theoretical cybernetics problems. XII International conference (Nijny Novgorod, 17–22 may 1999.). Part II. – Moscow: MSU. – 1999. – P. 245. (in Russian)

[18] Cheremushkin A. V. Methods of affine and linear classification of Boolean functions. // Trans. on discrete math. Vol.4. – Moscow.: Phiz.-math. lit. – 2001. – P. 273–314. (in Russian)

[19] Cheremushkin A. V. Uniquiness of product disjunctive decomposition of Boolean function into nonlinear irreducible factors. // Vestn. of Les's Moscow Univ. – Lesnoi vestnik. – 2004. – No. 4(35). – P. 86–190. (in Russian)

[20] Cheremushkin A. V. Problems of decomposition and linear classifisation of discrete functions. // Discr. Models in Control Syst. Theory: VI Int. Conf.: Moscow, 7-11 december 2004 / Eds. V. B. Alexeev, V. A. Zakharov, D. S. Romanov. — Moscow: Publ. CM&C dept. of MSU (licence ID No. 05899, 24.09.2001), 2004. - P. 88–92.P. 273–314. (in Russian)

[21] Cheremushkin A. V. On the functions with a trivial stabilizer group in generel affine groups. // Vestnic TSU. Suppl. Trans. Int. and Regional Conf., Simp. and Schools in TSU. – Tomsk: Publ. TSU. – No. 9(I). – August, 2004. – P. 41–44. (in Russian)

[22] Cheremushkin A. V. Decomposition and classification of discrete functions. – Moscow: TVP – OPPM, 2005. (to be appeared) (in Russian)

[23] Cheremushkin A. V. Linear and afiine classification of discrete functions. – Moscow: Mathematical cybernetics quiestions / Ed. O. B. Lupanov. – No. 14. – Moscow: Phyzmathlit, 2005. (to be appeared) (in Russian)

[24] Ashenhurst R. L. The application of counting techniques. // Proceedings of the Association for Computing Machinery, Pitsburg Meeting. – 1952. – pp. 293-305.

[25] Berlekamp E.R. and Welch L.R. Weight Distributions of the Cosets of the (32; 6) Reed-Muller Code. // IEEE Trans. Inform. Theory. – January 1972. – IT-18. – No. 1. – pp. 203–207.

[26] Biryukov A., De Canni'ere C., Braeken A., Preneel B. A Toolbox for Cryptanalysis: Linear and Affine Equivalence Algorithms. // EUROCRYPT'03. – LNCS 2656. – pp. 33-50.

[27] Braeken A., Borissov Y., Nikova S., Preneel B. Classification of Boolean Functions of 6 Variables or Less with Respect to Cryptographic Properties. `http://www.iacr.org/eprint/2004/`.

[28] Braeken A., Wolf C., Preneel B. Classification of Highly Nonlinear Boolean Power Functions with a Randomised Algorithm for Checking Normality. http://www.iacr.com/eprint/2004/214.

[29] Braeken A., Borissov Y., Nikova S., Preneel B. Classification of Cubic $(n-4)$-resilent Boolean Functions. http://www.iacr.com/eprint/2005/332.

[30] Brier E., Langevin P. Classification of Boolean Cubic Forms of Nine Variables. // 2003 Information Theory Workshop (ITW 2003). – IEEE Press, 2003. – pp. 179–182.

[31] deBruijn N. G. Generalization of Pólia's fundamental theorem in enumerative combinatorial analysis. // Nederl. Acad. Wetensch. Proc., Ser A. – vol.62. – Indag. Math. – 1959. – 21. – pp. 59–69.

[32] Canteaut A., Daum M., Dobbertin H, and Leander G. Normal and non-normal bent functions. // In Augot D., Charpin P., and Kabatianski G, editors. Workshop on Coding and Cryptography 2003. l'Ecole Superieure et d'Appliction des Transmissions, 2003. ISBN 2-7261- 1205-6. – 19 pages.

[33] Carlet C., Tarranikov Y. Covering sequences of Boolean functions and their cryptographic significance. Designes, Codes and Cryptography. – 2002. – Vol. 25. – pp. 263–279.

[34] Carlet C., Mesnager S. On the supports of Walsh transforms of Boolean functions. // url: http://eprint.iacr.org/2004/256.

[35] Carlet C. Vectorial Boolean functions for symmetric cryptography I. // http://www.cimpa-icpam.org/NotesCours/PDF/2005/Carlet05-5.pdf

[36] Carlet C., Charpin P. Cubic Boolean functions with highest resiliency. // IEEE Trans. Information Theory. — 2005. — Vol. 51. — No. 2. — 562–571.

[37] Courtois N., Meier W. Algebraic attacks on stream ciphers with linear feedback. // In Advances in Cryptology-EUROCRYPT 2003. Springer-Verlag. – 2003. – vol. LNCS 2656. – pp. 346-359.

[38] Denev J. D., Tonchev V. D. On the number of equivalence classes of Boolean functions under a transfofmation group. // IEEE Trans. Inform. Theory. – 1980. – v. 26. – No. 5. – pp. 625–626.

[39] Dixon L. E. Linear groups with exposition Galois field theory. – Leipzig, 1901. /2nd Ed. – Dover Publications, New York, 1958.

[40] Dobbertin H. Construction of Bent functions and balanced Boolean functions with high nonlinearity. // In Fast Software Encryption - FSE 1994, volume 1008 of Lecture Notes in Computer Science, pages 61-74. Bart Preneel, editor, Springer, 1994.

[41] Dobbertin H., Leander G. Cryptographer's Toolkit for Construction of 8-Bit Bent Functions. // url: http://eprint.iacr.org/2005/089.

[42] Fuller J., and Millan W. On Linear Redundancy in the AES S-Box. // FSE 2003. – LNCS 2887. – Springer-Verlag. – pp. 249–266.

[43] J. Fuller. Affine equivallence classes. http://www.isrc.qut.edu.au/people/fuller/.

[44] Harrison M. A. On the number of classes of (n,k)-switching networks. // J. Frankl. Inst. – 1963. – No. 4. – p. 313–327.

[45] Harrison M. A. The number of equivalence classes of Boolean functions under groups containing negation. // IEEE Trans. Electr. Comput. – 1963. – v. 12. – No. 5. – p. 559–561.

[46] Harrison M. A. On the classification of Boolean function by the general linear and affine groups. // J. Soc. for Indust. and Appl. Math. – 1964. – v. 12. – No. 2. – p. 285–299.

[47] Harrison M. A. Sur la classification des fonctioms logiques á plusyeurs valeurs. – Bull. Math. Soc. Sci. Math. de la R.S. de Roumantic. – 1969. – B (61). – No. 1. – pp. 41–54.

[48] Hou X.-D. Classification of cosets of the Reed-Muller code R(m-3,m). // Discrete Math. – Vol. 128. – 1994. – pp. 203–224.

[49] Hou X.-D. AGL$(m, 2)$ Acting on R(r,m)/R(s,m). // J.of Algebra. – 1995. – Vol. 171. – No. 3. – pp. 921–938.

[50] Hou X.-D. GL$(m, 2)$ Acting on R(r,m)/R(r-1,m). // Discrete Math. – Vol. 149. – 1996. – pp. 99–122.

[51] Hou X.-D. GL$(m, 2)$ Cubic bent functions. // Discrete Math. – Vol. 189. – 1998. – pp. 149–161.

[52] Kantor W. M, McDough T. P. On the maximality of PSL$(d+1, q)$, $d \geq 2$. // J. London Math. Soc. – Vol. 8. – No. 3. – p. 426.

[53] Lechner R. J. Affine equivalence of switching functions. – Ph. D. Dissertation, Harvard Univ., Cambridge. Mass., January 1963. / Submited to Bell Telephone Labs. as "Theory of switching" Harvard Computation Labs., Cambridge. Mass., Rept BL-33.

[54] Lechner R. J. A transform approach to login design. // IEEE Trans. Computers. – 1970. – v. C-19. – No. 7. – pp. 627–640.

[55] List R. On permutation groups containing PSL$_n(q)$ as a subgroup. // Geom. Dedic. – 1975. – Vol. 4. – No. 2–4. – p. 373-375.

[56] Maiorana J.A. A Classification of the Cosets of the Reed-Muller code R(1,6). // Mathematics of Computation. – July 1991. – Vol. 57. – No. 195. – pp. 403-414.

[57] Masaki S., Yoshiyuki I., Noburu T., Tadao K. Weight distribution of (128,64)–Reed–Muller Code. // IEEE Trans. Inform. Theory. – Vol. IT-17. – Sept., 1971. – pp. 627–628.

[58] Meier W., Pasalic E., Carlet C. Algebraic Attacks and Decomposition of Boolean Functions. // Eurocrypt 2004, LNCS 3027. – 2004. – pp.474–491.

[59] Meng Qing-shu, Yang min, Zhang huan-guo and Liu yu-zhen. Analysis of Affinely Equivalent Boolean Functions. // http://eprint.iacr.org/2005/025.

[60] Ninomia I. A study of the structures of boolean functions and its application to the synthesis of switching circuits. // IEEE Trans. Electronic Computers. – 1963. – v. EC-12. – p. 152.

[61] Rothaus O. S. On "bent" functions. // J. Combin. Theory. – 1976. – 20A. – pp. 300–306.

[62] Ryazanov B. V. Probabilistic methods in the theory of approximation of discrete functions. // In: Probabilistic Methods of Discrete Math.: Proc. 3rd Petrozavodsk Conf. TVP/VSP, Moskow/Utrecht, 1993, pp. 403–412.

[63] Sugita T., Kasami T., Fujiwara T. Weight distributions of the third and fifth order Reed-Muller codes of length 512. – Nara Inst. Sci. Tech. Report, Feb. 1996.

Subject Session "Mathematical Problems of Information Security"

Wreath Products in Stream Cipher Design

V. S. Anashin

Abstract

The paper develops a novel approach to stream cipher design: Both the state update function and the output function of the corresponding pseudorandom generators are compositions of arithmetic and bitwise logical operations, which are standard instructions of modern microprocessors. Moreover, both the state update function and the output function are being modified dynamically during the encryption. Also, these compositions could be keyed, so the only information available to an attacker is that these functions belong to some exponentially large class.

The paper shows that under rather loose conditions the output sequence is uniformly distributed, achieves maximum period length and has high linear complexity and ℓ-error linear complexity. Ciphers of this kind are flexible: One could choose a suitable combination of instructions to obtain due performance without affecting the quality of the output sequence. Finally, some evidence is given that a key recovery problem for (reasonably designed) stream ciphers of this kind is intractable up to plausible conjectures.

1. Introduction

A classical stream cipher is usually thought of as a pseudorandom generator which produces a keystream, that is, a binary random-looking string. Encryption procedure is just a bitwise addition modulo 2 (also called XORing) of the keystream to a plaintext, which is represented as a binary string either. That is, a pseudorandom generator is an algorithm that takes a short random string (*a key*, or *a seed*) and expands it into a very long random-looking string, a keystream.

To make software implementations of these algorithms platform-independent as well as to achieve high performance, the algorithms must use only those instructions that are common for contemporary processors. These instructions are numerical operations (addition, multiplication, subtraction,...) and logical ones (bitwise exclusive *or*, XOR,

bitwise *and*, AND, etc.). All these numerical and bitwise logical operations, and whence, all their compositions, belong to a special class of mappings from n-bit words into n-bit words: Each i^{th} bit of the output word depends only on bits $0, 1, \ldots, i$ of input words.[1] This fact underlies a number of results that enable one to determine whether a function of this kind is one-to-one, i.e., induces a permutation on n-bit words, or whether this permutation is a single cycle, or whether the function is balanced; that is, for each n-bit word the number of all its preimages is exactly the same, etc. Systematical studies of these properties for the above mentioned mappings were started by [9] and [3] (see also [4]) followed by [19], [5], [6], [7], [8], as well as by later works [17], [16], and [15].

The main goal of the paper is to present a mathematical background for a novel approach to the design of stream ciphers.[2] In this design, recurrence laws that define the key-stream are combinations of the above mentioned numerical and logical operations; moreover, these laws are being dynamically modified during encryption. Nevertheless, under minor restrictions we are able to prove that the key-stream has the longest (of possible) period, uniform distribution, and high linear complexity as well as high ℓ-error linear complexity and high 2-adic span. To give an idea of how these algorithms look like, consider the following illustrative example.

Let $m \equiv 3 \pmod 4$, $3 \le m \le \frac{2^n}{n}$. Take m *arbitrary* compositions $v_0(x), \ldots, v_{m-1}(x)$ of the above mentioned machine instructions (addition, multiplication, XOR, AND, etc.), then take another m *arbitrary* compositions $w_0(x), \ldots, w_{m-1}(x)$ of this kind. Arrange two arrays V and W writing these $v_j(x)$ and $w_j(x)$ to memory in *arbitrary* order. Now choose an arbitrary $x_0 \in \{0, 1, \ldots 2^n - 1\}$ as a seed. The generator calculates the recurrence sequence of states $x_{i+1} = (i \bmod m + x_i + 4 \cdot v_{i \bmod m}(x_i)) \bmod 2^n$ and outputs the sequence $z_i = (1 + \pi(x_i) + 4 \cdot w_{i \bmod m}(\pi(x_i))) \bmod 2^n$, where π is a bit order reverse permutation, which reads an n-bit number $z \in \{0, 1, \ldots, 2^n - 1\}$ in a reverse bit order; e.g., $\pi(0) = 0, \pi(1) = 2^{n-1}, \pi(2) = 2^{n-2}, \pi(3) = 2^{n-2} + 2^{n-1}$, etc. Then

[1] These mappings are well-known mathematical objects (however, under different names: Compatible mappings in algebra, determined functions in automata theory, triangle boolean mappings in the theory of Boolean functions, functions that satisfy Lipschitz condition with constant 1 in p-adic analysis) dating back to 1960^{th} [22], [24]. Usefulness of these mappings in cryptography has being directly pointed out since 1993 by V.S. Anashin [9], [3], [4], [5], [6], [7]. The name "T-functions" for these mappings was suggested by A. Klimov and A. Shamir in 2002 [17].

[2] This approach has been already resulted in a very fast and flexible stream cipher ABC v.2, see [10],[2].

the sequence $\{x_i\}$ of n-bit numbers is periodic; its shortest period is of length $2^n m$, and each number of $\{0, 1, \ldots, 2^n - 1\}$ occurs at the period exactly m times. Moreover, replacing each number x_i in $\{x_i\}$ by an n-bit word that is a base-2 expansion of x_i, we obtain by concatenation of these n-bit words a binary counterpart of the sequence $\{x_i\}$, i.e., a binary sequence $\{x_i\}'$ with a period of length $2^n mn$. This period is random in the sense of [18, Section 3.5, Definition Q1] (see (5) further); each k-tuple ($0 < k \le n$) occurs in this sequence $\{x_i\}'$ with frequency[3] $\frac{1}{2^k}$ exactly. The output sequence $\{z_i\}$ of numbers is also periodic; its shortest period is of length $2^n m$; each number of $\{0, 1, \ldots, 2^n - 1\}$ occurs at the period exactly m times. Finally, length of the shortest period of every binary subsequence $\{\delta_s(z_i) \colon i = 0, 1, 2, \ldots\}$ obtained by reading s^{th} bit $\delta_s(z_i)$ ($0 \le s \le n-1$) of each member of the sequence $\{z_i\}$ is a multiple of 2^n; linear complexity of this binary subsequence $\{\delta_s(z_i)\}$ (as well as linear complexity of binary counterparts $\{z_i\}'$ and $\{x_i\}'$) exceeds 2^{n-1}.

Ciphers of this kind are rather flexible. For instance, in the above example one can take $m = 2^k$ instead of odd $m \equiv 3 \pmod 4$ and replace $i \bmod m$ in the definition of the state transition functions by an arbitrary $c_i \in \{0, 1, \ldots, 2^k - 1\}$. To guarantee the above declared properties both of the state sequence and of the output sequence one must only demand that $c_0 + c_1 + \cdots + c_{m-1} \equiv 1 \pmod 2$. Moreover, one can take instead of π an arbitrary permutation of bits that takes the leftmost bit to the rightmost position (for instance, a circular 1-bit rotation towards higher order bits, which is also a standard instruction in modern microprocessors). Also, one can replace the second $+$ in the definition of the state transition and/or output functions with $\oplus$ (i.e., with XOR), or take the third summand in the form $2 \cdot (w(\pi(x) + 1) - w(\pi(x)))$ (or $2 \cdot (w(\pi(x) + 1) + \mathsf{NOT}(w(\pi(x))))$ instead of $4 \cdot w(\pi(x_i))$, etc. Once again we emphasize that both v and w could be arbitrary compositions of the above mentioned machine instructions (and derived ones); e.g., in the above example one might take[4]

$$v(x) = \left(1 + 2 \cdot \frac{(x \ \mathsf{AND}\ (x^2 + x^3)) \ \mathsf{OR}\ x^4}{3 + 4 \cdot (5 + 6x^5)^{x^6 \mathsf{XOR} x^7}}\right)^{7 + \frac{8x^8}{9 + 10x^9}}$$

We assume here and on that all the operands are non-negative integer rationals represented in their base-2 expansions; so, for instance,

[3] we count overlapping k-tuples either

[4] this example is of no practical value; it serves only to illustrate how 'crazy' the compositions could be

$2 = 1 \text{ XOR } 3 = 2 \text{ AND } 7 \equiv \text{NOT } 13 \pmod 8$, $\frac{1}{3} \equiv 3^{-1} \equiv 11 \equiv -5 \pmod{16}$, $3^{-\frac{1}{3}} \equiv 3^{11} \equiv 3^{-5} \equiv 11 \pmod{16}$, etc. Up to this agreement the functions v and w are well defined. The performance of the whole scheme depends only on the ratio of 'fast' and 'slow' operations in these compositions; one may vary this ratio in a wide range to achieve desirable speed.

The paper is organized as follows. Section 2. concerns basic facts about functions we use as 'building blocks' of our generators, Section 3. describes how to construct a generator out of these blocks, Section 4. studies properties of output sequences of these generators, and Section 5. gives some reasoning why (some of) these generators could be provably secure. Due to the space constraints, no proofs are given.

2. Preliminaries

Basically, the generator we consider in the paper is a finite automaton $\mathfrak{A} = \langle N, M, f, F, u_0\rangle$ with a finite state set N, state transition function $f : N \to N$, finite output alphabet M, output function $F : N \to M$ and an initial state (seed) $u_0 \in N$. Thus, this generator (see Figure 1) produces a sequence

$$\mathcal{S} = \{F(u_0), F(f(u_0)), F(f^{(2)}(u_0)), \dots, F(f^{(j)}(u_0)), \dots\}$$

over the set M, where

$$f^{(j)}(u_0) = \underbrace{f(\dots f(}_{j \text{ times}} u_0)\dots) \quad (j = 1, 2, \dots); \qquad f^{(0)}(u_0) = u_0.$$

Automata of the form $\mathfrak{A}$ could be used either as pseudorandom generators per se, or as components of more complicated pseudorandom generators, the so called *counter-dependent generators* (see Figure 2); the latter produce sequences $\{z_0, z_1, z_2, \dots\}$ over M according to the rule

$$z_0 = F_0(u_0), u_1 = f_0(u_0); \dots z_i = F_i(u_i), u_{i+1} = f_i(u_i); \dots \qquad (1)$$

That is, at the $(i+1)^{\text{th}}$ step the automaton $\mathfrak{A}_i = \langle N, M, f_i, F_i, u_i\rangle$ is applied to the state $u_i \in N$, producing a new state $u_{i+1} = f_i(u_i) \in N$, and outputting a symbol $z_i = F_i(u_i) \in M$.

Now we give a more formal

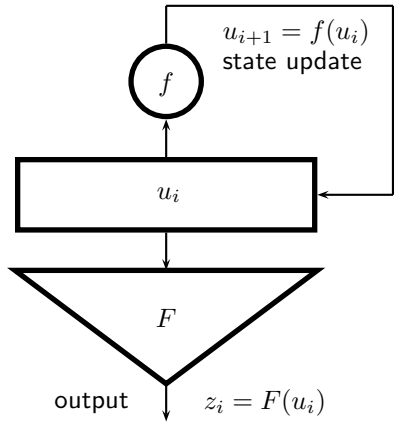

Figure 1. Ordinary PRNG

Definition 1. Let $\mathfrak{A}_j = \langle N, M, f_j, F_j\rangle$ be a family of automata with the same state set N and the same output alphabet M indexed by elements of a non-empty (possibly, countably infinite) set J (members of the family need not be necessarily pairwise distinct). Let $T\colon J \to J$ be an arbitrary mapping. A *wreath product* of the family $\{\mathfrak{A}_j\}$ of automata with respect to the mapping T is an automaton with the state set $N \times J$, state transition function $\breve{f}(j, z) = (f_j(z), T(j))$ and output function $\breve{F}(j, z) = F_j(z)$. The state transition function $\breve{f}(j, z) = (f_j(z), T(j))$ is called a *wreath product of a family of mappings* $\{f_j\colon j \in J\}$ *with respect to the mapping* T [5]. We call f_j (resp., F_j) *clock* state update (resp., output) functions.

It worth notice here that if $J = \mathbb{N}_0$ and F_i does not depend on i, this construction gives us a number of examples of counter-dependent generators in the sense of [23, Definition 2.4], where the notion of a counter-dependent generator was originally introduced. However, we use this notion in a broader sense in comparison with that of [23]: In our counter-dependent generators not only the state transition function, but also the output function depends on i. Moreover, in [23] only a special case of counter-dependent generators is studied; namely, counter-assisted generators and their cascaded and two-step modifications. A

[5] cf. *skew shift* in ergodic theory; cf. round function in the Feistel network. We are using a term from group theory.

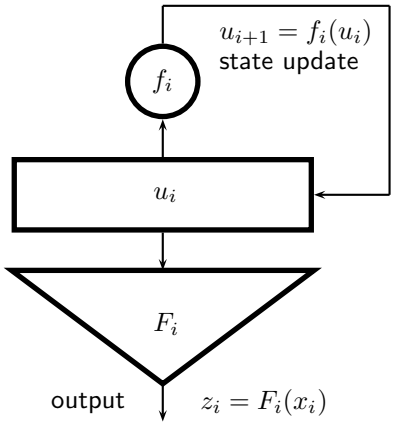

Figure 2. Counter-dependent PRNG

state transition function of a counter-assisted generator is of the form $f_i(x) = i \star h(x)$, where $\star$ is a binary quasigroup operation (in particular, group operation, e.g., + or XOR), and $h(x)$ does not depend on i. An output function of a counter-assisted generator does not depend on i either. Finally, our constructions provide long period, uniform distribution, and high linear complexity of output sequences; cf. [23], where only the diversity is guaranteed.

Throughout the paper we assume that $N = \mathbb{I}_n(p) = \{0, 1, \ldots, p^n - 1\}$, $M = \mathbb{I}_m(p)$, $m \le n$, where p is a prime. Moreover, mainly we are focused on the case $p = 2$ as the most suitable for computer implementations. It is convenient to think of elements $z \in \mathbb{I}_n(p)$ as base-p expansions of rational integers:

$$z = \delta_0^p(z) + \delta_1^p(z) \cdot p + \cdots + \delta_{n-1}^p(z) \cdot p^{n-1};$$

here $\delta_j^p(z) \in \{0, 1, \ldots, p-1\}$. For $p = 2$ we usually omit the superscript, when this does not lead to misunderstanding. Further we usually identify $\mathbb{I}_n(p)$ with the ring $\mathbb{Z}/p^n$ of residues modulo p^n.

As said above, we consider bitwise logical operators as functions defined on the set $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$ Machine instructions SHR_m and SHL_m — an m-bit right shift ($\cdot \upharpoonright m$, which is a multiplication by 2^m) and an m-bit left shift ($\cdot \upharpoonleft m$, integer division by 2^m, i.e., $\lfloor \frac{\cdot}{2^m} \rfloor$, with $\lfloor \alpha \rfloor$ being the greatest rational integer that does not exceed α) are defined on $\mathbb{N}_0$ either. *Note that since this moment throughout the paper we represent integers i in reverse bit order — less significant bits left, according to their occurrences in 2-adic canonical representation of $i = \delta_0(i) + \delta_1(i) \cdot 2 + \delta_2(i) \cdot 4 + \ldots$*; so 0011 is 12, and not 3. Moreover, one may think about these logical and machine operators, as well as of numerical, i.e., arithmetic ones (addition, multiplication, etc.), as of functions that are defined on (and valuated in) the set $\mathbb{Z}_2$ of all 2-adic integers[6] (see [3, 5]), e.g., $x \,\mathsf{OR}\, y = (\delta_0(x) \vee \delta_0(y)) + (\delta_1(x) \vee \delta_1(y)) \cdot 2 + (\delta_2(x) \vee \delta_2(y)) \cdot 2^2 + \ldots$

A common feature of the above mentioned operations is that they all, with exception of shifts towards less significant bits and circular rotations[7], are *compatible*, i.e., $\omega(u, v) \equiv \omega(u_1, v_1) \pmod{2^r}$ whenever both congruences $u \equiv u_1 \pmod{2^r}$ and $v \equiv v_1 \pmod{2^r}$ hold simultaneously. The notion of compatible mapping could be naturally generalized to multivariate mappings $(\mathbb{Z}/p^l)^t \to (\mathbb{Z}/p^l)^s$ and $(\mathbb{Z}_p)^t \to (\mathbb{Z}_p)^s$ over a residue ring modulo p^l (resp., the ring $\mathbb{Z}_p$ of p-adic integers). Obviously, a composition of compatible mappings is a compatible mapping. We list now some important examples of compatible operators $(\mathbb{Z}_p)^2 \to \mathbb{Z}_p$, p prime (see [5]). Part of them originates from arithmetic operations:

$$\begin{aligned}
&\text{multiplication, } \cdot : \ (u, v) \mapsto uv;\\
&\text{addition, } + : \ (u, v) \mapsto u + v;\\
&\text{subtraction, } - : \ (u, v) \mapsto u - v;\\
&\text{exponentiation, } \uparrow_p : \ (u, v) \mapsto u \uparrow_p v = (1 + pu)^v; \text{ in particular,}\\
&\text{raising to negative powers, } u \uparrow_p (-r) = (1 + pu)^{-r}, r \in \mathbb{N}; \text{ and}\\
&\text{division, } /_p : u/_p v = u \cdot (v \uparrow_p (-1)) = \frac{u}{1 + pv}.
\end{aligned} \tag{2}$$

The other part originates from digitwise logical operations of p-valued logic:

$$\begin{aligned}
&\text{digitwise multiplication } u \odot_p v : \delta_j(u \odot_p v) \equiv \delta_j(u)\delta_j(v) \pmod{p};\\
&\text{digitwise addition } u \oplus_p v : \delta_j(u \oplus_p v) \equiv \delta_j(u) + \delta_j(v) \pmod{p};\\
&\text{digitwise subtraction } u \ominus_p v : \delta_j(u \ominus_p v) \equiv \delta_j(u) - \delta_j(v) \pmod{p}.
\end{aligned} \tag{3}$$

[6] The latter ones within the context of this paper could be thought of as countable infinite binary sequences with members indexed by $0, 1, 2, \ldots$; $\mathbb{Z}_2$ is a metric space with respect to the 2-adic norm $\|\alpha\|_2 = 2^{-k}$, where k is the number of the first zero members of the sequence $\alpha \in \mathbb{Z}_2$: $\|0\| = \|000\ldots\|_2 = 0$, $\|1\| = \|100\ldots\|_2 = 1$, $\|2\| = \|010\ldots\|_2 = \frac{1}{2}$, etc.

[7] nevertheless, the both are used in further constructions

Here $\delta_j(z)$ $(j = 0, 1, 2, \ldots)$ stands for the j^{th} digit of z in its base-p expansion.

More compatible mappings could be derived from the above mentioned ones. For instance, a reduction modulo p^n, $n \in \mathbb{N}$, is $u \bmod p^n = u \odot_p \frac{p^n-1}{p-1}$, an l-step shift towards more significant digits is just a multiplication by p^l, etc. Obviously, $u \odot_2 v = u \text{ AND } v$, $u \oplus_2 v = u \text{ XOR } v$. Further in case $p = 2$ we omit subscripts of the corresponding operators.

In case $p = 2$ compatible mappings could be characterized in terms of Boolean functions. Namely, each mapping $T\colon \mathbb{Z}/2^n \to \mathbb{Z}/2^n$ could be considered as an ensemble of n Boolean functions τ_i^T, $i = 0, 1, 2, \ldots, n-1$, in n Boolean variables $\chi_0, \ldots, \chi_{n-1}$ by assuming $\chi_i = \delta_i(u)$, $\tau_i^T(\chi_0, \ldots, \chi_{n-1}) = \delta_i(T(u))$ for u running from 0 to $2^n - 1$. The following proposition holds.

Proposition 1. ([3, Proposition 3.9]) *A mapping $T\colon \mathbb{Z}/2^n \to \mathbb{Z}/2^n$ (resp., a mapping $T\colon \mathbb{Z}_2 \to \mathbb{Z}_2$) is compatible iff each Boolean function $\tau_i^T(\chi_0, \chi_1, \ldots) = \delta_i(T(u))$, $i = 0, 1, 2, \ldots$, does not depend on the variables $\chi_j = \delta_j(u)$ for $j > i$.*

Note. Mappings satisfying conditions of the proposition are also known in the theory of Boolean functions as *triangle* mappings; the term *T-functions* is used in [17], [16], [15] instead. For multivariate mappings Proposition 1 holds either: A mapping

$$T = (t_1, \ldots, t_s)\colon (\mathbb{Z}_2)^{(r)} \to (\mathbb{Z}_2)^{(s)}$$

is compatible iff each Boolean function

$$\tau_i^{t_j}(\chi_{1,0}, \chi_{1,1}, \ldots, \chi_{r,0}, \chi_{r,1}, \ldots) = \delta_i(t_k(u, \ldots, u_r))$$

$(i \in \mathbb{N}_0, k = 0, 1, \ldots, s)$ does not depend on the variables $\chi_{\ell,j} = \delta_j(u_\ell)$ for $j > i$ $(\ell = 1, 2, \ldots, r)$.

Now, given a compatible mapping $T\colon \mathbb{Z}_2 \to \mathbb{Z}_2$, one can define an induced mapping $T \bmod 2^n\colon \mathbb{Z}/2^n \to \mathbb{Z}/2^n$ assuming $(T \bmod 2^n)(z) = T(z) \bmod 2^n = (T(z)) \text{ AND } (2^n - 1)$ for $z = 0, 1, \ldots, 2^n - 1$. Obviously, $T \bmod 2^n$ is also compatible. For odd prime p, as well as for multivariate case $T\colon (\mathbb{Z}_p)^s \to (\mathbb{Z}_p)^t$ an induced mapping $T \bmod p^n$ could be defined by analogy.

Definition 2. (See [5]). We call a compatible mapping $T\colon \mathbb{Z}_p \to \mathbb{Z}_p$ *bijective modulo p^n* iff the induced mapping $T \bmod p^n$ is a permutation on $\mathbb{Z}/p^n$; we call T *transitive modulo p^n*, iff $T \bmod p^n$ is a permutation with a single cycle. We say that T is *measure-preserving* (respectively, *ergodic*), iff T is bijective (respectively, transitive) modulo p^n for all $n \in \mathbb{N}$. We call a compatible mapping $T\colon (\mathbb{Z}_p)^s \to (\mathbb{Z}_p)^t$ *balanced modulo p^n* iff the induced mapping $T \bmod p^n$ maps $(\mathbb{Z}/p^n)^s$ onto $(\mathbb{Z}/p^n)^t$, and each element of $(\mathbb{Z}/p^n)^t$ has the same number of preimages in $(\mathbb{Z}/p^n)^s$. Also, the mapping $T\colon (\mathbb{Z}_p)^s \to (\mathbb{Z}_p)^t$ is called *measure-preserving* iff it is balanced modulo p^n for all $n \in \mathbb{N}$.[8]

Both transitive modulo p^n and balanced modulo p^n mappings could be used as building blocks of pseudorandom generators to provide both long period and uniform distribution of output sequences. The following obvious proposition holds.

Proposition 2. *If the state transition function f of the automaton $\mathfrak{A}$ is transitive on the state set N, i.e., if f is a permutation with a single cycle of length $|N|$; if, further, $|M|$ is a factor of $|N|$, and if the output function $F : N \to M$ is balanced (i.e., $|F^{-1}(s)| = |F^{-1}(t)|$ for all $s, t \in M$), or, in particular, bijective, then the output sequence $\mathcal{S}$ of the automaton $\mathfrak{A}$ is purely periodic with a period of length $|N|$ (i.e., maximum possible), and each element of M occurs at the period the same number of times: $\frac{|N|}{|M|}$ exactly. That is, the output sequence $\mathcal{S}$ is uniformly distributed.*

Definition 3. Further in the paper we call a sequence $\mathcal{S} = \{s_i \in M\}$ over a finite set M purely periodic with a period of length t iff $s_{i+t} = s_i$ for all $i = 0, 1, 2, \ldots$. The sequence $\mathcal{S}$ is called *strictly uniformly distributed* iff it is purely periodic with a period of length t, and every element of M occurs at the period the same number of times, i.e., exactly $\frac{t}{|M|}$. A sequence $\{s_i \in \mathbb{Z}_p\}$ of p-adic integers is called *strictly uniformly distributed modulo p^k* iff the sequence $\{s_i \bmod p^k\}$ of residues modulo p^k is strictly uniformly distributed over a residue ring $\mathbb{Z}/p^k$.

Note. A sequence $\{s_i \in \mathbb{Z}_p\colon i = 0, 1, 2, \ldots\}$ of p-adic integers is uniformly distributed (with respect to the normalized Haar measure μ on $\mathbb{Z}_p$) [9] iff it is uniformly distributed modulo p^k for all $k = 1, 2, \ldots$; that is, for every $a \in \mathbb{Z}/p^k$ relative numbers of occurrences of a in the initial segment of length ℓ in the sequence $\{s_i \bmod p^k\}$ of residues modulo p^k

[8] The terms measure-preserving and ergodic originate from the theory of dynamical systems. Namely, a mapping $T\colon \mathbb{Z}_p \to \mathbb{Z}_p$ is compatible iff it satisfies Lipschitz condition with a constant 1 with respect to the p-adic metric; T defines a dynamics on the measurable space $\mathbb{Z}_p$ with respect to the normalized Haar measure. The mapping T is, e.g., ergodic with respect to this measure (in the sense of the theory of dynamical systems) iff it satisfies Definition 2, see [5] for details.

[9] i.e., $\mu(a + p^k\mathbb{Z}_p) = p^{-k}$ for all $a \in \mathbb{Z}_p$ and all $k = 0, 1, 2, \ldots$.

are asymptotically equal, i.e., $\lim_{\ell\to\infty}\frac{A(a,\ell)}{\ell}=\frac{1}{p^k}$, where $A(a,\ell)=|\{s_i\equiv a \pmod{p^k}\colon i<\ell\}|$ (see [20] for details). So strictly uniformly distributed sequences are uniformly distributed in the common meaning of the theory of distribution of sequences.

Thus, assuming $N=\mathbb{Z}/2^n, M=\mathbb{Z}/2^m, n=km$, $f=\overline{f}=\widetilde{f} \bmod 2^n$ and $F=\overline{F}=\widetilde{F} \bmod 2^m$, where the function $\widetilde{f}:\mathbb{Z}_2\to\mathbb{Z}_2$ is compatible and ergodic, and the function $\widetilde{F}:(\mathbb{Z}_2)^k\to\mathbb{Z}_2$ is compatible and measure-preserving, we obtain an automaton that generates a uniformly distributed periodic sequence, and length of a period of this sequence is 2^n. That is, each element of $\mathbb{Z}/2^m$ occurs at the period the same number of times (namely, 2^{n-m}). Obviously, the conclusion holds if one takes as F an arbitrary composition of the function $\overline{F}=\widetilde{F} \bmod 2^m$ with a measure-preserving function: For instance, one may put $F(i)=\overline{F}(\pi(i))$ or $F(i)=\delta_j(i)$, etc. Thus, Proposition 2 makes it possible to vary both the state transition and the output functions (for instance, to make them key-dependent, or in order to achieve better performance[10]) *leaving the output sequence uniformly distributed*.

There exists an easy way to construct a measure preserving or ergodic mapping out of an arbitrary compatible mapping, i.e., out of an arbitrary composition of both arithmetic (2) and logical (3) operators.

Proposition 3 ([5], Lemma 2.1 and Theorem 2.5)**.** *Let Δ be a difference operator, i.e., $\Delta g(x)=g(x+1)-g(x)$ by the definition. Let, further, p be a prime, let c be a coprime with p, $\gcd(c,p)=1$, and let $g\colon\mathbb{Z}_p\to\mathbb{Z}_p$ be a compatible mapping. Then the mapping $z\mapsto c+z+p\cdot\Delta g(z)$ $(z\in\mathbb{Z}_p)$ is ergodic, and the mapping $z\mapsto d+cx+p\cdot g(x)$ preserves measure for an arbitrary d. Moreover, if $p=2$, then the converse also holds: Each compatible and ergodic (respectively, each compatible and measure preserving) mapping $z\mapsto f(z)$ $(z\in\mathbb{Z}_2)$ could be represented as $f(x)=1+x+2\cdot\Delta g(x)$ (respectively, as $f(x)=d+x+2\cdot g(x)$) for suitable $d\in\mathbb{Z}_2$ and compatible $g\colon\mathbb{Z}_2\to\mathbb{Z}_2$.*

Corollary 1. *Let $p=2$, and let f be a compatible and ergodic mapping of $\mathbb{Z}_2$ onto itself. Then for each $n=1,2,\ldots$ the state transition function $f \bmod 2^n$ could be represented as a finite composition of bitwise logical and arithmetic operators.*

[10] e.g., in [17] there was introduced a fast generator of this kind:

$$f(x)=(x+(x^2 \text{ OR } C)) \bmod 2^{2n}, F(x)=\lfloor\frac{x}{2^n}\rfloor \bmod 2^n$$

For the sequel we need one more representation, in a Boolean form (see Proposition 1). The following theorem is just a restatement of a known result from the theory of Boolean functions, the so-called bijectivity/transitivity criterion for triangle Boolean mappings. However, the criterion belongs to the mathematical folklore; thus it is difficult to attribute it to somebody, yet a reader could find a proof in, e.g., [3, Lemma 4.8]. Recall that every Boolean function $\psi(\chi_0,\ldots\chi_n)$ in the Boolean variables $\chi_0,\ldots\chi_n$ admits a unique representation in the form

$$\psi(\chi_0,\ldots\chi_n)\equiv\sum_{\varepsilon_0,\ldots,\varepsilon_n\in\{0,1\}}\xi_{\varepsilon_0,\ldots,\varepsilon_n}\chi_0^{\varepsilon_0}\cdots\chi_i^{\varepsilon_n} \pmod 2,$$

where $\xi_{\varepsilon_0,\ldots,\varepsilon_n}\in\{0,1\}$; the sum in the right hand part is called an algebraic normal form (ANF) of the Boolean function ψ. The degree $\deg\psi$ is $\max\{\varepsilon_0+\cdots+\varepsilon_n\colon\xi_{\varepsilon_0,\ldots,\varepsilon_n}=1\}$.

Theorem 1. *A mapping $T\colon\mathbb{Z}_2\to\mathbb{Z}_2$ is compatible and measure-preserving iff for each $i=0,1,\ldots$ the ANF of the Boolean function $\tau_i^T=\delta_i(T)$ in Boolean variables $\chi_0,\ldots,\chi_i$ could be represented as*

$$\tau_i^T(\chi_0,\ldots,\chi_i)=\chi_i+\varphi_i^T(\chi_0,\ldots,\chi_{i-1}),$$

where φ_i^T is a Boolean function. The mapping T is compatible and ergodic iff, additionally, the Boolean function φ_i^T is of odd weight, that is, takes value 1 exactly at the odd number of points $(\varepsilon_0,\ldots,\varepsilon_{i-1})$, where $\varepsilon_j\in\{0,1\}$ for $j=0,1,\ldots,i-1$. The latter holds if and only if $\varphi_0^T=1$ and degree of φ_i^T for $i\geq 1$ is exactly i, that is, the ANF of φ_i^T contains a monomial $\chi_0\cdots\chi_{i-1}$.

Corollary 2. *There are exactly 2^{2^n-n-1} compatible and transitive mappings of $\mathbb{Z}/2^n$ onto $\mathbb{Z}/2^n$.*

From Theorem 1 follows an easy way to produce new ergodic functions out of given ones:

Proposition 4. *For any ergodic f and any compatible v the following functions are ergodic: $f(x+4\cdot v(x))$, $f(x\oplus(4\cdot v(x)))$, $f(x)+4\cdot v(x)$, and $f(x)\oplus(4\cdot v(x))$.*

With the use of Theorem 1 one can determine whether a given compatible mapping f preserves measure (or is ergodic) assuming it is bijective (respectively, transitive) modulo 2^n and studying behaviour of the Boolean function $\delta_n(f)$. This approach is called a bit-slice analysis

in [17], [16], and [15]. More 'analytic' techniques based on p-adic differential calculus and Mahler interpolation series were developed in [9], [3], and [5]; see also [21],[19] and [7] for various examples of compatible and ergodic functions, for instance:

- (see [9], [3]) The function $f(x) = a + a_1(x \oplus b_1) + \cdots + a_k(x \oplus b_k)$ is ergodic iff it is transitive modulo 4;
- (see [9], [3]) The function $f(x) = a + a_0 \cdot \delta_0(x) + a_1 \cdot \delta_1(x) + \cdots$ is compatible and ergodic iff $a \equiv 1 \pmod 2$, $a_0 \equiv 1 \pmod 4$, and $a_i \equiv 0 \pmod{2^i}$, $a_i \not\equiv 0 \pmod{2^{i+1}}$ for $i = 1, 2, \ldots$;
- (see [19]) The function

$$f(x) = (\ldots((((x + c_0) \oplus d_0) + c_1) \oplus d_1) + \cdots + c_m) \oplus d_m,$$

  is ergodic iff f is transitive modulo 4;
- (see [17]) The function $f(x) = x + ((x^2) \mathbin{\mathsf{OR}} c)$ is ergodic iff $c \equiv 5 \pmod 8$ or $c \equiv 7 \pmod 8$ (an equivalent statement — iff f is transitive modulo 8);
- (see [21]) The polynomial $f(x) = a_0 + a_1x + \cdots + a_dx^d$ with integral coefficients is ergodic iff the following congruences hold simultaneously:

$$\begin{gathered} a_3 + a_5 + a_7 + a_9 + \cdots \equiv 2a_2 \pmod 4; \\ a_4 + a_6 + a_8 + \cdots \equiv a_1 + a_2 - 1 \pmod 4; \\ a_1 \equiv 1 \pmod 2; \quad a_0 \equiv 1 \pmod 2 \end{gathered}$$

  (an equivalent statement — iff f is transitive modulo 8);
- (see [5]) A polynomial of degree d with rational (and not necessarily integral) coefficients is integer-valued, compatible, and ergodic iff f takes integral values at the points

$$0, 1, \ldots, 2^{\lfloor \log_2(\deg f) \rfloor + 3} - 1,$$

  and the mapping

$$z \mapsto f(z) \bmod 2^{\lfloor \log_2(\deg f) \rfloor + 3},$$

  is compatible and transitive on the residue class ring $\mathbb{Z}/2^{\lfloor \log_2 d \rfloor + 3}$ (i.e., modulo the biggest power of 2 not exceeding $8d$);
- (see [9], [3]) The entire function $f(x) = \frac{u(x)}{1+2\cdot v(x)}$, where $u(x), v(x)$ are polynomials with integral coefficients, is ergodic iff it is transitive modulo 8;
- (see [7, Example 3.6]) The function $f(x) = ax + a^x$ is ergodic iff a is odd (an equivalent statement — iff f is transitive modulo 2).

A multivariate case was studied in [15], [8]; see also [5, Theorem 3.11]. Multivariate ergodic mappings could be of use in order to produce longer periods out of shorter words operations: For instance, to obtain a period of length 2^{256} one may use either univariate ergodic functions (hence, 256-bit operands) or he may use 8-variate ergodic functions and work with 32-bit words. Multivariate ergodic mappings of [15] are conjugate to univariate ones (see [8]); hence *despite all further results are stated for a univariate case, they hold for these multivariate mappings as well.* Thus a designer could use further constructions either with longer words organized into 1-dimensional arrays, or with shorter words organized into arrays of bigger dimensions.

3. Constructions

In this section we introduce a method to construct counter dependent pseudorandom generators out of ergodic and measure-preserving mappings. The method guarantees that output sequences of these generators are always strictly uniformly distributed. Actually, all these constructions are wreath products of automata in the sense of Definition 1; the following results give us conditions these automata should satisfy to produce a uniformly distributed output sequence. Our main technical tool is the following

Theorem 2. *Let $\mathcal{G} = \{g_0, \ldots, g_{m-1}\}$ be a finite sequence of compatible measure preserving mappings of $\mathbb{Z}_2$ onto itself such that*

(1) *the sequence $\{(g_{i \bmod m}(0)) \bmod 2 \colon i = 0, 1, 2, \ldots\}$ is purely periodic, its shortest period is of length m;*

(2) $\sum_{i=0}^{m-1} g_i(0) \equiv 1 \pmod 2$*;*

(3) $\sum_{j=0}^{m-1} \sum_{z=0}^{2^k-1} g_j(z) \equiv 2^k \pmod{2^{k+1}}$ *for all $k = 1, 2, \ldots$.*

Then the recurrence sequence $\mathcal{Z}$ defined by the relation

$$x_{i+1} = g_{i \bmod m}(x_i)$$

is strictly uniformly distributed modulo 2^n for all $n = 1, 2, \ldots$: That is, modulo each 2^n the sequence $\mathcal{Z}$ is purely periodic, its shortest period is of length $2^n m$, and each element of $\mathbb{Z}/2^n$ occurs at the period exactly m times.

Note. In view of Theorem 1 condition (3) of Theorem 2 could be replaced by the equivalent condition

$$\sum_{j=0}^{m-1} \mathrm{Coef}_{0,\dots,k-1}(\varphi_k^j) \equiv 1 \pmod 2 \qquad (k = 1, 2, \dots),$$

where $\mathrm{Coef}_{0,\dots,k-1}(\varphi)$ is a coefficient of the monomial $\chi_0 \cdots \chi_{k-1}$ in the Boolean polynomial φ.

It turns out that the sequence $\mathcal{Z}$ of Theorem 2 is just the sequence $\mathcal{Y}$ of the following

Lemma 1. *Let $c_0, \dots, c_{m-1}$ be a finite sequence of 2-adic integers, and let $g_0, \dots, g_{m-1}$ be a finite sequence of compatible mappings of $\mathbb{Z}_2$ onto itself such that*

(i) $g_j(x) \equiv x + c_j \pmod 2$ *for* $j = 0, 1, \dots, m-1$,

(ii) $\sum_{j=0}^{m-1} c_j \equiv 1 \pmod 2$,

(iii) *the sequence* $\{c_{i \bmod m} \bmod 2 \colon i = 0, 1, 2, \dots\}$ *is purely periodic, its shortest period is of length* m,

(iv) $\delta_k(g_j(z)) \equiv \zeta_k + \varphi_k^j(\zeta_0, \dots, \zeta_{k-1}) \pmod 2$, $k = 1, 2, \dots$, *where* $\zeta_r = \delta_r(z)$, $r = 0, 1, 2, \dots$,

(v) *for each* $k = 1, 2, \dots$ *an odd number of Boolean polynomials* φ_k^j *in the Boolean variables* $\zeta_0, \dots, \zeta_{k-1}$ *are of odd weight.*

Then the recurrence sequence $\mathcal{Y} = \{x_i \in \mathbb{Z}_2\}$ defined by the relation $x_{i+1} = g_{i \bmod m}(x_i)$ is strictly uniformly distributed: It is purely periodic modulo 2^k for all $k = 1, 2, \dots$; its shortest period is of length $2^k m$; each element of $\mathbb{Z}/2^k$ occurs at the period exactly m times. Moreover,

(1) *the sequence $\mathcal{D}_s = \{\delta_s(x_i) \colon i = 0, 1, 2, \dots\}$ is purely periodic; it has a period of length $2^{s+1}m$,*

(2) $\delta_s(x_{i+2^s m}) \equiv \delta_s(x_i) + 1 \pmod 2$ *for all* $s = 0, 1, \dots, k-1$, $i = 0, 1, 2, \dots$,

(3) *for each* $t = 1, 2, \dots, k$ *and each* $r = 0, 1, 2, \dots$ *the sequence*

$$x_r \bmod 2^t, x_{r+m} \bmod 2^t, x_{r+2m} \bmod 2^t, \dots$$

is purely periodic, its shortest period is of length 2^t, each element of $\mathbb{Z}/2^t$ occurs at the period exactly once.

Note 1. Assuming $m = 1$ in Theorem 2 one obtains ergodicity criterion (Theorem 1).

Corollary 3. *Let a finite sequence of mappings $\{g_0, \dots, g_{m-1}\}$ of $\mathbb{Z}_2$ into itself satisfy conditions of Theorem 2, and let $\{F_0, \dots, F_{m-1}\}$ be an arbitrary finite sequence of balanced (and not necessarily compatible) mappings of $\mathbb{Z}/2^n$ ($n \geq 1$) onto $\mathbb{Z}/2^k$, $1 \leq k \leq n$. Then the sequence $\mathcal{F} = \{F_{i \bmod m}(x_i) \colon i = 0, 1, 2 \dots\}$, where $x_{i+1} = g_{i \bmod m}(x_i) \bmod 2^n$, is strictly uniformly distributed over $\mathbb{Z}/2^k$: It is purely periodic with a period of length $2^n m$, and each element of $\mathbb{Z}/2^k$ occurs at the period exactly $2^{n-k}m$ times.*

Theorem 2 and Lemma 1 together with Corollary 3 enables one to construct a counter-dependent generator out of the following components:

- A sequence $c_0, \dots, c_{m-1}$ of integers, which we call a *control sequence.*
- A sequence $h_0, \dots, h_{m-1}$ of compatible mappings, which is used to form a sequence of clock state update functions g_i (see, e.g., Example 1).
- A sequence $H_0, \dots, H_{m-1}$ of compatible mappings to produce clock output functions F_i (see, e.g., Proposition 7).

Note that ergodic functions that are needed to meet conditions of Proposition 7 or Example 1(3) could be produced out of compatible ones with the use of Propositions 3 or 4. A control sequence could be produced by an external generator (which in turn could be a generator of the kind considered in this paper), or it could be just a queue the state update and output functions are called from a look-up table. The functions h_i and/or H_i could be either precomputed to arrange that look-up table, or they could be produced on-the-fly in a form that is determined by a control sequence. This form may also look 'crazy', e.g.,

$$h_i(x) = (\cdots((u_0(\delta_0(c_i)) \bigcirc_{\delta_1(c_i),\delta_2(c_i)} u_1(\delta_3(c_i))) \bigcirc_{\delta_4(c_i),\delta_5(c_i)} u_2(\delta_6(c_i)))\cdots, \tag{4}$$

where $u_j(0) = x$, the variable, and $u_j(1)$ is a constant (which is determined by c_i, or is read from a precomputed look-up table, etc.), while (say) $\bigcirc_{0,0} = +$, an integer addition, $\bigcirc_{1,0} = \cdot$, an integer multiplication, $\bigcirc_{0,1} = \mathsf{XOR}$, $\bigcirc_{1,1} = \mathsf{AND}$. There is absolutely no matter what these h_i and H_i look like or how they are obtained, *the above stated results give a general method to combine all the data together to produce a uniformly distributed output sequence of a maximum period length.*

Example 1. These are obtained with the use of Lemma 1, Theorem 1, Proposition 4, and (7).

(1) A control sequence could be produced by the generator $\mathfrak{A} = \langle \mathbb{Z}/2^s, \mathbb{Z}/2^s, f, F, u_0\rangle$ (see Section 2) with ergodic state update function f and measure-preserving output function F. Then length of the shortest period of the control sequence is $m = 2^s$, see Proposition 2. Take m arbitrary ergodic functions $h_0, \dots, h_{m-1}$ and arbitrary odd $k \in \{0, 1, \dots, m-1\}$, and put $\breve{g}_0(x) = x \oplus (x+1) \oplus h_0(x), \dots, \breve{g}_{k-1} = x \oplus (x+1) \oplus h_{k-1}(x)$, $\breve{g}_k = h_k, \dots, \breve{g}_{m-1} = h_{m-1}$, $g_i = \breve{g}_{c_i \bmod m}$ for $i = 0, 1, 2, \dots$. In other words, in this case the control sequence just define the queue the functions $\breve{g}_j$ are called, thus producing the output sequence

$$x_0, x_1 = \breve{g}_{c_0}(x_0) \bmod 2^n, x_2 = \breve{g}_{c_1}(x_1) \bmod 2^n, \dots$$

Obviously, in this item a control sequence could be an arbitrary permutation of $0, 1, \dots, 2^s - 1$, and not necessarily an output of the generator $\mathfrak{A}$.

(2) Now let $\{c_0, \dots, c_{m-1}\}$ be an arbitrary sequence of length $m = 2^s$, i.e., $c_0, \dots, c_{m-1}$ are not necessarily pairwise distinct. Let $\{h_0, \dots, h_{m-1}\}$ be arbitrary compatible and ergodic mappings. For $0 \le j \le m-1$ put $g_j(x) = c_j + h_j(x)$. [11] These mappings g_j satisfy conditions of Theorem 2 if and only if $\sum_{j=0}^{2^m-1} c_j \equiv 1 \pmod 2$.

(3) For $m > 1$ odd let $\{h_0, \dots, h_{m-1}\}$ be a finite sequence of compatible and ergodic mappings; let $\{c_0, \dots, c_{m-1}\}$ be a finite sequence of integers such that

- $\sum_{j=0}^{m-1} c_j \equiv 0 \pmod 2$, and
- the sequence $\{c_{i \bmod m} \bmod 2 \colon i = 0, 1, 2, \dots\}$ is purely periodic with the shortest period of length m.

Put $g_j(x) = c_j \oplus h_j(x)$ (respectively, $g_j(x) = c_j + h_j(x)$). Then g_j satisfy conditions of Theorem 2.

(4) The conditions of (3) are satisfied in the case $m = 2^s - 1$ and $\{c_0, \dots, c_{m-1}\}$ is the output sequence of a maximum period linear feedback shift register over $\mathbb{Z}/2$ with s cells.

[11] one may also put $g_j(x) = (c_j + x) \oplus (2 \cdot h_j(x))$.

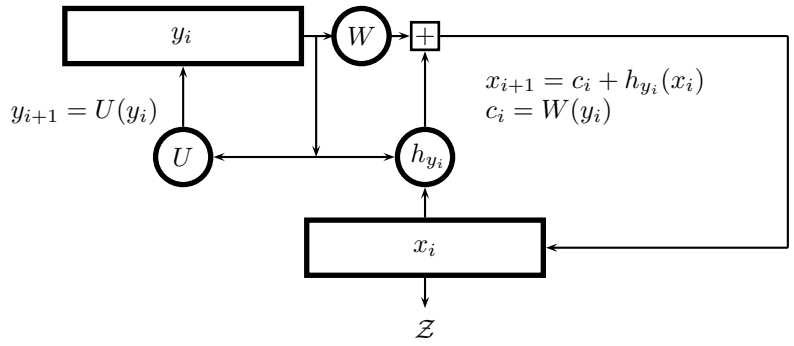

Figure 3. Wreath product basic circuit of Example 1, (2)–(4).

A basic circle illustrating this example wreath products is given at Figure 3. A number of counter dependent generators could be derived from Example 1 by taking explicit expressions for involved mappings. For instance, one can obtain the following result, which is a variation of theme of [16, Theorem 3]). Take odd $m > 1$ and consider a finite sequence $C_0, \dots, C_{m-1}$ of integers such that $\delta_0(C_j) = 1$ and $\delta_2(C_j) = 1$, $j = 0, 1, \dots, m-1$. Let a sequence $\{c_j \colon j = 0, 1, 2, \dots\}$ satisfy conditions of Example 1(3). Then the sequence $\{x_{i+1} = (x_i + c_i + (x_i^2 \mathbin{\mathsf{OR}} C_i)) \bmod 2^n \colon i = 0, 1, 2, \dots\}$ is purely periodic modulo 2^k for all $k = 1, 2, \dots$ with the shortest period of length $2^k m$, and each element of $\mathbb{Z}/2^k$ occurs at the period exactly m times. This is a stronger claim in comparison with that of [16, Theorem 3]): Not only the sequence of pairs (y_i, x_i) defined by $y_{i+1} = (y_i + 1) \bmod m$; $x_{i+1} = (x_i + c_i + (x_i^2 \mathbin{\mathsf{OR}} C_{y_i})) \bmod 2^n$ is periodic with a period of length $2^n m$, yet length of the shortest period of the sequence $\{x_i\}$ is $2^n m$. The latter could never be achieved under conditions of Theorem 3 of [16]: They imply that the length of the shortest period of the sequence $\{x_i \pmod 2\}$ is 2, and not $2m$.

4. Properties of output sequences

Distribution of k-tuples

The output sequence $\mathcal{Z}$ of any wreath product of automata that satisfy Theorem 2 is strictly uniformly distributed as a sequence over $\mathbb{Z}/2^n$ for all n. That is, each sequence $\mathcal{Z}_n$ of residues modulo 2^n of members of the sequence $\mathcal{Z}$ is purely periodic, and each element of $\mathbb{Z}/2^n$ occurs at the period the same number of times. However, when this sequence $\mathcal{Z}_n$ is used as a key-stream, that is, as a binary sequence $\mathcal{Z}'_n$ obtained by

a concatenation of successive n-bit words of $\mathcal{Z}$, it is important to know how n-tuples are distributed in this binary sequence. Yet strict uniform distribution of an arbitrary sequence $\mathcal{T}$ as a sequence over $\mathbb{Z}/2^n$ does not necessarily imply uniform distribution of n-tuples, if this sequence is considered as a binary sequence $\mathcal{T}'$.

For instance, let $\mathcal{T} = 023102310231\ldots$. This sequence is strictly uniformly distributed over $\mathbb{Z}/4$; the length of its shortest period is 4. Its binary representation is $\mathcal{T}'_2 = 000111100001111000011110\ldots$ Considering $\mathcal{T}$ as a sequence over $\mathbb{Z}/4$, each number of $\{0,1,2,3\}$ occurs in the sequence with the same frequency $\frac{1}{4}$. Yet if we consider $\mathcal{T}$ in its binary form $\mathcal{T}'_2$, then 00 (as well as 11) occurs in this sequence with frequency $\frac{3}{8}$, whereas 01 (as well as 10) occurs with frequency $\frac{1}{8}$.

In this subsection we show that such an effect does not take place for output sequences of automata described in Theorem 2, Lemma 1, and Example 1: *Considering any of these sequences in a binary form, a distribution of k-tuples is uniform, for all $k \le n$.* Now we state this property formally.

Consider a (binary) *n-cycle* $C = (\varepsilon_0\varepsilon_1\ldots\varepsilon_{n-1})$, i.e., an oriented graph on vertices $\{a_0, a_1, \ldots, a_{n-1}\}$ and edges

$$\{(a_0,a_1),(a_1,a_2),\ldots,(a_{n-2},a_{n-1}),(a_{n-1},a_0)\},$$

where each vertex a_j is labelled with $\varepsilon_j \in \{0,1\}$, $j = 0,1,\ldots,n-1$. (Note that then $(\varepsilon_0\varepsilon_1\ldots\varepsilon_{n-1}) = (\varepsilon_{n-1}\varepsilon_0\ldots\varepsilon_{n-2}) = \ldots$, etc.). Clearly, each purely periodic sequence $\mathcal{S}$ over $\mathbb{Z}/2$ with period $\alpha_0\ldots\alpha_{n-1}$ of length n could be related to a binary n-cycle $C(\mathcal{S}) = (\alpha_0\ldots\alpha_{n-1})$. Conversely, to each binary n-cycle $(\alpha_0\ldots\alpha_{n-1})$ we could relate n purely periodic binary sequences with periods of length n: Those are n shifted versions of the sequence

$$\alpha_0\ldots\alpha_{n-1}\alpha_0\ldots\alpha_{n-1}\ldots.$$

Further, *a k-chain in a binary n-cycle C* is a binary string $\beta_0\ldots\beta_{k-1}$, $k < n$, that satisfies the following condition: There exists $j \in \{0,1,\ldots,n-1\}$ such that $\beta_i = \varepsilon_{(i+j) \bmod n}$ for $i = 0,1,\ldots,k-1$. Thus, a k-chain is just a string of length k of labels that corresponds to a chain of length k in a graph C. We call a binary n-cycle C *k-full*, if each k-chain occurs in the graph C the same number $r > 0$ of times.

Clearly, if C is k-full, then $n = 2^k r$. For instance, a well-known De Bruijn sequence is an n-full 2^n-cycle. Clearly enough that a k-full n-cycle is $(k-1)$-full: Each $(k-1)$-chain occurs in C exactly $2r$ times, etc. Thus, if an n-cycle $C(\mathcal{S})$ is k-full, then each m-tuple (where $1 \le m \le k$) occurs in the sequence $\mathcal{S}$ with the same probability (limit frequency) $\frac{1}{2^m}$. That is, the sequence $\mathcal{S}$ is *k-distributed*, see [18, Section 3.5, Definition D].

Definition 4. A purely periodic binary sequence $\mathcal{S}$ with the shortest period of length N is said to be *strictly k-distributed* iff the corresponding N-cycle $C(\mathcal{S})$ is k-full.

Thus, if a sequence $\mathcal{S}$ is strictly k-distributed, then it is strictly s-distributed, for all positive $s \le k$.

Theorem 3. *For the sequence $\mathcal{Z}$ of Theorem 2 each binary sequence $\mathcal{Z}'_n$ is strictly k-distributed for all $k = 1,2,\ldots,n$.*

Note 2. Theorem 3 remains true for the sequence $\mathcal{F}$ of Corollary 3, where $F_j(x) = \left\lfloor \frac{x}{2^{n-k}} \right\rfloor \bmod 2^k$, $j = 0,1,\ldots,m-1$, a truncation of $(n-k)$ less significant bits. Namely, *a binary representation $\mathcal{F}'_n$ of the sequence $\mathcal{F}$ is a purely periodic strictly k-distributed binary sequence with a period of length $2^n mk$.*

Theorem 3 treats an output sequence of a counter-dependent automaton as an infinite (though, a periodic) binary sequence. However, in cryptography only a part of a period is used during encryption. So it is natural to ask how 'random' is a finite segment (namely, the period) of this infinite sequence. According to [18, Section 3.5, Definition Q1] a finite binary sequence $\varepsilon_0\varepsilon_1\ldots\varepsilon_{N-1}$ of length N is said to be random, iff

$$\left|\frac{\nu(\beta_0\ldots\beta_{k-1})}{N} - \frac{1}{2^k}\right| \le \frac{1}{\sqrt{N}} \tag{5}$$

for all $0 < k \le \log_2 N$, where $\nu(\beta_0\ldots\beta_{k-1})$ is the number of occurrences of a binary word $\beta_0\ldots\beta_{k-1}$ in a binary word $\varepsilon_0\varepsilon_1\ldots\varepsilon_{N-1}$. If a finite sequence is random in the sense of this Definition Q1 of [18], we shall say that this sequence *satisfies* Q1. We shall also say that an *infinite periodic sequence satisfy* Q1 iff its shortest period satisfies Q1. Note that, contrasting to the case of strict k-distribution, which implies strict $(k-1)$-distribution, it is not enough to demonstrate only that (5) holds for $k = \lfloor \log_2 N \rfloor$ to prove a finite sequence of length N satisfies Q1: For instance, the sequence 1111111100000111 satisfies (5) for $k = \lfloor \log_2 N \rfloor = 4$ and does not satisfy (5) for $k = 3$.

Corollary 4. *The sequence $\mathcal{Z}'_n$ of Theorem 3 satisfies Q1 if $m \le \frac{2^n}{n}$. Moreover, in this case under the conditions of Note 2 the output binary sequence still satisfies Q1 if one truncates $0 \le k \le \frac{n}{2} - \log_2 \frac{n}{2}$ lower order bits (that is, if one uses clock output functions F_j of Note 2).*

We note here that according to Corollary 4 a control sequence of a counter-dependent automaton (see Theorem 2, Lemma 1, Corollary 3, and the text and examples thereafter) may not satisfy Q1 at all, yet nevertheless a corresponding output sequence necessarily satisfies Q1. Thus, *with the use of wreath product techniques one could stretch 'non-randomly looking' sequences to 'randomly looking' ones.*

Structure

A recurrence sequence could be 'very uniformly distributed', yet nevertheless could have some mathematical structure that might be used by an attacker to break the cipher. For instance, a clock sequence $x_i = i$ is uniformly distributed in $\mathbb{Z}_2$; moreover, its counterpart in the field $\mathbb{R}$ of real numbers, the so-called Van der Corput sequence $u_i = i \cdot 2^{-\lfloor \log_2 i \rfloor - 1}$, has the least (of the known) discrepancy, see [20]. We are going to study what structure could have sequences outputted by our counter-dependent generators.

Theorem 2 immediately implies that the j^{th} *coordinate sequence* $\delta_j(\mathcal{Z}) = \{\delta_j(x_i)\colon i = 0, 1, 2, \ldots\}$ $(j = 0, 1, 2, \ldots)$ *of the sequence* $\mathcal{Z}$, i.e., a sequence formed by all j^{th} bits of members of the sequence $\mathcal{Z}$, has a period not longer than $m \cdot 2^{j+1}$. Moreover, the following could be easily proved:

Proposition 5.

(1) *The j^{th} coordinate sequence $\delta_j(\mathcal{Z})$ is a purely periodic binary sequence with a period of length $2^{j+1}m$, and*

(2) *the second half of the period is a bitwise negation of the first half:* $\delta_j(x_{i+2^j m}) \equiv \delta_j(x_i) + 1 \pmod 2$, $i = 0, 1, 2, \ldots$

This means that the j^{th} coordinate sequence of the sequence of states of a counter-dependent generator is completely determined by the first half of its period; so, intuitively, it is as 'complex' as the first half of its period. Thus we ought to understand what sequences of length $2^j m$ occur as the first half of the period of the j^{th} coordinate sequence.

For $j = 0$ (and $m > 1$) the answer immediately follows from Theorem 2 and Lemma 1 — any binary sequence $c_0, \ldots, c_{m-1}$ such that $\sum_{j=0}^{m-1} c_j \equiv 1 \pmod 2$ does. It turns out that for $j > 0$ *any binary sequence could be produced as the first half of the period of the j^{th} coordinate sequence independently of other coordinate sequences.*

More formally, to each sequence $\mathcal{Z}$ described by Theorem 2 we associate a sequence $\Gamma(\mathcal{Z}) = \{\gamma_1, \gamma_2, \ldots\}$ of non-negative rational integers γ_j such that $0 \le \gamma_j \le 2^{2^j m} - 1$ and the base-2 expansion of γ_j agrees with the first half of the period of the j^{th} coordinate sequence $\delta_j(\mathcal{Z})$ for all $j = 1, 2, \ldots$; that is

$$\gamma_j = \delta_j(x_0) + 2 \cdot \delta_j(x_1) + 4 \cdot \delta_j(x_2) + \cdots + 2^{2^j m - 1} \cdot \delta_j(x_{2^j m - 1}),$$

where x_0 is an initial state; $x_{i+1} = g_{i \bmod m}(x_i)$, $i = 0, 1, 2, \ldots$. Now we take an arbitrary sequence $\Gamma(\mathcal{Z}) = \{\gamma_1, \gamma_2, \ldots\}$ of non-negative rational integers γ_j such that $0 \le \gamma_j \le 2^{2^j m} - 1$ and wonder whether this sequence could be so associated to some sequence $\mathcal{Z}$ described by Theorem 2.

The answer is *yes*. Namely, the following theorem holds.

Theorem 4. *Let $m > 1$ be a rational integer, and let $\Gamma = \{\gamma_1, \gamma_2, \ldots\}$ be an arbitrary sequence over $\mathbb{N}_0$ such that $\gamma_j \in \{1, 2, \ldots, 2^{2^j m} - 1\}$ for all $j = 1, 2, \ldots$. Then there exist a finite sequence $\mathcal{G} = \{g_0, \ldots, g_{m-1}\}$ of compatible measure preserving mappings of $\mathbb{Z}_2$ onto itself and a 2-adic integer $x_0 = z \in \mathbb{Z}_2$ such that $\mathcal{G}$ satisfies conditions of Theorem 2, and the base-2 expansion of γ_j agrees with the first $2^j m$ terms of the sequence $\delta_j(\mathcal{Z})$ for all $j = 1, 2, \ldots$, where the recurrence sequence $\mathcal{Z} = \{x_0, x_1, \ldots \in \mathbb{Z}_2\}$ is defined by the recurrence relation $x_{i+1} = g_{i \bmod m}(x_i)$, $(i = 0, 1, 2, \ldots)$. In the case $m = 1$ the assertion holds for an arbitrary $\Gamma = \{\gamma_0, \gamma_1, \ldots\}$, where $\gamma_j \in \{1, 2, \ldots, 2^{2^j} - 1\}$, $j = 0, 1, 2, \ldots$.*

Linear complexity

The latter is an important cryptographic measure of complexity of a binary sequence; being a number of cells of the shortest linear feedback shift register (LFSR) that outputs the given sequence[12] it estimates dimensions of a linear system an attacker must solve to obtain initial state.

Theorem 5. *For $\mathcal{Z}$ and m of Theorem 2 let $\mathcal{Z}_j = \delta_j(\mathcal{Z})$, $j > 0$, be the j^{th} coordinate sequence. Represent $m = 2^k r$, where r is odd. Then length of the shortest period of $\mathcal{Z}_j$ is $2^{k+j+1}s$ for some $s \in \{1, 2, \ldots, r\}$, and both extreme cases $s = 1$ and $s = r$ occur: For every sequence $s_1, s_2, \ldots$ over a set $\{1, r\}$ there exists a sequence $\mathcal{Z}$ of Theorem 2 such that length of the shortest period of $\mathcal{Z}_j$ is $2^{k+j+1}s_j$, $(j = 1, 2, \ldots)$. Moreover, linear complexity $\Psi_2(\mathcal{Z}_j)$ of the sequence $\mathcal{Z}_j$ satisfies the following inequality:*

$$2^{k+j} + 1 \le \Psi_2(\mathcal{Z}_j) \le 2^{k+j} r + 1.$$

[12] i.e., degree of the minimal polynomial over $Z/2$ of the given sequence

Both these bounds are sharp: For every sequence $t_1, t_2, \ldots$ over a set $\{1, r\}$ there exists a sequence $\mathcal{Z}$ of Theorem 2 such that linear complexity of $\mathcal{Z}_j$ is exactly $2^{k+j}t_j + 1$, $(j = 1, 2, \ldots)$.

Note. Somewhat similar estimates hold for 2-adic span (see definition in [14]), one more cryptographic measure of complexity of a sequence. We have to omit exact statements due to space limitations.

Whereas the linear complexity of a binary sequence $\mathcal{X}$ is the length of the shortest LFSR that produces $\mathcal{X}$, the *ℓ-error linear complexity* is the length of the shortest LFSR that produces a sequence with almost the same (with the exception of not more than ℓ members) period as that of $\mathcal{X}$; that is, the two periods coincide everywhere but at $t \le \ell$ places. Obviously, a random sequence of length L coincides with a sequence that has a period of length L approximately at $\frac{L}{2}$ places. That is, the ℓ-error linear complexity makes sense only for $\ell < \frac{L}{2}$. The following proposition holds.

Proposition 6. *Let $\mathcal{Z}$ be a sequence of Theorem 2, and let $m = 2^s > 1$. Then for ℓ less than the half of the length of the shortest period of the j-th coordinate sequence $\delta_j(\mathcal{Z})$, the ℓ-error linear complexity of $\delta_j(\mathcal{Z})$ exceeds 2^{j+m-1}, the half of the length of its shortest period.*

From Theorem 5 it follows that the less is j, the shorter is a period (and the smaller is linear complexity) of the coordinate sequence $\mathcal{Z}_j$. This could be improved by truncation of less significant bits (see Corollary 4) or, if necessary, with the use of clock output functions of special kind:

Proposition 7. *Let $H_i\colon \mathbb{Z}_2 \to \mathbb{Z}_2$ $(i = 0, 1, 2, \ldots, m-1)$ be compatible and ergodic mappings. For $x \in \{0, 1, \ldots, 2^n - 1\}$ let $F_i(x) = (H_i(\pi(x))) \bmod 2^n$, where π is a permutation of bits of $x \in \mathbb{Z}/2^n$ such that $\delta_0(\pi(x)) = \delta_{n-1}(x)$. Consider a sequence $\mathcal{F}$ of Corollary 3. Then the shortest period of the j^{th} coordinate sequence $\mathcal{F}_j = \delta_j(\mathcal{F})$ $(j = 0, 1, 2, \ldots, n-1)$ is of length $2^n k_j$ for a suitable $1 \le k_j \le m$. Moreover, linear complexity of the sequence $\mathcal{F}_j$ exceeds 2^{n-1}.*

Note. In view of Note 1, all the results of Section 4. remain true for compatible mappings $T\colon \mathbb{Z}_2 \to \mathbb{Z}_2$ (i.e., for T-functions) either.

5. Security issues

The paper introduces design techniques that guarantees in advance that the so constructed generator, which dynamically modifies itself during encryption, will meet certain important cryptographic properties; namely, long period, uniform distribution and high linear complexity of the output sequence. The techniques can not guarantee per se that every such cipher will be secure — obvious degenerative cases exist. On the other hand, if clock state update functions g_i are chosen arbitrarily under the conditions of Theorem 2, and clock output functions F_i just truncate k low order bits, $k \approx \frac{n}{2}$ (see Corollary 4), Theorem 4 leaves no chance to an attacker to break such a scheme. Yet in practice we can not choose g_i arbitrarily; restrictions are determined by concrete implementations, which are not discussed here.

In this section we are going to give some evidence that with the use of the techniques described above it might be possible to design stream ciphers such that the problem of their key recovery is intractable up to the following conjecture: Choose (randomly and independently) $k \le n$ ANF's ψ_i in n Boolean variables $\chi_0, \ldots, \chi_{n-1}$ from the class of ANF's with polynomially restricted number of monomials. Consider a mapping $F\colon \mathbb{Z}/2^n \to \mathbb{Z}/2^k$:

$$F(x) = F(\chi_0, \ldots, \chi_{n-1}) = \psi_0(\chi_0, \ldots, \chi_{n-1}) \\ \oplus \psi_1(\chi_0, \ldots, \chi_{n-1}) \cdot 2 \oplus \cdots \oplus \psi_{k-1}(\chi_0, \ldots, \chi_{n-1}) \cdot 2^{k-1},$$

where $\chi_j = \delta_j(x)$ for $x \in \mathbb{Z}/2^n$. We conjecture that *this function F is one-way*, that is, one could invert it (i.e., could find an F-preimage in case it exists) only with a negligible in n probability. Note that to find any F-preimage, i.e., to solve an equation $F(x) = y$ in unknown x one has to solve a system of k Boolean equations in n variables. Yet *to determine whether k ANF have common zero is an NP-complete problem*, see, e.g., [13, Appendix A, Section A7.2, Problem ANT-9].

Of course, it is not sufficient to conjecture F is one-way in case we only know that the problem of whether F-preimage exists is NP-complete; it must be hard in average to invert F. However, to our best knowledge, no polynomial-time algorithms that solve random systems of k Boolean equations in n variables for so restricted k are known. The best known results are polynomial-time algorithms that solve so-called overdefined Boolean systems of degree not more than 2, i.e., systems where the number of equations is greater than the number of unknowns and where each ANF is at most quadratic, see [11], [12].

Proceeding with the above plausible conjecture, to each ANF ψ_i, $i = 0, 1, 2, \ldots, k-1$ we relate a mapping $\Psi_i\colon \mathbb{Z}_2 \to \mathbb{Z}_2$ in the following way: $\Psi_i(x) = \psi_i(\delta_0(x), \ldots, \delta_{n-1}(x)) \in \{0, 1\} \subset \mathbb{Z}_2$. Now to each above

mapping F we relate a mapping

$$f_F(x) = (1+x) \oplus 2^{n+1} \cdot F(x)$$
$$= (1+x) \oplus 2^{n+1} \cdot \Psi_0(x) \oplus 2^{n+2} \cdot \Psi_1(x) \oplus \cdots \oplus 2^{n+k} \cdot \Psi_{k-1}(x)$$

of $\mathbb{Z}_2$ onto itself. Clearly,

$$\delta_j(f_F(x)) = \begin{cases} 1 \oplus \delta_0(x), & \text{if } j = 0; \\ \delta_j(x) \oplus \delta_0(x) \cdots \delta_{j-1}(x), & \text{if } 0 < j \le n; \\ \delta_j(x) \oplus \delta_0(x) \cdots \delta_{j-1}(x) & \\ \quad \oplus \psi_{j-n-1}(\delta_0(x), \ldots, \delta_{n-1}(x)), & \text{if } n+1 \le j \le n+k. \end{cases}$$

In view of Theorem 1 the mapping $f_F \colon \mathbb{Z}_2 \to \mathbb{Z}_2$ is compatible and ergodic for any choice of ANF's $\psi_0, \ldots, \psi_{k-1}$.

Now for $m = 2^n$ and $i = 0, 1, 2, \ldots, m-1$ choose arbitrarily and independently mappings $F_i \colon \mathbb{Z}/2^n \to \mathbb{Z}/2^k$ of the above kind. Put $d_0 = \ldots = d_{2^n-3} = 0$, $d_{2^n-2} = d_{2^n-1} = 1$, and consider a recurrence sequence of states $x_{i+1} = d_{i \bmod m} \oplus f_{F_i \bmod m}(x_i)$ and a corresponding output sequence $g(x_0), g(x_1), \ldots$ over $\mathbb{Z}/2^k$, where $g(x) = \lfloor \frac{x}{2^{n+1}} \rfloor \bmod 2^k$, a truncation. In view of Example 1 the output sequence satisfy Corollary 3.

We shall always take a key $z \in \{0, 1, \ldots, 2^n - 1\}$ as an initial state x_0. Let z be the only information that is not known to an attacker, let everything else, i.e., n, k, f_{F_i}, d_i, and g, as well as the first s members of the output sequence $\{y_i\}$, be known to him. Since $\delta_0(x) \cdots \delta_{j-1}(x) = 1$ iff $x \equiv -1 \pmod{2^j}$, with probability $1 - \epsilon$ (where ϵ is negligible if s is a polynomial in n) he obtains a sequence[13]:

$$y_0 = F_0(z), y_0 \oplus y_1 = F_1(z+1), \ldots, y_{s-2} \oplus y_{s-1} = F_{s-1}(z+s-1) \quad (6)$$

To find z the attacker may try to solve any of these equations; he could do it with a negligible advantage, since F_i is one-way. Of course, the attacker may try to express $z+i$ as a collection of ANF's $\delta_0(z+i), \ldots, \delta_{n-1}(z+i)$ in the variables $\chi_0 = \delta_0(z), \ldots, \chi_{n-1} = \delta_{n-1}(z)$, then substitute these ANF's for the variables into the ANF's that define mappings F_i, to obtain an overdefined system (6) in unknowns $\chi_0, \ldots, \chi_{n-1}$. However, the known formula (see, e.g., [1] and fix an obvious misprint there)

$$\delta_j(z+i) \equiv \chi_j + \delta_j(i) + \sum_{r=0}^{j-1} \delta_r(i) \cdot \chi_r \prod_{t=r+1}^{j-1} (\delta_t(i) + \chi_t) \pmod 2; \quad (7)$$

[13] which is pseudorandom even if $F = F_0 = F_1 = \ldots$, under additional conjecture (how plausible is it?) that the function F constructed above is a pseudorandom function

implies that the number of monomials in the equations of the obtained system will be, generally speaking, exponential in n; to say nothing of that the number of operations to make these substitutions and then to collect similar terms is also exponential in n, unless the degree of all ANF's that define all F_i is bounded by a constant (the latter is not a case according to our assumptions).

Finally, our assumption that the attacker knows all F_i seems to be too strong: It is more practical to assume that he does not know F_i in (6), since given clock output (and/or clock state update) functions as explicit compositions of arithmetical and bitwise logical operators, 'normally' it is infeasible to express these functions in the Boolean form (Proposition 1): Corresponding ANF's 'as a rule' are sums of exponential in n number of monomials, cf. (7). Moreover, if these clock output functions F_i and/or clock state update functions f_i are determined by a key-dependent control sequence (say, which is produced by a generator with unknown initial state), see Section 3, then the explicit forms of the mentioned compositions are also unknown. So in general an attacker has to find an initial state u_0 having only a segment $z_j, z_{j+1}, \ldots$ of the output sequence formed according to the rule (1), where both f_i and F_i are not known to him. An 'algebraic' way to do this by guessing f_i and F_i and solving corresponding systems of equations seems to be hopeless in view of Corollary 2 and the above discussion. The results of preceding sections[14] give us reasons to conjecture that under common tests the sequence $z_j, z_{j+1}, \ldots$ behaves like a random one, so 'statistical' methods of breaking such (reasonably designed) ciphers seem to be ineffective as well.

References

[1] R. C. Alperin. p-adic binomial coefficients mod$p$. *The Amer. Math. Month.*, 92(8):576–578, 1985.

[2] V. Anashin, A. Bogdanov, I. Kizhvatov. ABC: A new fast flexible stream cipher, version 2. Available from `http://crypto.rsuh.ru/papers/abc-spec-v2.pdf`, 2005.

[3] V. S. Anashin. Uniformly distributed sequences of p-adic integers. *Mathematical Notes*, 55(2):109–133, 1994.

[4] V. S. Anashin. Uniformly distributed sequences in computer algebra, or how to constuct program generators of random numbers. *J. Math. Sci.*, 89(4):1355–1390, 1998.

[14] as well as computer experiments: Output sequences of explicit generators of the kind considered in the paper passed both DIEHARD and NIST test suites

[5] V. S. Anashin. Uniformly distributed sequences of p-adic integers, II. *Discrete Math. Appl.*, 12(6):527–590, 2002. A preprint available from `http://arXiv.org/math.NT/0209407`.

[6] V. S. Anashin. On finite pseudorandom sequences. In *Kolmogorov and contemporary mathematics*, p. 382–383, Moscow, June 2003. Russian Academy of Sciences, Moscow State University. Abstracts of the International Conference.

[7] V. S. Anashin. Pseudorandom number generation by p-adic ergodic transformations. Available from `http://arxiv.org/abs/cs.CR/0401030`, January 2004.

[8] V. S. Anashin. Pseudorandom number generation by p-adic ergodic transformations: an addendum. `http://arxiv.org/abs/cs.CR/0402060`, February 2004.

[9] Vladimir Anashin. Uniformly distributed sequences over p-adic integers. In I. Shparlinsky A. J. van der Poorten and H. G. Zimmer, editors, *Number theoretic and algebraic methods in computer science. Proceedings of the International Conference (Moscow, June–July, 1993)*, p. 1–18. World Scientific, 1995.

[10] Vladimir Anashin, Andrey Bogdanov, Ilya Kizhvatov. Increasing the ABC stream cipher period. Technical report, ECRYPT, July 2005. `http://www.ecrypt.eu.org/stream/papersdir/050.pdf`.

[11] M. Bardet, J.-C. Faugère, B. Salvy. Complexity of Gröbner basis computation for semi-regular overdetermined sequences over $\mathbb{F}_2$ with solutions in $\mathbb{F}_2$. Available from `http://www.inria.fr/rrrt/rr-5049.html`, 2004.

[12] N. Courtois, A. Klimov, J. Patarin, A. Shamir. Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In *Eurocrypt 2000*, v. 1807 of *Lect. Notes Comp. Sci.*, p. 392–407. Springer-Verlag, 2000.

[13] M. R. Garey, D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman and Co., 1979.

[14] A. Klapper, M. Goresky. Feedback shift registers, 2-adic span, and combiners with memory. *J. Cryptology*, 10:111–147, 1997.

[15] A. Klimov, A. Shamir. New cryptographic primitives based on multiword T-functions. In Bimal Roy and Willi Meier, editors, *Fast Software Encryption: 11th International Workshop, FSE 2004, Delhi, India, February 5–7, 2004. Revised Papers*, p. 1–15. Springer-Verlag GmbH, 2004.

[16] A. Klimov, A. Shamir. Cryptographic applications of T-functions. In *Selected Areas in Cryptography-2003*, 2003.

[17] A. Klimov, A. Shamir. A new class of invertible mappings. In B. S. Kaliski Jr. et al., editors, *Cryptographic Hardware and Embedded Systems 2002*, v. 2523 of *Lect. Notes in Comp. Sci.*, p. 470–483. Springer-Verlag, 2003.

[18] D. Knuth. *The Art of Computer Programming*, volume 2. Addison-Wesley, third edition, 1998.

[19] L. Kotomina. Fast nonlinear congruential generators. Diploma Thesis, Russian State University for the Humanities, Moscow, 1999. (in Russian).

[20] L. Kuipers, H. Niederreiter. *Uniform Distribution of Sequences*. John Wiley & Sons, N.Y. etc., 1974.

[21] M. V. Larin. Transitive polynomial transformations of residue class rings. *Discrete Mathematics and Applications*, 12(2):141–154, 2002.

[22] Hans Lausch, Wilfried Nöbauer. *Algebra of Polynomials*. North-Holl. Publ. Co, American Elsevier Publ. Co, 1973.

[23] A. Shamir, B. Tsaban. Guaranteeing the diversity of number generators. *Information and Computation*, 171:350–363, 2001. Available from `http://arXiv.org/abs/cs.CR/0112014`.

[24] S. V. Yablonsky. Basic notions of cybernetics. In *Problems of Cybernetics*. Fizmatgiz, 1959. (in Russian).

A Perfect Multi-Secret Sharing Scheme Based on Linear Transformations over Finite Commutative Chain Ring

A. N. Alekseychuk, L. V. Skrypnik, A. L. Voloshin

1. A multi-secret sharing scheme (multi-SSS) [1, 2] is a cryptographical protocol to share simultaneously several secret keys (secrets) among a set of participants in such a way that only predefined subsets (coalitions) of participants, pooling together their components, can reconstruct some predefined keys. If the participants of each coalition in a multi-SSS can obtain no posterior information on remaining keys (to which, according to the protocol, they may not access), then the multi-SSS is called perfect [2].

A trivial way to construct a multi-SSS consists of independent constructing of several simple secret sharing schemes (SSS) (see, for example, [3], chapter 5), each of these SSS can be separately used for sharing of its "own" secret key. Usually, this solution has bad practical characteristics because the participants of such multi-SSS should store a lot of secret information.

Some different ways to construct SSS based on linear transformations (or codes) over finite fields or vector spaces were studied in [3, 4, 5, 6] and some other works. In [7] the "vector construction" of a (in general case) non-perfect SSS over a Galois ring is proposed and partial description of the access hierarchy [8] of such secret sharing scheme is obtained.

In this paper we propose a construction of perfect multi-SSS differed from multi-secret sharing schemes introduced in [1, 2]. The proposed construction is a direct generalization of vector secret sharing schemes over finite fields [4] and Galois rings [7]. Next, we give a description of access hierarchies of proposed multi-SSS and obtain necessary and sufficient conditions, under which for arbitrary predefined access hierarchy such multi-SSS exists. At last, we propose an algorithm for constructing multi-SSS for a given access hierarchy. This algorithm generalizes a well-known algorithm for constructing vector SSS (for a predefined access structure) over a finite field [6].

2. Let R be a finite commutative local principal ideal ring (commutative chain ring) with the radical J and the residue field $\overline{R} = R/J = \mathrm{GF}(q)$ [9]. Denote by d the nilpotent index of radical J. Let's fix an element $a \in J \setminus J^2$ and arbitrary map $\gamma\colon \overline{R} \to R$ such that $\gamma(r+J) \equiv r \pmod J$ for any $r \in R$. According to [9], ideals of R will form a chain: $R \supset J \supset J^2 \supset \ldots \supset J^d = 0$. In this case the equalities $J^k = Ra^k$, $|J^k| = q^{d-k}$, $k \in \overline{0,d}$ are hold and each element $r \in R$ can be uniquely represented at the form

$$r = r[0] + r[1]a + \ldots + r[d-1]a^{d-1}, \tag{1}$$

where $r[j] \in \gamma(\overline{R})$, $j \in \overline{0,d-1}$.

Let us give a formal definition of multi-SSS that will be proposed.

Let $S = \overline{R}^d$ be a set of all d-tuples of secret keys (each of them is an element from the field $\overline{R}$), which should be shared between the participants from the set $P = \{1, 2, \ldots, n\}$, $n \geq 2$. Let's fix a $k \times (n+1)$ matrix

$$G = \left(\begin{array}{c|c} 1 & \\ 0 & \\ \vdots & G' \\ 0 & \end{array}\right), \tag{2}$$

over the ring R, where $k \geq 2$. Let $0, 1, \ldots, n$ be column numbers of G (from left to right).

Now we define a multi-SSS $\sigma(G)$ based on the matrix (2) as follows.

Let $(s_j : j \in \overline{0,d-1})$ be an arbitrary element from S. In order to share the tuple s_j, $j \in \overline{0,d-1}$ the dealer of multi-SSS $\sigma(G)$

(a) calculates the element

$$s = \sum_{j=0}^{d-1} \gamma(s_j)a^j \in R; \tag{3}$$

(b) chooses independently, randomly, and equiprobable elements $a_1, \ldots, a_{k-1} \in R$ and calculates the vector $(s, \pi_1, \ldots, \pi_n) = (s, a_1, \ldots, a_{k-1})G$. Last n coordinates of this vector are shares of the keys tuple $(s_j : j \in \overline{0,d-1})$. The element $\pi_i \in R$ is distributed to the i-th participant, $i \in \overline{1,n}$.

For any $j \in \overline{0,d}$ denote by $\widetilde{\Sigma}_j$ the collection of all sets A of participants, which can uniquely reconstruct from its shares the secret keys

$s_0, s_1, \ldots, s_{d-j-1}$ and only them. Let's note that

$$\widetilde{\Sigma}_{j_1} \cap \widetilde{\Sigma}_{j_2} = \emptyset, j_1, j_2 \in \overline{0,d}, j_1 \neq j_2, \bigcup_{j=0}^{d} \widetilde{\Sigma}_j = 2^P,$$

where 2^P denote the set of all subsets of P. According to terminology in [8], the set family $\widetilde{\Sigma} = \{\widetilde{\Sigma}_j : j \in \overline{0,d}\}$ is called an access hierarchy for the multi-SSS $\sigma(G)$.

As the following Theorem states, the participants in arbitrary coalition $A \in \widetilde{\Sigma}_j$ $(j \in \overline{0,d})$ can obtain no posterior information about the secret keys s_l for $l \in \overline{d-j, d-1}$. So, the multi-SSS $\sigma(G)$ is a (combinatorial) perfect multi-secret sharing scheme.

Theorem 1. *Let M be a R-module generated by rows of the matrix (2). For any $U \subseteq P_0 \stackrel{\text{def}}{=} P \cup 0$ let's denote by $\|M_U\|$ the number of all different vectors that occurs in columns with numbers from the set U of a $|M| \times (n+1)$ array with rows from the module M. Suppose $j \in \overline{0,d}$ and $\widetilde{\Sigma}_j \neq \emptyset$. Then for any $A \in \widetilde{\Sigma}_j$ the equality $\|M_{A\cup 0}\| = q^j \|M_A\|$ is hold.*

The described multi-SSS $\sigma(G)$ is called a linear multi-secret sharing scheme over ring R. Note that in particular case $d = 1$ this multi-SSS is a well-known vector SSS over the field GF(q) [4]. If R is a Galois ring, then the step (b) of proposed algorithm is similar to the share calculation procedure of (one) secret key $s \in R$ in a non-perfect SSS described in [7].

Let us denote by G_A the sub-matrix contained in columns of G with numbers from the set $A \subseteq P$. Let's denote by $G_i^{\downarrow}$ the i-th column of the matrix G, $i \in \overline{0,n}$; in particular, $G_0^{\downarrow} = (1, 0, \ldots, 0)^T \in R^{(k)}$.

The following Theorem generalizes one result of [7] and gives full description of the access hierarchy for multi-SSS $\sigma(G)$.

Theorem 2. *Let $A \subseteq P$, $j \in \overline{0,d}$. Then $A \in \widetilde{\Sigma}_j$ if and only if j is the least integer from 0 to d, for which the following system of linear equations (SLE) over ring R is consistent:*

$$G_A x^{\downarrow} = a^j G_0^{\downarrow}. \tag{4}$$

Let's note that the statement of Theorem 2 gives an algorithm for computing the secret keys s_l, $l \in \overline{0, d-j-1}$ by participants from arbitrary coalition $A \in \widetilde{\Sigma}_j$, $j \in \overline{0, d-1}$. This algorithm is obviously based on the equalities (2)–(4) and on the fact of uniqueness representation of elements from R at the form (1).

3. Now, let $\Sigma = \{\Sigma(i) : i \in \overline{0,d}\}$ be a collection of pairwise not intersected subsets of 2^P (the case $\Sigma(i) = \emptyset$ is not excepted) and $\bigcup_{j=0}^{d} \Sigma(i) = 2^P$.

Let's obtain a necessary and sufficient condition that Σ is an access hierarchy for some linear multi-SSS over the ring R and construct a matrix G, according to (2), that determines this multi-SSS (if it exists).

Let's denote by $A_{j,1}, \ldots, A_{j,r_j}$ all minimal elements (by including) of the set $\Sigma(j)$, $j \in \overline{0,d}$.

Theorem 3. *There exists a linear multi-SSS over the ring R for a access hierarchy equal to Σ if and only if*

(1) $\emptyset \in \Sigma(d)$;

(2) *for any $j \in \overline{0,d}$ the set family $\Delta(j) \stackrel{\text{def}}{=} \bigcup_{l=0}^{j} \Sigma(l)$ is monotone, i.e. $A \in \Delta(j)$, $B \in 2^P$, $A \subseteq B$ implies $B \in \Delta(j)$;*

(3) *there exists a matrix C over R with $r = r_0 + \ldots + r_{d-1}$ rows $\overrightarrow{c_{j,l}} = (a^j, \overrightarrow{f_{j,l}})$, $\overrightarrow{f_{j,l}} \in R^n$, $j \in \overline{0, d-1}$, $l \in \overline{1, r_j}$, such that*

   (a) *for any $j \in \overline{0, d-1}$, $l \in \overline{1, r_j}$ the support of $\overrightarrow{f_{j,l}}$ is equal to $A_{j,l}$;*

   (b) *for any $j \in \overline{0, d-1}$ and arbitrary maximal (by including) element X of $2^P \setminus \Delta(j)$ the following SLE over ring R is consistent:*

$$C_{\overline{X}} x^{\downarrow} = a^{d-(j+1)} C_0^{\downarrow},$$

   where $C_{\overline{X}}$ denote the sub-matrix of C containing in its columns with numbers from the set $\overline{X} \stackrel{\text{def}}{=} P \setminus X$, $C_0^{\downarrow}$ is column of C with the number 0.

As conditions (1)–(3) are hold, the rows of a matrix G that determines a multi-SSS for the access hierarchy Σ can be obtained as elements of a minimal generating system of R-module of all solutions of SLE $Cx^{\downarrow} = 0^{\downarrow}$.

Let us note, that in particular case $d = 1$ the statement of Theorem 3 follows from results of [6].

We propose an algorithm, which checks up existence and (in positive case) constructs the required matrix G that determines a linear multi-SSS for a predefined access hierarchy Σ. This algorithm consists of a preliminary evaluations phase and a row forming procedure. The row forming procedure constructs row by row (by the depth-first search) a matrix C that satisfies conditions (a), (b) from Theorem 3. The time complexity of this procedure is equal to

$$T = O\left(n^2 d^n \sum_{j=0}^{d-1} |\overline{\Delta}_j^1| q^{\left(\sum_{j=0}^{d-1}\sum_{l=1}^{r_j}|A_{j,l}|-n\right)}\right)$$

arithmetical operations in R, where $\overline{\Delta}_j^1$ denote the set of all maximal elements of $2^P \setminus \Delta(j)$, $j \in \overline{0, d-1}$.

4. From a practical viewpoint, there is an important problem to obtaining a necessary and sufficient condition under which a linear multi-secret sharing scheme is optimal (among all perfect multi-SSS) for a given access hierarchy, i.e. satisfy the minimality condition for the greatest length of participants shares. Now we have not obtained complete solution of this problem yet, however we can construct some large families of optimally linear multi-SSS over any finite chain ring.

References

[1] Jackson W.-A., Martin K. M., O'Keefe C. M. Multisecret threshold schemes // Advances in Cryptology — CRYPTO'93. Lecture Notes in Computer Science. V. 773. P. 126–135.

[2] Blundo C., De Santis A., Di Crescenzo G., Gaggia A. G., Vaccaro U. Multi-secret sharing schemes // Advances in Cryptology — CRYPTO'95. Lecture Notes in Computer Science. V. 832. P. 150–163.

[3] Vvedenie v kriptographiyu (Yaschenko V. V., ed.). M.: CheRo, 1999. 272 pp.

[4] Brickell E. F. Some ideal secret sharing schemes // J. Combin. Math. and Combin. Comput. 1989. No. 9. P. 105–113.

[5] Blakley G. R., Kabatianski G. A. Linear algebra approach to secret sharing schemes // Preproc. of Workshop on Information Protection.: Moscow, 1993.

[6] van Dijk M. A linear construction of perfect secret sharing schemes // Advances in Cryptology — EUROCRYPT'94. Lecture Notes in Comput. Science. V. 950. P. 23–34.

[7] Ashikhmin A., Barg A. Minimal vectors in linear codes // IEEE Trans. on Inform. Theory. 1998. V. 5. P. 2010–2018.

[8] Kurosawa K., Okada K., Sakano K., Ogata W., Tsujii S. Nonperfect secret sharing schemes and matroids // Advances in Cryptology — EUROCRYPT'93. Lecture Notes in Comput. Science. V. 765. P. 126–141.

[9] Nechaev A. A. Konechie koltsa glavnih idealov // Math. coll. 1973. V. 91. No. 3. P. 350–366.

Gradient Statistical Attack to Block Ciphers

B. Ya. Ryabko, V. A. Monarev, A. N. Fionov, Yu. I. Shokin

Abstract

An attack to block ciphers based on detecting deviations from randomness inside round transformations is suggested. We show the ability of mounting this attack with respect to the ciphers, for which no attacks are known better than the exhaustive key search. The experiments with RC5 block cipher are presented. The implementation is based on the new statistical tests suggested recently by the authors.

1. Introduction

Cryptanalysis of block ciphers attracts much research, and new results in the field are always beneficial for improving constructions of the ciphers. Sometimes the complexity of a new attack (measured in the number of memory units and operations required for mounting such an attack) might be quite large. Nevertheless, even if a relatively small decrease in the attack complexity is achieved, in comparison with previously known methods, this can motivate further development of the cipher design. Thus linear cryptanalysis of DES (see [1]) requires 2^{43} known plaitext–ciphertext pairs and is generally considered infeasible in practice. But it has made an important impact on the design principles of modern block ciphers, that are now resistant to this kind of attack. In this contribution, we suggest a new attack for block ciphers, referred to as a "gradient statistical attack". We show the ability of mounting this attack with respect to the ciphers, for which no attacks are known better than the exhaustive key search.

Consider a (symmetric-key) block cipher with the blocklength n, key length s, and encryption function $E(x, K)$, where $x \in \{0,1\}^n$ denotes a

Supported by RFBR; grant no. 03-01-00495

plaintext block, and $K \in \{0,1\}^s$ a secret key. The typical values of n and s for modern block ciphers are $n = 64$ or $n = 128$, $s = 128$ bits. The majority of block ciphers are iterated, i.e., involve many rounds of transformations usually bracketed by some prologue and epilogue. Either of these, in turn, can sometimes be divided into a number of more simple steps. In respect of this iterated structure, the secret key K is expanded into a sequence of subkeys (or round keys) $k_1, k_2, \ldots, k_t$, where t is the number of "simple steps" in a block cipher. Denote by x_0 the initial state of block x, and by x_i the state after the i-th step. So the complete encryption is $x_t = E(x_0, K)$ and this can be written as

$$x_1 = E_1(x_0, k_1), \quad \ldots, \quad x_t = E_t(x_{t-1}, k_t), \tag{1}$$

where E_i denotes the encrypting transformation at the i-th step.

Example 1. Consider the cipher RC5 [2] with the blocklength 64 and number of rounds r. The encrypting process, with reference to (1), is as follows:

input: (a, b)	$x_0 = (a, b)$
prologue:	
$a \leftarrow a + k_1$	$x_1 = E_1(x_0, k_1)$
$b \leftarrow b + k_2$	$x_2 = E_2(x_1, k_2)$
round 1:	
$a \leftarrow ((a \oplus b) \hookleftarrow b) + k_3$	$x_3 = E_3(x_2, k_3)$
$b \leftarrow ((b \oplus a) \hookleftarrow a) + k_4$	$x_4 = E_4(x_3, k_4)$
$\ldots$	$\ldots$
round r:	
$a \leftarrow ((a \oplus b) \hookleftarrow b) + k_{2r+1}$	
$b \leftarrow ((b \oplus a) \hookleftarrow a) + k_{2r+2}$	$x_t = E_t(x_{t-1}, k_t)$
output: (a, b)	$x_t = (a, b)$

Here $t = 2r + 2$, the length of each subkey is 32 bits. Many other ciphers, including RC6 and AES, may also be represented by (1) with relatively small, e.g., 32 bits or less, subkeys.

We suggest a chosen-plaintext attack for the cipher which can be represented by (1) with relatively small subkeys. Denote the lengths of the secret key and each subkey by $|K|$ and $|k|$, respectively. The exhaustive key-search requires $O(2^{|K|})$ operations (decrypt with $K = 0, 1, \ldots$ until a known x is obtained). Meanwhile, the suggested attack requires

$O(mt2^{|k|})$ operations, where m is the number of ciphertext blocks sufficient for statistical analysis. The attack succeeds in finding the correct subkeys (instead of K itself), provided the statistical test is able to detect deviations from randomness in a sequence of m blocks. It is essential that we use new efficient statistical tests recently suggested in [3, 4].

The experimental part of our research was first concentrated on RC5 block cipher. Currently, we came to the point that RC5 with 8 rounds can be broken by our attack with 2^{33} chosen plaintext–ciphertext pairs.

2. Description of the Attack

The suggested attack belongs to a class of chosen–plaintext attacks. Upon that kind of attack, a cryptanalyst is able to input any information to the cipher and observe the corresponding output. Her aim is to recover the secret key or, which is almost the same, the round keys. Such attacks are of practical interest and it is assumed that the block ciphers must be secure against them.

We consider the block ciphers that can be described by (1) (this includes many of the modern ciphers, if not the all). Notice that the corresponding to (1) sequence for decryption is

$$x_{t-1} = D_t(x_t, k_t), \quad \ldots, \quad x_0 = D_1(x_1, k_1), \tag{2}$$

where D_i denotes the decrypting transformation inverse to E_i.

One of the requirements to a block cipher is that, given a sequence of different blocks as input, the cipher must output a sequence of bits that looks like random. A truly-random sequence may be defined as one generated by a Bernoulli source with equal probabilities of 0's and 1's. We shall loosely call bit sequences "more random" or "less random" depending on how much they differ from a truly-random sequence. One way to measure randomness is to use some statistic on the sequence with the property that less random sequences have greater statistics (to within some probability of error in decision). This may be a well-known x^2 statistic subjected to χ^2 distribution. Denote such a statistic by $\gamma(x)$, where x is a bit sequence.

Denote by $\alpha_1, \alpha_2, \ldots, \alpha_m$ a sequence of input blocks. Let all the blocks be non-random and pairwise different. A possible example would be $\alpha_1 = 1, \alpha_2 = 2, \ldots, \alpha_m = m$, where the numbers are written as n-bit words. For a good block cipher, the encrypted sequence

$$E(\alpha_1, K), \quad E(\alpha_2, K), \quad \ldots, \quad E(\alpha_m, K)$$

must look like random, for any K. Now recall (1). Apply only one step of encryption to the input sequence, denoting the result by $\beta_1, \beta_2, \ldots, \beta_m$:

$$\beta_1 = E_1(\alpha_1, k_1), \quad \ldots, \quad \beta_m = E_1(\alpha_m, k_1).$$

We claim that the sequence β is more random than the sequence α, i.e., $\gamma(\beta) < \gamma(\alpha)$. After the second step of encryption, the sequence

$$E_2(\beta_1, k_2), \quad E_2(\beta_2, k_2), \quad \ldots, \quad E_2(\beta_m, k_2)$$

is more random than β, and so on. Each step of encryption increases the degree of randomness.

Notice the obvious consequence: in decryption according to (2), the randomness of the data decreases from step to step. For example, the sequence

$$D_1(\beta_1, k_1), \quad D_1(\beta_2, k_1), \quad \ldots, \quad D_1(\beta_m, k_1),$$

which is α, is less random than β. But what is important: this is true only if the decryption is done with the valid key. If the key is not valid, denote it by k'_1, then the sequence

$$\alpha'_1 = D_1(\beta_1, k'_1), \quad \ldots, \quad \alpha'_m = D_1(\beta_m, k'_1)$$

will be *more* random than β, $\gamma(\alpha') < \gamma(\beta)$. This is because decrypting with a different key corresponds to further encrypting with that key, which is the well-known multiple encryption principle. So, generally, decryption with an invalid round key increases randomness, while decryption with the valid round key decreases randomness. This difference can be detected by a statistical test.

The suggested gradient statistical attack is mounted as follows. First encrypt the sequence $\alpha_1, \alpha_2, \ldots, \alpha_m$, defined above. Denote the output sequence by ω,

$$\omega_1 = E(\alpha_1, K), \quad \ldots, \quad \omega_m = E(\alpha_m, K).$$

(Recall that the cipher involves t rounds or steps, and the length of subkey at each step is $|k|$.)

Now begin the main procedure of key search. For all $u \in \{0,1\}^{|k|}$ compute a sequence

$$\Gamma_t(u) = D_t(\omega_1, u), \quad D_t(\omega_2, u), \quad \ldots, \quad D_t(\omega_m, u)$$

and estimate its randomness, i.e., compute $\gamma(\Gamma_t(u))$. Find such u^* for which $\gamma(\Gamma_t(u^*))$ is maximal. Assume that unknown subkey $k_t = u^*$. Note that the number of operations at this stage is $O(m2^{|k|})$.

After that, based upon the sequence $\Gamma_t(k_t)$, repeat the similar computations to find subkey k_{t-1}. Using $\Gamma_{t-1}(k_{t-1})$ find k_{t-2}, and so on downto k_1. The total number of operations to recover all subkeys is $O(mt2^{|k|})$.

3. Experiments with RC5

Our experiments were planned as follows. First, we analyzed the degree of randomness of encrypted sequences as a function of the number of steps of encryption. The goal was to find the maximal number of steps at which the tests can tell the encrypted sequence from truly random. Second, we examined the claim our attack is based upon, namely, whether decrypting with a wrong subkey increases randomness in comparison with decrypting using a valid subkey, more exactly, whether these are distinguishable by the tests. Third, based on the results obtained, we started the suggested attack (now in progress). The experiments were carried out on a multiprocessor system containing 10 1-GHz Alpha processors with 1G bytes of memory per device.

To test the statistical properties of RC5, the input sequence

$$\alpha_1\alpha_2\ldots\alpha_m$$

was used with sufficiently large m and several randomly selected keys (Table 1). We can see that the encrypted sequence is stably distinguished from truly random up to the 15th step (with significance level 0.01) which corresponds to the 8th round of RC5.

Table 1. The number of sequences declared non-random

t	m	Number of keys	Number of non-random outputs
10	2^{28}	30	30
11	2^{29}	22	10
12	2^{31}	6	6
13	2^{32}	6	6
14	2^{32}	6	5
15	2^{33}	3	3

To examine the distinguishability of the sequences decrypted with a valid and invalid subkeys, we used the initial sequence of length $m = 2^{24}$ encrypted with 8 and 9 steps under a randomly selected key K and corresponding subkeys k_8 and k_9. In each case the output sequence was decrypted one step downwards with the valid and 5 randomly chosen invalid subkeys $u_1, \ldots, u_5$. All computations repeated 10 times. Table 2 shows the number of cases in which the sequence was declared non-random. One can see that the situations of decrypting using valid and invalid keys are distinguished reliably (10 against 4 and 5 against 0).

Table 2. The number of sequences declared non-random

	valid subkey	u_1	u_2	u_3	u_4	u_5
$t = 8$	10	4	4	4	3	3
$t = 9$	5	0	0	0	0	0

The experiments confirm our assumptions as to principal possibility of the suggested gradient statistical attack. First, the encrypted sequence gets more random as the number of rounds increases. And second, a sequence decrypted with invalid subkey is more random than one decrypted with valid subkey, and the test can detect this. To the date, RC5 with up to 5 rounds is broken under chosen-plaintext attack using our gradient statistical analysis.

References

[1] Menzes A., van Oorschot P., Vanstone S. Handbook of Applied Cryptography. CRC Press, 1996.

[2] Rivest R. L. The RC5 encryption algorithm // B. Preneel, editor. Fast Software Encryption. Second International Workshop (LNCS 1008), P. 86–96. Springer-Verlag, 1995.

[3] Ryabko B. Ya., Stognienko V. S., Shokin Yu. I. A new test for randomness and its application to some cryptographic problems // Journal of Statistical Planning and Inference. V. 123, N. 2. 2003. P. 365–376.

[4] B. Ya. Ryabko, V. A. Monarev. Using information theory approach to randomness testing // Journal of Statistical Planning and Inference. V. 133, N 1. 2005. P. 95–110.

Upper Bounds for Average Differential Approximation Probabilities of Boolean Maps

L. V. Koval'chuk

1. Introduction

Most of the papers devoted to block cipher safety investigations by linear and differential cryptanalysis study SPN or Feistel ciphers where s-blocks are the only non-linear transformations and a key summator realizes binary vector bitwise Boolean addition. Mathematical techniques for estimation of such ciphers safety against the mentioned procedures were developed in [1, 2, 3, 4, 5] and some other papers.

Nevertheless, some modern ciphers (e.g. [6, 7]) are constructed on different principle; in particular, the key summator realizes the addition modulo 2^{16} and 2^{32}, respectively. The existing methods for safety estimation of classical block ciphers ([1, 2, 3, 4, 5]), generally speaking, are appeared to be not applicable for such ciphers.

In [8] new number parameters for s-boxes were introduced, suitable for Feistel ciphers of GOST 28147-89 type, allowing to obtain analytical expressions for upper bounds of average differential and linear characteristic probabilities.

The paper proposes a some set of new upper bounds of average differential approximation probabilities for maps on $\{0,1\}^m$, which are a composition of key summator modulo 2^n and substitution block (for different group operations on the domain and range of such maps).

2. Estimation of Average Differential Approximation Probabilities for Key Summator Modulo 2^m and Substitution Block Composition

Below hereto after we use the following definition:

$$V_m = \{0,1\}^m,\ m \in \mathbb{N};$$
$$f_k(x) = \varphi(x+k),\ x,k \in V_m, \tag{1}$$

where under addition we mean the addition modulo 2^m, and a function $\varphi : V_m \to V_m$ has the following property:

$$\varphi(x_1,x_2) = 2^{m-t}\varphi_2(x_2) + \varphi_1(x_1), \tag{2}$$

where $x_2 \in V_t$, $x_1 \in V_{m-t}$, $\varphi_1 \colon V_{m-t} \to V_{m-t}$, $\varphi_2 \colon V_t \to V_t$ are bijective maps, addition is realized modulo 2^m.

Let us consider the following values:

$$d_{f_k}(\alpha,\beta) = 2^{-m}\sum_{x\in V_m}\delta\left(f_k(x\circ\alpha)\bullet f_k^*(x),\beta\right); \tag{3}$$

$$D_f(\alpha,\beta) = 2^{-m}\sum_{k\in V_m} d_{f_r}(\alpha,\beta), \tag{4}$$

where δ is a Kronecker symbol, the operation "$\circ$" and "$\bullet$" imply some group operations defined on V_m, $f_k^*(x)$ is an inverse element for $f_k(x)$ in respect to the group operation "$\bullet$".

We also apply notations "$\oplus$" and "$+$" for addition modulo 2 and addition modulo 2^l, where the value of l becomes clear from the context. Then we construct the upper estimations for $D_f(\alpha,\beta)$ and $\max\limits_{\alpha,\beta\neq 0} D_f(\alpha,\beta)$ for different operations "$\circ$" and "$\bullet$".

Theorem 1. *With our notations the following inequalities are true:*

(1) $D_f(\alpha,\beta) \le W^{\varphi_2}(\alpha_2,\beta_2)D_{\varphi_1}(\alpha_1,\beta_1)$, *where "$\circ$" and "$\bullet$" denote the addition modulo 2^m;*

$$W^{\varphi_2}(\alpha_2,\beta_2) = 2^{-t}\max_{\eta,\nu\in V_1}\left\{\sum_{x_2\in V_t}\delta\left(\varphi_2(x_2+\alpha_2+\nu)-\varphi_2(x_2)-\eta,\beta_2\right)\right\},$$
$$\alpha = (\alpha_2,\alpha_1),\ \beta = (\beta_2,\beta_1),\ \alpha_1,\beta_1 \in V_{m-t},\ \alpha_2,\beta_2 \in V_t;$$

(2) $D_f(\alpha,\beta) \le U^{\varphi_2}(\alpha_2,\beta_2)D_{\varphi_1}(\alpha_1,\beta_1)$, *where "$\circ$" denotes the addition modulo* 2^m, *"$\bullet$" denotes the addition modulo* 2*;*

$$U^{\varphi_2}(\alpha_2,\beta_2) = 2^{-t}\max_{\nu\in V_1}\left\{\sum_{x_2\in V_t}\delta\left(\varphi_2(x_2+\alpha_2+\nu)\oplus\varphi_2(x_2),\beta_2\right)\right\},$$
$$\alpha=(\alpha_2,\alpha_1),\ \beta=(\beta_2,\beta_1),\ \alpha_1,\beta_1\in V_{m-t},\ \alpha_2,\beta_2\in V_t;$$

(3) $D_f(\alpha,\beta) \le V^{\varphi_2}(\alpha_2,\beta_2)D_{\varphi_1}(\alpha_1,\beta_1)$, *where "$\circ$" denotes the addition modulo* 2, *"$\bullet$" denotes the addition modulo* 2^m*;*

$$V^{\varphi_2}(\alpha_2,\beta_2) = 2^{-2t}\max_{\eta,\mu,\nu\in V_1}\left\{\sum_{x_2,k_2\in V_t}\delta(\varphi_2((x_2\oplus\alpha_2)+k_2+\mu) - \varphi_2(x_2+k_2+\nu)-\eta,\beta_2)\right\},$$
$$\alpha=(\alpha_2,\alpha_1),\ \beta=(\beta_2,\beta_1),\ \alpha_1,\beta_1\in V_{m-t},\ \alpha_2,\beta_2\in V_t;$$

(4) $D_f(\alpha,\beta) \le Y^{\varphi_2}(\alpha_2,\beta_2)D_{\varphi_1}(\alpha_1,\beta_1)$, *where "$\circ$" and "$\bullet$" denote the addition modulo* 2*;*

$$Y^{\varphi_2}(\alpha_2,\beta_2) = 2^{-2t}\max_{\eta,\nu\in V_1}\left\{\sum_{x_2,k_2\in V_t}\delta(\varphi_2((x_2\oplus\alpha_2)+k_2+\nu) \oplus\varphi_2(x_2+k_2+\eta),\beta_2)\right\},$$
$$\alpha=(\alpha_2,\alpha_1),\ \beta=(\beta_2,\beta_1),\ \alpha_1,\beta_1\in V_{m-t},\ \alpha_2,\beta_2\in V_t.$$

Corollary 1. *Let* $m=pt$, $\alpha=(\alpha_p,\dots,\alpha_1)$, $\beta=(\beta_p,\dots,\beta_1)$, $\alpha_i,\beta_i\in V_t$,

$$\varphi(\alpha)=\sum_{i=1}^{p}2^{(i-1)t}\varphi_i(\alpha_i), \tag{5}$$

where $\varphi_i\colon V_t\to V_t$ *are bijective maps,* $i=\overline{1,p}$, *addition in* (5) *is realized modulo* 2^m*. Then under theorem 1 conditions and notations the following inequalities are true, respectively:*

(1) $D_f(\alpha,\beta) \le \prod_{i=2}^{p} W^{\varphi_i}(\alpha_i,\beta_i)D_{\varphi_1}(\alpha_1,\beta_1)$,

(2) $D_f(\alpha,\beta) \le \prod_{i=2}^{p} U^{\varphi_i}(\alpha_i,\beta_i)D_{\varphi_1}(\alpha_1,\beta_1)$,

(3) $D_f(\alpha,\beta) \le \prod_{i=2}^{p} V^{\varphi_i}(\alpha_i,\beta_i)D_{\varphi_1}(\alpha_1,\beta_1)$,

(4) $D_f(\alpha,\beta) \le \prod_{i=2}^{p} Y^{\varphi_i}(\alpha_i,\beta_i)D_{\varphi_1}(\alpha_1,\beta_1)$.

Table 1 shows the results of statistical estimation for probability distributions of parameters $W^\varphi, U^\varphi, V^\varphi, Y^\varphi$ for $\varphi: V_4\to V_4$ (where φ is chosen equiprobably on V_4)

Table 1. The Statistical estimation results for probability distributions of parameters $W^\varphi, U^\varphi, V^\varphi, Y^\varphi$ (for 10^4 substitutions φ on V_4)

Parameter value intervals	Number of substitution for W^φ	Number of substitution for U^φ	Number of substitution for V^φ	Number of substitution for Y^φ
0.00–0.05	0	0	0	0
0.05–0.10	0	0	0	0
0.10–0.15	24	0	784	0
0.15–0.20	3899	225	7075	325
0.20–0.25	4650	5627	1851	5998
0.25–0.30	0	0	12	0
0.30–0.35	1196	1360	245	1065
0.35–0.40	200	2423	29	2274
0.40–0.45	28	20	3	5
0.45–0.50	3	310	1	301
0.50–0.55	0	0	0	0
0.55–0.60	0	0	0	0
0.60–0.65	0	30	0	28
0.65–0.70	0	0	0	0
0.70–0.75	0	5	0	4
0.75–0.80	0	0	0	0
0.80–0.85	0	0	0	0
0.85–0.90	0	0	0	0
0.90–0.95	0	0	0	0
0.95–1.00	0	0	0	0

3. Average Differential Approximation Probability Estimations in the Feistel Scheme

Let $f_k(x,y)=(y, x\oplus\varphi(y+k))$, where $x,y,k\in V_m$, φ has the property (2).We define the following operations on the set V_{2m} :

$$v\circ u=(v^L\oplus u^L, v^R+u^R),$$
$$v\bullet u=(v^L+u^L, v^R\oplus u^R),$$

where $v=(v^L,v^R)$, $u=(u^L,u^R)$, $v^L,v^R,u^L,u^R\in V_m$, "$\oplus$", "$+$" denotes the addition modulo 2 and modulo 2^m, respectively.

Like in the previous section, we consider the values

$$d_{f_k}(\alpha,\beta) = 2^{-2m} \sum_{x\in V_{2m}} \delta\left(f_k(x\circ\alpha)\bullet f_k^*(x),\beta\right); \tag{6}$$

$$D_f(\alpha,\beta) = 2^{-m} \sum_{k\in V_m} d_{f_k}(\alpha,\beta). \tag{7}$$

Lemma 1. *$D_f(\alpha,\beta) = d_{f_k}(\alpha,\beta) = \delta(\alpha^R,\beta^L)\, d_\varphi(\alpha^R,\beta^R-\alpha^L)$, $\forall k\in V_m$, where $d_\varphi(a,b) = 2^{-m}\sum\limits_{x\in V_m}\delta\left(\varphi(x+a)-\varphi(x),b\right)$ and does not depend on k for any $a,b\in V_m$.*

Theorem 2. *$d_\varphi(a,b)\le\Delta^{\varphi_2}(a_2,b_2)d_{\varphi_1}(a_1,b_1)$, where*

$$\Delta^{\varphi_2}(a_2,b_2) = 2^{-t}\max_{\eta,\nu\in V_1}\left\{\sum_{x_2\in V_t}\delta\left(\varphi_2(x_2+a_2+\nu)-\varphi_2(x_2)-\eta,b_2\right)\right\},$$

$a=(a_2,a_1),\ b=(b_2,b_1),\ a_1,b_1\in V_{m-t},\ a_2,b_2\in V_t$.

References

[1] Biham E., Shamir A. Differential cryptanalysis of DES-like cryptosystems // Journal of Cryptology. 1991. V. 4. No. 1. P. 3–72.

[2] Matsui M. Linear cryptanalysis methods for DES cipher // Advances in Cryptology — EUROCRYPT93. Proceedings. Springer Verlag, 1994. P. 386–397.

[3] Knudsen L. R. Practically secure Feistel cipher // Fast Software Encryption (FSE94). Proceedings. Springer Verlag, 1994. P. 211–221.

[4] Kanda M. Practical security evaluation against differential and linear cryptanalyses for Feistel ciphers with SPN round function // Selected Areas in Cryptography (SAC 2000). Proceedings. Springer Verlag, 2001. P. 324–338.

[5] Vaudenay S. Decorrelation: a theory for block cipher security // J. of Cryptology. 2003. V. 16. No. 4. P. 249–286.

[6] Gosudarstvennyi Standart 28147-89. Cryptographic Protection for Data Processing Systems. Government Committee of the USSR for Standarts, 1989.

[7] Lai X., Massey J. L., Murphy S. Markov ciphers and differential cryptanalysis // Advances in Cryptology — EUROCRYPT91. Proceedings. Springer Verlag, 1991. P. 17–38.

[8] Alexeichuk A., Kovalchuk L. Upper Bounds of Maximum Values of Average Differential and Linear Characteristic Probabilities of Feistel Cipher with Adder Modulo 2^m // International conference "Modern problems and new flows in a probability theory". Chernnivtsi, June 19–26, 2005. P. 9–10.

On Constructions of Endomorphic Perfect Ciphers

S. S. Konovalova, S. S. Titov

According to Shannon's theorem, perfect ciphers are solely gamming stream ciphers with a random equiprobable gamma [1, 2]. It is therefore important to study constructions which generate families of such unbreakable ciphers.

This paper is devoted to solving the problems of constructing endomorphic perfect ciphers, both classical and unbreakable, which generalize Shannon's theorem to other types of cryptoattacks [1, 2, 3, 4, 5, 6]. With regard to classical linear endomorphic perfect ciphers described in 1987 by Western cryptologists [5], the following problems have been stated in [4] (see the definitions below and in [4]):

1. Is a robustly perfect bilinear cipher of Construction 1 a multiplication cipher?
2. Is any robustly perfect bilinear cipher a multiplication cipher?
3. Is any robustly perfect linear cipher a bilinear cipher?

In [7], these problems were solved, i.e. a positive answer is given to the first question and a negative answer to the second and third questions. Below, in the first section of the paper, a brief summary of these results is presented based on the theory of finite planes.

The second part explores unbreakable perfect ciphers, namely, $U(L)$- and $O(L)$-immune ciphers [4]. As is shown in [4], the main problem is the construction of endomorphic $U(L)$- and $O(L)$-immune ciphers. The paper presents the constructions of $U(2)$, $U(3)$- and $O(2)$, $O(3)$-immune ciphers based on finite planes and analogues of linear-fractional functions and shows a relationship between them.

We will adhere to the notions, terminology and methods described in [4]. Note only that the enciphering matrix in Construction 1 is a Hankel matrix: see [6], p. 218 and [3, 8].

Let us introduce a few notions: the enciphering rule for Shannon's perfect cipher is given by the equation $y = x*k$, where y is the ciphertext, x is the plaintext, k is the encryption key, $*$ is multiplication in the

corresponding quasigroup. The sets of X plaintexts and Y ciphertexts (cipherquantities) are regarded in this paper as subsets of the vector spaces over the finite field F, and F^r as the space of row vectors of length $r \in \mathbb{N}$ over the field F.

It would be natural to regard the cipher as linear over the field F if $X = F^m$, $Y = F^n$ $(m, n \in \mathbb{N})$ and for each k the encryption operation is linear in x. However, perfect ciphers that would be linear in this sense do not exist (see pp. 66–68 in [4]), but one could construct a perfect cipher that would be linear over F by modifying the definition's conditions: $X = F^m \setminus \{0\}$, $Y = F^n \setminus \{0\}$ and for each $k \neq 0$ the encryption operation is linear in x. Hereinafter a linear cipher will mean a cipher which meets the above conditions. For such a cipher the enciphering rule may be given by the matrix M_k of dimensions $m \times n$. A cipher which is linear over F is robustly perfect [3] if and only if the following conditions are satisfied:

1. For any x, $y \in F^m \setminus \{0\}$ there exists (uniquely) the key $k \in K$ that meets satisfies the condition $y = xM_k$
2. The distribution of the key $P(K)$ is uniform.

We will call a linear cipher bilinear over F, if $X = F^m \setminus \{0\}$, $Y = F^n \setminus \{0\}$, $K = F^s \setminus \{0\}$ for certain m, n, $s \in \mathbb{N}$; $x \in X$ and each element of the matrix M_k is linear in k.

Commonly, three constructions are considered [4] which enable one to construct perfect ciphers. Let $\overrightarrow{x} = (x_1, \ldots, x_m)$, $\overrightarrow{k} = (k_1, \ldots, k_m)$ are non-zero elements of the field $\mathrm{GF}(q^m)$ represented in coordinate form.

Construction 1.

Step 1. Let $\overrightarrow{k} = (k_1, \ldots, k_m) \neq 0$ be the initial vector of a linear recurrent sequence with the maximum possible period over the field $F = \mathrm{GF}(q)$.

Step 2. Using the recursion law let us express each of the following $m - 1$ signs k_{m+1}, $k_{m+2}, \ldots, k_{2m-1}$ of the LFSR as linear combinations of variables $k_1, \ldots, k_m$.

Step 3. As the i-th row of the matrix M_k we take the vector $(k_i, \ldots, k_{i+m-1})$, each coordinate k_j, $j > m$ of which is written as a linear combination (obtained at step 2) of variables $k_1, \ldots, k_m$.

Construction 2. Let us define the encryption rule in accordance with the relationship $y = x \cdot k$ in the field $\mathrm{GF}(q^m)$.

Construction 3. Let us define the encryption rule in accordance with the relationship $y' = x' \cdot k'$ in the field $\mathrm{GF}(q^m)$, where $x' = xA$, $k' = kB$, $y' = yC$, and A, B, C are nonsingular matrices $m \times m$ over $\mathrm{GF}(q)$.

Definition 1. Construction 3 gives a minimal robustly perfect *multiplication* cipher which is bilinear over F.

1. Solution of three problems on three constructions of perfect ciphers

Problem 1. Let us demonstrate that a robustly perfect cipher constructed with the help of Construction 1 is a multiplication one.

Let the bilinear cipher encryption equation be represented by $y = xM_k$, where $\overrightarrow{y} = (y_0, y_1, \ldots, y_{n-1})$, $\overrightarrow{x} = (x_0, x_1, \ldots, x_{n-1})$ are vectors, $M = M_k$ is a square matrix $n \times n$ built with the help of Construction 1.

In accordance with [8] and Chapter Six in [3], let us introduce a companion matrix S of the polynomial $f(x) = a_0 + a_1x^1 + \ldots + a_{n-1}x^{n-1} + x^n$ over the field F as

$$S = \begin{pmatrix} 0 & \ldots & 0 & -a_0 \\ 1 & \ldots & 0 & -a_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 1 & -a_{n-1} \end{pmatrix}$$

We obtain that any key vector $(k_{0+i}, \ldots, k_{n-1+i})$ is related to the vector $(k_{1+i}, \ldots, k_{n+i})$ with the equality

$$(k_{1+i}, \ldots, k_{n+i}) = (k_{0+i}, \ldots, k_{n-1+i})S.$$

Therefore the matrix M_k is composed of rows of the type k, kS, $kS^2, \ldots, kS^{n-1}$ where k is the key vector-row $\overrightarrow{k} = (k_0, k_1, \ldots, k_{n-1})$.

Hence it follows that the matrix M_k is nonsingular for any $k \neq 0$. Since the polynomial f is primitive, then the matrix S generates the field $\mathrm{GF}(2^n)$ of matrix polynomials from S, whereas its degrees generate the multiplicative group of this field, which is cyclic Hence, for any $k \neq 0$ it is possible to determine a degree m such that $k = k_oS^m$, where $k_0 = \overrightarrow{k_o} = (1, 0, \ldots, 0)$ is the "initial" key row, as noted in [4] following [9]. Therefore the matrix M_k is composed of rows of the type k_0S^m, k_0S^{m+1}, $k_0S^{m+2}, \ldots$, where $m = m(k)$ by virtue of this table of correspondence between the degrees m and the key rows k. The vector space of all n-bit rows x may be isomorphically inserted into the vector space of the degrees S over the first row: $\overrightarrow{x} = \overrightarrow{e}S^\ell$, where $\overrightarrow{e} = \overrightarrow{k_0}$. Similar insertion could be performed also for the vector space of the vectors k and y: $\overrightarrow{k} = \overrightarrow{e}S^m \Leftrightarrow M(\overrightarrow{k}) = M_eS^m$, $\overrightarrow{y} = \overrightarrow{e}S^s$,

$\phi(S^s) = \overrightarrow{y}$, $\phi(S^m) = \overrightarrow{k}$, where M_e is the matrix M_k with the vector $\overrightarrow{e}$ in the first row. If $y = x * k$, then $\overrightarrow{y} = \overrightarrow{x}M(\overrightarrow{k})$ for $\overrightarrow{x} = \overrightarrow{e}S^\ell$ and $M_k = M_eS^m$, $\overrightarrow{y} = \overrightarrow{e}S^s$, i.e. for $\overrightarrow{x} = \phi(S^\ell)$ and $\overrightarrow{k} = \phi(S^m)$ we obtain $\overrightarrow{y} = \overrightarrow{e}S^\ell M_eS^m$, $\overrightarrow{y} = \overrightarrow{e}S^{t+m}$, and since ϕ is an isomorphism of spaces, there can be found a degree t such that $\overrightarrow{e}S^\ell M_e = \overrightarrow{e}S^t$. setting $\overrightarrow{u} = \overrightarrow{e}S^t$, we obtain $\phi(S^t) = \overrightarrow{u} \Leftrightarrow \overrightarrow{u} = \overrightarrow{x}M_e$, and if we consider the representation of the elements of the field $\mathrm{GF}(2^n)$ as vectors — the first rows of the degrees of the matrix S, then it turns out that $y = u \cdot k$ in $\mathrm{GF}(2^m)$, i.e. $S^s = S^tS^m$, where $u = xM_e$, which proves the multiplicativity of the cipher since multiplication by M_e is a linear transformation A of the degrees S^ℓ into the space of degrees S^t, and $x * k = \phi(S^\ell) * \phi(S^m) = \phi(S^t \cdot S^m) = \phi((S^\ell)A \cdot S^m)$. Thus, this establishes

Theorem 1. *Construction 1 sets a multiplication cipher of the type $y = u \cdot k$, $u = xM_e$.*

It is convenient to consider the recurrence relation $k_{i+3} = k_{i+2} + k_i$ [4] as an example. Let LFSR be given by the primitive polynomial $f(x) = x^3 + x^2 + 1$. The enciphering matrix has the following form:

$$M_k = \begin{pmatrix} k_1 & k_2 & k_3 \\ k_2 & k_3 & k_1 + k_3 \\ k_3 & k_1 + k_3 & k_1 + k_2 + k_3 \end{pmatrix}$$

The companion matrix and the relationship between its degrees and vectors $\overrightarrow{k}$ and $\overrightarrow{x}$:

$$S = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad M_k = \begin{pmatrix} kE \\ kS \\ kS^2 \end{pmatrix} \quad k_0 = (1,0,0) \quad M_k = \begin{pmatrix} k_0S^m \\ k_0S^{m+1} \\ k_0S^{m+2} \end{pmatrix}$$

$$S^2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad S^3 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \quad S^4 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

$$S^5 = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad S^6 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad S^7 = E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$\overrightarrow{e}S^0M_e = \overrightarrow{e}EM_e = \overrightarrow{e}M_e = \overrightarrow{e} = \overrightarrow{e}S^0$, so that $(S^0)A = S^0$; $\overrightarrow{e}S^6M_e = \overrightarrow{e}S^1$, and therefore $(S^6)A = S^1$; $\overrightarrow{e}S^1M_e = \overrightarrow{e}S^2$, so that

$(S^1)A = S^2$, and since the basis expansion is $S^2 = S^6 + S^1$ (the basis is taken as S^0, S^6, S^1), then $(S^1)A = S^6 + S^1$; therefore the matrix A of linear transformation $S^\ell \longmapsto S^t$ (i.e. $\overrightarrow{x} \longmapsto \overrightarrow{u}$) is given by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = M_e$$

k	m
001	1
010	6
011	2
100	0
101	5
110	4
111	3

x	ℓ
001	1
010	6
011	2
100	0
101	5
110	4
111	3

x	u
001	011
010	001
011	010
100	100
101	111
110	101
111	110

One can easily check that the Cayley table and the table of multiplication in the field of degrees S agree very closely.

Note. The isotopy constructed brings Construction 1 to a multiplicative cipher not only for a primitive but also for any irreducible polynomial.

Indeed, from the analysis of the above agreement it becomes clear that the primitivity of the polynomial is not critical; what is important is only the availability of a table of re-encrypting isomorphism, which, obviously, is true of any irreducible polynomial.

An example is conveniently provided by the polynomial

$$f(x) = x^4 + x^3 + x^2 + x + 1.$$

By virtue of its non-primitivity the companion matrix of the polynomial will have only five and not $2^4 - 1 = 15$ various non-zero degrees; it is, therefore, impossible to obtain with their help all 15 key vectors $\overrightarrow{k}$ and $\overrightarrow{x}$. But the vector space of all 4-bit rows x may be isomorphically inserted into the linear space of the matrices $4 * 4$, setting the insertion by the formula $\varphi(x) = a + bS + cS^2 + dS^3$, where a, b, c, d are elements of the field $\mathbb{Z}_2$. Then the first row of these matrices will be the vector $\overrightarrow{x}$. Below is presented a table showing correspondence between combinations of the elements a, b, c, d and the first row (vector $\overrightarrow{x}$) of the corresponding matrix $\varphi(x)$:

abcd	x
0001	0110
0010	0011
0011	0101

0100	0001
0101	0111
0110	0010
0111	0100

1000	1000
1001	1110
1010	1011
1011	1101

1100	1001
1101	1111
1110	1010
1111	1100

Thus, indeed each of the 15 vectors $\overrightarrow{x}$ is encoded with one set of numerals a, b, c, d, the table of multiplication in the field and in the quasigroup corresponding to each other by virtue of our isotopy set by the matrix

$$M_e = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Problem 2. For solving we use simple observation and Albert's theorem.

Observation 1. A set of matrices M_k leads to a perfect cipher if and only if $\det[M(k') - M(k'')] \neq 0$ for any different non-zero k' and k''.

The matrix M_k is an isomorphic image of the vector k. If this determinant is equal to zero, then, even if $k' \neq k''$, there will exist a non-zero vector $\overrightarrow{x}$ such that $\overrightarrow{x}M(k') = \overrightarrow{x}M(k'')$, which does not meet the definition of perfect cipher.

Theorem (Albert's)**.** *If a quasigroup with a unity is isotopic to the group, then it is isomorphic to it (and hence associative). See, e.g.* [12, 15].

Based on the results of Observation 1, a linear perfect cipher is a special case of the Veblen-Wedderburn system [11], and therefore the solution to the second problem is any such system possessing the property of bilateral distributivity and not reducible to a field. This is, for instance, Albert's semifield and Donald Knuth's semifield (see [11]).

Let us take Donald Knuth's five-dimensional semifield (it is characterized by bilateral distributivity), which is given by the irreducible polynomial $f(x) = x^5 + x^2 + 1$ so that $x_0 = 1$, x, x^2, x^3, x^4 is the basis $\mathrm{GF}(2^5)$, and by the following multiplication table for basis elements:

1	x	x^2	x^3	x^4
x	x^2	x^3	x^4	$x+1$
x^2	x^3	x^4	x^2+1	$x^4+x^3+x^2+x$
x^3	x^4	x^2+1	x^3+x	x^4+x^2+x
x^4	$x+1$	$x^4+x^3+x^2+x$	x^4+x^2+x	x^3+x^2

The encryption matrix is given by:

$$M_k = \begin{pmatrix} k_1 & k_2 & k_3 & k_4 & k_5 \\ k_2 & k_3 & k_4 & k_5 & k_3+k_1 \\ k_3 & k_4 & k_5 & k_3+k_1 & k_4+k_2 \\ k_4 & k_5 & k_3+k_1 & k_4+k_2 & k_5+k_3 \\ k_5 & k_3+k_1 & k_4+k_2 & k_5+k_3 & k_4+k_3+k_1 \end{pmatrix}$$

According to Albert's theorem these semifields provide examples of bilinear non-multiplicative ciphers owing to their non-associativity: $A = x(x^2x^2) = x+1, B = (xx^2)x^2 = x^2+1 \Rightarrow A \neq B$ [11, 7]. In contrast, a multiplicative cipher is given by a quasigroup which is isotopic to the multiplicative group of the field (isotopy is set by linear transformations of A, B, C [4]). The Veblen-Wedderburn systems which are not reduced to fields provide non-Desarguesian finite planes [11]. Such systems do exist, see examples above; hence we have established

Theorem 2. *There exist bilinear non-multiplicative perfect ciphers.*

Problem 3. Let us prove that not any perfect linear cipher is a bilinear cipher. To this end it would suffice to subject the key k to a non-linear reversible transformation, which greatly outnumber the linear ones over the field [10].

For example $y = x \cdot k^h$, $h \neq 1$, $h \in \mathbb{Z}$. For $h = -1$ decryption takes place rather than encryption (the phenomenon of parastrophy). Over the field of Characteristic Two one can take only odd h. It is obvious that, on condition of reciprocal simplicity of h with $2^m - 1$, it is possible to construct a cipher which is linear but not bilinear [7]. Thus, true is

Theorem 3. *Any isotopy of the multiplication group of the field which is non-linear for the key k provides an example of a linear but not bilinear perfect cipher.*

The solution of the third problem is also the Hall system [11], the advantage of which over the above-considered one (where a left-hand unity is present) is that it contains a two-sided unity. Let $f(x) = (x^2 - rx - s)$ be a second-order polynomial which is non-reducible over the field F. The enciphering matrix corresponding to the Hall system is given by [7]:

$$M_k = \begin{pmatrix} k_0 & k_1 \\ \frac{-k_0^2 + rk_0 + s}{k_1} & -k_0 + r \end{pmatrix}, \quad \text{for } k_1 \neq 0$$

$$M(k_0, 0) = k_0 E \quad \text{for } k_1 = 0,$$

where the key $k = \overrightarrow{k} = (k_0, k_1)$.

The eigen-value polynomial of such matrix is the initial

$$\det(\lambda E - M) = \chi_m(\lambda) = f(\lambda) = \lambda^2 - r\lambda - s \neq 0$$

because $f(x)$ is irreducible. Let us now check, in accordance with Observation 1, the determinant of the difference of non-diagonal matrices for different k:

$$M(x, y) = \begin{pmatrix} x & y \\ \frac{-x^2+rx+s}{y} & -x+r \end{pmatrix},$$

$$M(u, v) = \begin{pmatrix} u & v \\ \frac{-u^2+ru+s}{v} & -u+r \end{pmatrix},$$

$$\det[M(x,y) - M(u,v)] = \det \begin{pmatrix} x-u & y-v \\ \frac{-x^2+rx+s}{y} - \frac{-u^2+ru+s}{v} & -x+u \end{pmatrix}$$
$$= \frac{-(xv-yu)^2 + r[(xv-yu)(v-y)] + s(v-y)^2}{yv} \neq 0,$$

because otherwise $z = \frac{xv-yu}{v-y}$ would be the root of the equation $z^2 - rz - s = 0$ belonging to the field GF(q); however, according to the condition, the polynomial $f(z) = z^2 - rz - s$ is irreducible over this field. Hence we established

Theorem 4. *The Hall system provides an example of a linear but not bilinear perfect cipher based on a quasigroup with a two-sided unity which is two-dimensional over a given field.*

Let us now take advantage of the relationship as established [7] between perfect ciphers [4] and finite planes [10, 11] for studying unbreakable perfect ciphers [4].

2. Constructions of perfect ciphers which are immune to other types of crypto-attacks

Let us now go over to considering $U(L)$ and $O(L)$-immune ciphers, and also perpendicular $PA_1(L, \lambda, \mu)$ and cyclic perpendicular arrays $CPA_1(L, \lambda, \mu)$ since the construction of such ciphers is equivalent to constructing perpendicular arrays [4].

Definition 2. $U(L)$-immune cipher is a cipher which is immune to attacks based on unordered L-tuple set of ciphertexts obtained on one key.

The problem of constructing $U(L)$-immune ciphers is equivalent to that of construction perpendicular arrays of special type.

Definition 3. The perpendicular array $PA_\omega(t, \lambda, \mu)$ is a matrix A of the dimensions $\omega \cdot C_\mu^t \times \lambda$, with elements from the set Y of power μ, each row of which consists of λ various elements, and any t of different elements of the set Y are contained exactly in ω rows of the submatrix composed of any t columns of the matrix A.

Hereinafter we will be concerned with only minimal endomorphic ciphers, i.e. ciphers constructed on arrays with $\lambda = \mu$, since according to Theorem 4.2.4 in [4] a study of $U(L)$-immune ciphers with $\lambda < \mu$ and a minimal number of keys π is reduced to the case where $\lambda = \mu$ (on $O(L)$ see p. 152 in [4]).

Theorem (A.Yu.Zubov [4], theorem (4.2.8))**.** *If there exists a perpendicular array $PA_\omega(t, \lambda, \mu)$, then there is also a $U(t)$-immune cipher with the parameters $|X| = \lambda$, $|Y| = \mu$, $|K| = \omega \cdot C_\mu^t$.*

Definition 4. The cyclic perpendicular array $CPA_\omega(t, \lambda, \mu)$ is a perpendicular array $PA_\omega(t, \lambda, \mu)$ which contains as rows with each row all of its cyclic shifts [4].

The construction of a cyclic $U(2)$-immune cipher is reduced to the problem of placing the queens so that they would not threaten each other on a cylindrical chess board $n \times n$ [14]. From the results of [14] it follows, in particular, that $CPA_1(2, 9, 9)$ does not exist. By considering arrays where enciphering is given by the formula $y = i \cdot x + j$ with the key $k = (i, j)$, it was shown [4] that there exists an array $CPA_1(2, 5, 5) = CPA_1(3, 5, 5)$.

Definition 5. The $O(L)$-immune cipher is a cipher immune to attacks based on an ordered L-tuple set of ciphertexts obtained on one key. In this case, in the encryption table $A_\omega(L, \lambda, \mu)$ for $\omega = 1$ any row should contain only one vector from any L of its elements contained in any L columns. As in the case of $U(L)$-immune ciphers, for creating $O(L)$-immune ciphers a matrix with $\lambda = \mu$ is constructed (see p. 152 in [4]). For minimum values of the parameters of an endomorphic $O(L)$-immune cipher we have: $\lambda = \mu$, $\pi = (\lambda!)/((\lambda - L)!)$ [4]. Note that the construction

of an $O(L)$-immune cipher leads to a secret sharing scheme $(L, 2L-1)$ [6]. More opportunities for creating $U(L)$ and $O(L)$-immune ciphers are provided by non-cyclic arrays.

In what follows, a relationship with the first part of this paper is established:

Observation 2. The construction of an endomorphic $O(2)$-immune cipher is reduced to constructing a finite (affine) plane: any two x will have a pair y corresponding to them one-to-one, which determines two points on the plane through which one can draw a unique straight line, and, conversely, any two non-parallel straight lines intersect at a unique point.

A more detailed reasoning does not seem to be necessary and, in fact, it would be a repetition of the introductory notes from the corresponding sections of the classical texts [10, 11, 13].

Thus, related to each linear perfect cipher could be an endomorphic $O(2)$-immune cipher by the formula $\overrightarrow{y} = \overrightarrow{x}M_{\overrightarrow{k}} + \overrightarrow{\ell}$, where $M_{\overrightarrow{k}}$ is an enciphering matrix linear cipher on the key $\overrightarrow{k}$. It is just natural to call such a cipher linear.

Arrays constructed on the Veblen-Wedderburn systems which are not reduced to fields (see above) provide examples of linear $O(2)$-immune ciphers with interesting properties.

Observation 3. A set of matrices M_k brings about a $U(2)$-immune cipher if and only if $\det[M(k')\pm M(k'')] \neq 0$ for any different non-zero k' and k''.

If, in the set of matrices, there are matrices $M(k'')$ and $M(k')$, with $M(k'') = -M(k')$, then $\det[M(k') + (-M(k'))] = 0$, which contradicts Observation 1.

The relationship between $U(2)$-immune and $O(2)$-immune ciphers is rather simple:

Theorem 5. *An endomorphic $U(2)$-immune cipher with the enciphering equation $\overrightarrow{y} = \overrightarrow{x}M_{\overrightarrow{k}} + \overrightarrow{\ell}$, where $M_{\overrightarrow{k}}$ is set of matrices, may be complemented to a linear $O(2)$-immune one; and conversely, if together with each matrix M_k a $O(2)$-immune cipher contains an $-M_k$ matrix, then one matrix from any such pair (any!) may be removed, obtaining a linear $U(2)$-immune cipher.*

Indeed, for constructing a $U(2)$-cipher one needs a set of matrices M_k. If we add to it a set of matrices $-M_k$, the resulting arrays will be an $O(2)$-immune cipher because $\det[\pm M(k') - (\pm M(k''))] = \det[\pm(M(k') \pm M(k''))] = \pm\det[M(k') \pm M(k'')] \neq 0$ (proceeding from

Observation 3). The reverse conclusion, in fact, would repeat the reasonings from [4] in p. 92–96.

Note that the limitation in the reverse statement is essential. The counter example is an affine plane of the Hall system for the polynomial $f(x) = x^2+x+2$ which is primitive over the field GF(3). It is not difficult to check that it is impossible to separate an $U(2)$-immune cipher out of the corresponding linear but not bilinear $O(2)$-immune cipher, and it is not for every enciphering matrix M_k that an enciphering matrix $(-M_k)$ is available.

In [4] in p. 91–92 mention is made of the existence of $PA_1(3, 8, 8)$. It may be implemented in an affine plane according to the formula $\overrightarrow{y} = \overrightarrow{x} \cdot \overrightarrow{k} + \overrightarrow{\ell}$, where multiplication takes place in the field GF(8) by the table of the degrees of the polynomial $f(x) = x^3 + x + 1$ or $f(x) = x^3 + x^2 + 1$. The sought-for array is a cipher which is both $U(3)$- and $O(2)$-immune.

The issue of the possibility of separating a $U(3)$-immune cipher out of a $O(3)$-immune cipher requires separate special consideration.

Families of $O(3)$-immune ciphers may be constructed as encryption arrays based on linear-fractional functions of the type

$$f(x) = (ax + b)/(cx + d),$$

adding to its elements ∞ and defining the values of the function for the following values of x: $f(\infty) = a/c$, $f(-d/c) = \infty$. In [4] (p. 149), [9] it is mentioned that a projective linear group $\mathrm{PGL}(2, \lambda)$ is sharply 3-transitive and therefore may be used for constructing a $O(3)$-immune cipher. Representing this group as linear-fractional transformations seems to be sufficiently natural and geometrically clear [10].

For obtaining new $O(3)$-immune ciphers one should give up the group structure of the set of keys (see the recommendation in p. 151 in [4]). In addition to linear-fractional transformations of a finite field, which is equivalent to using a sharply 3-transitive group $\mathrm{PGL}(2, \lambda)$, we could suggest generalized linear-fractional functions which do not form a group and which are determined through ternary operation of an arbitrary finite affine plane (see [11, 13]): $y = x \cdot m \circ b$ if point (x,y) lies on the straight line from F_m (a family of parallel straight lines) passing through the point $(0,b)$. This operation is determined for any x, m and b, chosen from the set of $\lambda = \mu$ elements. Multiplication xm and summation $x+b$ are determined as special cases of the ternary operation as follows: $xm = x \cdot m \circ 0$, $x+b = x \cdot 1 \circ b$. For a ternary operation, five properties are postulated [11, 13], which are equivalent, as is known, to setting a finite

plane. A linear-fractional function for this ternary operation may be naturally set as follows: $f(x) = (x \cdot a \circ b)/(x \cdot c \circ d)$. Division takes place in the quasigroup of non-zero elements with a multiplication operation: for $c = 0$ $f(x) = (x \cdot a \circ b)/d = x \cdot a' \circ b'$, $f(\infty) = \infty$; for $c \neq 0$ it is sufficient to take $c = 1$ and $f(x) = (x \cdot a' \circ b')/(x \cdot 1 \circ d') = (x \cdot a' \circ b')/(x + d')$, $f(\infty) = a'$, $f(x_0) = \infty$ for $x_0 \cdot 1 \circ d' = x_0 + d' = 0$. This element x_0 is unique. It is indeed very easy to check, using the postulates of ternary operation, that the summation operation determines the quasigroup on the set of all elements, and the multiplication operation determines the quasigroup on the set of non-zero elements; both the right and left divisions are, therefore, unambiguously executable.

However, the above described construction does not always lead to a $O(3)$-immune cipher; for instance, the Hall system does not always lead to it. The property to be an $O(3)$-cipher geometrically means the presence on the plane of projections of special type [10]. It is essential that the linear or linear-fractional function graph passes through any three given points. The projective linear group $\mathrm{PGL}(2, \lambda)$ possesses such projections by virtue of the Pappus configuration, the presence of which is equivalent to the commutativity of the multiplication in the field $\mathrm{GF}(\lambda)$ [10]. Nevertheless this construction provides a result not only for finite fields:

Theorem 6. *In a commutative semifield the list of linear and linear-fractional functions forms an enciphering table for an endomorphic $O(3)$-immune cipher.*

This theorem is proven by direct reasonings, the latter being true of not only the field but also the commutative semifield. The scheme and the idea of the demonstration ares as follows: if there does not exist a linear function passing through three given points, then for the coefficients of the linear-fractional function passing through them we obtain a set of linear equations the determinant of which has the geometric meaning of the "area" of this triangle and is not equal to zero by virtue of computation, which may be performed without using associativity because one of its columns consists of the unities of this semifield.

Conclusion

The use of approaches from geometric algebra has enabled us to solve a number of problems relating to the construction of perfect ciphers, both classical and unbreakable, and to expand the variety of known ciphers.

The development of such algebraic methods will make it possible to go over to studying more unbreakable perfect ciphers.

The authors are grateful to V. V. Yashchenko, M. M. Glukhov and A. A. Makhnev for their attention to this work.

References

[1] Alferov A. P., Zubov A. Yu., Kuzmin A. S., Cheremushkin A. V. Fundamentals of cryptography. Moscow: Gelios ARV, 2001. 480 p. [in Russian]

[2] Shannon C. Works on theory of information and cybernetics. Moscow: IL, 1963. 830 p.

[3] Babash A. V., Shankin G. P. Cryptography. (Series "Aspects of Protection") Ed. Sherstyuk V. P., Primenko E. A. Moscow: SOLON-R, 2002. 512 p. [in Russian]

[4] Zubov A. Yu. Perfect ciphers. Moscow: Gelios ARV, 2003. 160 p. [in Russian]

[5] Massey J., Maurer U., Wang M. Non-expanding, key minimal, robustly-perfect, linear and bilinear ciphers. Procceedings of Crypto'87. 1987. p. 237–247.

[6] Kharin Yu. S., Bernik V. I., Matveyev G. V., Agiyevich S. V. Mathematical and computer basics of cryptology. Minsk: Novoye Znaniye, 2003. 382 p. [in Russian]

[7] Konovalova S. S., Titov S. S. Three problems on three constructions of perfect ciphers. Issues in theoretical and applied mathematics. Proceedings of 36-th Regional Youth Conference. Ekaterinburg: UrO RAN, 2005. P. 37-41. [in Russian]

[8] Gantmakher F. R. Theory of matrices. Moscow: Nauka, 1967. 576 p. [in Russian]

[9] Glukhov M. M., Yelizarov V. P., Nechayev A. A. Algebra (textbook). Moscow: v/ch 33965, 1990. [in Russian]

[10] Artin E. Geometric algebra. Moscow: Nauka, 1969. 284 p. [in Russian]

[11] Hall M. Combinatorics. Moscow: Mir, 1970. 424 p. [in Russian]

[12] Belousov V. D. Basics of the theory of quasigroups and loops. Moscow: Nauka, 1967. 223 p. [in Russian]

[13] Skornyakov L. A. Projective planes. UMN. 6:6 (46), 1951. P. 112-154. [in Russian]

[14] Grebenshchikova N. V., Korepanova N. V., Rusina I. S., Titov S. S. Variants of placement of queens on a cylindrical board. Young scients for transport: Proceedings of the 4-th Science and Technology Conference. Ekaterinburg: UrGUPS. 2003. P. 359-363.

[15] Belousov V. D., Belyavskaya G. B. Latin squares, quasigroups and their applications. Kishinev: Shtinitsa, 1989. 80 p. [in Russian]

Testing of Pseudo-Random Generators by MTD Models

Yu. S. Kharin, A. N. Yarmola

1. Introduction

One of the most important problems of cryptographic information security is statistical testing of pseudo-random sequences [1]

$$x_t \in \mathcal{A} = \{0, 1, \dots, N-1\}, 2 \le N < \infty, t \in \mathbf{N}.$$

Statistical test is a decision rule, which allows according to the observed sample $x_1, \dots, x_T \in \mathcal{A}$ of size T and the given accuracy ε to confirm the hypothesis H_0 ($\{x_t\}$ is the uniformly distributed random sequence(UDRS) (i.e. $\{x_t\}$ is a sequence of independent uniformly distributed random variables) or to take the alternative hypothesis H_1. A review of existing statistical tests in [2] reveals the following facts: 1) the majority of the known tests are based on testing only one of the probabilistic properties of the UDRS; 2) a lot of tests are heuristic and don't fix the set of alternatives; 3) a lot of tests don't have any performance analysis. The problems of development of adequate probabilistic models to describe deviations H_1 from the UDRS model and construction of algorithms to detect and to estimate these deviations are very topical.

This paper is devoted to the indicated problems for the situations where H_1 is specified by high-order Markov dependencies in $\{x_t\}$ and is based on the Raftery's MTD model of discrete-valued time series [3].

2. Properties of the MTD model

Let $\{x_t \in \mathcal{A} : t \in \mathbf{N}\}$ be an s-th order Markov chain, $2 \le s < \infty$, on the probabilistic space $(\Omega, \mathcal{F}, \mathbf{P})$, with a finite state space $\mathcal{A} = \{0, \dots, N-1\}$, $2 \le N < \infty$; $P = (p_{i_0,\dots,i_s})$ be a transition-probability matrix, $p_{i_0,\dots,i_s} = \mathbf{P}\{x_t = i_s \,|\, x_{t-1} = i_{s-1}, \dots, x_{t-s} = i_0\}$, $i_0, \dots, i_s \in \mathcal{A}$, $t > s$. The MTD model was introduced in 1985 by Raftery [3] for the modeling of high-order Markov chains in discrete time. The MTD model

is defined by the special form of the transition-probability matrix P:

$$p_{i_0,\dots,i_s} = \sum_{j=0}^{s-1} \lambda_j q_{i_j i_s}, \quad i_0, \dots, i_s \in \mathcal{A}, \tag{1}$$

where, $Q = (q_{ik})$, $i, k \in \mathcal{A}$, is a stochastic $(N \times N)$-matrix; $\lambda = (\lambda_0, \dots, \lambda_{s-1})'$ is an s-vector, $\lambda_0 + \dots + \lambda_{s-1} = 1$, $\lambda_0 > 0$, $\lambda_j \ge 0$, $j = 1, \dots, s-1$.

An important development of the MTD model is the MTDg model, that uses different transition matrices for the lags [4]:

$$p_{i_0,\dots,i_s} = \sum_{j=0}^{s-1} \lambda_j q^{(j)}_{i_j i_s}, \quad i_0, \dots, i_s \in \mathcal{A}, \tag{2}$$

where $Q^{(j)} = (q^{(j)}_{ik})$, $j \in \{0, \dots, N-1\}$ are stochastic $(N \times N)$-matrices.

Let us start with the analysis of the conditions of ergodicity for the MTDg model. The sequence $\{X_t = (x_{t-(s-1)}, \dots, x_t)' : t > s\}$ is the first order Markov chain with the transition matrix $P_X = (p^X_{ik})$,

$$p^X_{ik} = \begin{cases} p_{i_0,\dots,i_{s-1},k_{s-1}}, & \text{if } k_0 = i_1, \ \dots, \ k_{s-2} = i_{s-1}, \\ 0, & \text{otherwise,} \end{cases}$$

$i, k \in \mathcal{A}^s$.

Lemma 1. *Let $Q^{(0)}$ be an ergodic transition matrix. Then in the case of the MTDg model, the matrix P_X is also ergodic.*

Lemma 2. *In the case of the MTD model, if and only if $Q^{(0)}$ is an ergodic matrix, then the matrix P_X is ergodic.*

Now we study probabilistic properties of the MTDg and the MTD models. Let $\pi^{(t)} = (\pi^{(t)}_0, \dots, \pi^{(t)}_{N-1})'$ be one-dimensional marginal probability distribution of x_t, $t \in \mathbf{N}$, $\pi^{(t)}_i = \mathbf{P}\{x_t = i\}$, $i \in \mathcal{A}$; $\Pi^{(t)} = \left(\pi^{(t)}_{i_1,\dots,i_s}\right)$ be s-dimensional probability distribution of the random vector X_t, $\pi^{(t)}_{i_1,\dots,i_s} = \mathbf{P}\{x_{t-(s-1)} = i_1, \dots, x_t = i_s\}$, $i_1, \dots, i_s \in \mathcal{A}$; $\Pi^* = \left(\pi^*_{i_1,\dots,i_s}\right)$ $i_1, \dots, i_s \in \mathcal{A}$, be s-dimensional stationary probability distribution of Markov chain [5]; $\pi^* = (\pi^*_0, \dots, \pi^*_{N-1})'$ be one-dimensional stationary probability distribution.

Theorem 1. *Let $\{x_t\}$ be a discrete-valued time series satisfying the MTDg model (2). One-dimensional marginal probability distributions $\{\pi^{(t)}\}$ satisfy the equations:*

$$\pi^{(t)} = \sum_{j=0}^{s-1} \lambda_j \left(Q^{(j)}\right)' \pi^{(t-s+j)}, \quad t > s. \tag{3}$$

Theorem 2. *Let one-dimensional marginal probability distributions of the observed Markov chain satisfy the equations:*

$$\pi^{(t)} = \sum_{j=0}^{s-1} \left(A^{(j)}\right)' \pi^{(t-s+j)}, \quad \forall t > s, \tag{4}$$

where $A^{(j)} = (a^j_{ik})$, $a^{(j)}_{ik} \geq 0$, $i,k \in \mathcal{A}$, $j = 0,\ldots,s-1$. Then there exists the vector $\lambda = (\lambda_0,\ldots,\lambda_{s-1})'$, $\lambda_j \geq 0$, $\lambda_0 + \ldots + \lambda_{s-1} = 1$, and stochastic $(N \times N)$-matrices $\{Q^{(j)}\}$, $j = 0,\ldots,s-1$, such that the transition-probability matrix P satisfies the equation (2).

It follows from Theorems 1 and 2 that the property (3) is the identification property of the MTDg model.

Lemma 3. *In the case of the MTDg model, for all $t \geq 2s$ ($i_1,\ldots,i_s \in \mathcal{A}$):*

$$\pi^{(t)}_{i_1,\ldots,i_s} = \prod_{l=0}^{s-1} \left(\sum_{j=l+1}^{s-1} \lambda_j q^{(j)}_{i_{j-l},i_{s-l}} + \sum_{j=0}^{l} \lambda_j \sum_{r=0}^{N-1} q^{(j)}_{ri_{s-l}} \pi^{(t-l-s+j)}_r \right). \tag{5}$$

Theorem 3. *Under the conditions of Lemma 1, the stationary distribution Π^* is ($i_1,\ldots,i_s \in \mathcal{A}$):*

$$\pi^*_{i_1,\ldots,i_s} = \prod_{l=0}^{s-1} \left(\pi^*_{i_{s-l}} + \sum_{j=l+1}^{s-1} \lambda_j \left(q^{(j)}_{i_{j-l}i_{s-l}} - \sum_{r=0}^{N-1} q^{(j)}_{ri_{s-l}} \pi^*_r. \right) \right). \tag{6}$$

We need the following corollary for the estimation of parameters of the MTD model.

Corollary 1. *Under the conditions of Lemma 1, for marginal two-dimensional stationary distribution $\Pi^*(m) = (\pi^*_{ik}(m))$ of the random vector $(x_{t-m}, x_t)'$, $1 \leq m \leq s$, in the case of the MTDg model, it holds:*

$$\pi^*_{ki}(m) = \pi^*_k \pi^*_i + \pi^*_k \lambda_{s-m} \left(q^{(s-m)}_{ki} - \sum_{r=0}^{N-1} q^{(s-m)}_{ri} \pi^*_r \right), \quad i,k \in \mathcal{A}; \tag{7}$$

in particular, in the case of the MTD model:

$$\pi^*_{ki}(m) = \pi^*_k \pi^*_i + \pi^*_k \lambda_{s-m} \left(q_{ki} - \pi^*_i \right), \quad i,k \in \mathcal{A}. \tag{8}$$

3. Statistical estimation of parameters and hypothesis testing

Consider the problem of statistical estimation of parameters of MTD model. Let $X = (x_1,\ldots,x_T)$ be a sample of size T of discrete-valued time series satisfying the MTD model. Now we introduce new estimators based on the property (8) of stationary distributions:

$$\hat{q}_{ki} = \left\{ \sum_{j=1}^{s} \hat{\pi}_{ki}(j)/\hat{\pi}_k - (s-1)\hat{\pi}_i, \text{ if } \hat{\pi}_k > 0; \quad 1/N, \text{ otherwise} \right\}, \tag{9}$$

where

$$\hat{\pi}_i = \sum_{t=s+1}^{T-s+1} \mathbf{I}\{x_t = i\}/(T-2s+1),$$

$$\hat{\pi}_{ki}(j) = \sum_{t=s+j}^{T-s+j} \mathbf{I}\{x_{t-j} = k\}\mathbf{I}\{x_t = i\}/(T-2s+1),$$

$i,k \in \mathcal{A}$, $j = 1,\ldots,s$; $\mathbf{I}\{a\} = \{1, \text{ if } a \text{ is true}; \ 0, \text{ otherwise}\}$;

$$\hat{\lambda} = \arg\min_{\lambda} \sum_{i,k\in\mathcal{A}} \sum_{j=0}^{s-1} \left(z_{ki}(j) - \lambda_j d_{ki} \right)^2, \tag{10}$$

where $z_{ki}(j) = \hat{\pi}_{ki}(s-j)/\hat{\pi}_k - \hat{\pi}_i$, $i,k \in \mathcal{A}$, $j = 0,\ldots,s-1$; $d_{ki} = \hat{q}_{ki} - \hat{\pi}_i$, $i,k \in \mathcal{A}$.

Theorem 4. *In the case of the MTD model, under the conditions of Lemma 2, statistics (9), (10) are asymptotically ($T \to \infty$) unbiased and consistent estimators of Q, λ. Moreover, if $T > 2s$, then the matrix $\hat{Q}$ is a stochastic matrix.*

Unfortunately, we can't use this method to estimate the parameters for the MTDg model. Furthermore, the following theorem holds.

Theorem 5. *Let $\pi^*(j_1,\ldots,j_m)$ be any m-dimensional probability distribution, $1 \leq j_1 < \ldots < j_m \leq s$, $m < s$. Then there exist infinitely many sets of parameters $\{\lambda, Q^{(0)},\ldots,Q^{(s-1)}\}$, such that $\pi^*(j_1,\ldots,j_m)$ is the stationary distribution of the random vector $(x_t, x_{t-j_1},\ldots,x_{t-j_m})'$ for the MTDg model, or there is no set of parameters $\{\lambda, Q^{(0)},\ldots,Q^{(s-1)}\}$ generating this probability distribution $\pi^*(j_1,\ldots,j_m)$.*

We use the estimators (9), (10) to calculate ML-estimates $\tilde{Q}$, $\tilde{\lambda}$ for parameters of the MTD model. The log-likelihood function is

$$l(Q,\lambda) = \sum_{t=s+1}^{T} \ln\left(\sum_{j=0}^{s-1} \lambda_j q_{x_{t-s+j},x_t}\right). \qquad (11)$$

To calculate ML-estimates we have to maximize the log-likelihood function (11). Because of non-linearity of the log-likelihood function (11), the iterative algorithm of estimation of parameters for the MTD model was introduced in [6] by Berchtold. Unfortunately, the initial values chosen in [6] are not consistent estimators. This property decreases effectiveness of the algorithm, because when the sample size T increases, the number of iterations necessary to calculate estimates does not decrease, and closeness of the estimates to the true values of parameters is not guaranteed. That is why we use the estimators (9), (10) as better initial values for the Berchtold's algorithm.

The constructed estimates $\tilde{Q}$, $\tilde{\lambda}$ are useful in testing of pseudo-random generators. In terms of the MTD model the hypothesis $H_0 = \{x_t \text{ is the UDRS}\}$ is equivalent to the hypothesis: $H_0 = \{q_{ki} = 1/N,\ i,k \in \mathcal{A}\}$. To confirm the hypothesis H_0 or the alternative hypothesis $H_1 = \bar{H}_0$ we use the likelihood ratio test:

$$\lambda_T(X) = 2(l(\tilde{Q},\tilde{\lambda}) + T\ln N),$$

which asymptotically ($T \to \infty$) has the significance level $\varepsilon \in (0,1)$:

$$d = d(X) = \{0,\ \lambda_T(X) < \Delta_\varepsilon; \quad 1,\ \lambda_T(X) \geq \Delta_\varepsilon\}, \qquad (12)$$

where Δ_ε is the quantile of the level ε of the chi-square distribution with $N(N-1)$ degrees of freedom.

Table 1. Experiments results

Registers initial values	$\lambda_T(X)$	$d(X)$
$R_1 = 1$, $R_2 = 2$, $R_3 = 3$	3.43	0
$R_1 = 67BA$, $R_2 = 395AB$, $R_3 = BEBE8$	6.48	1
$R_1 = 67BA$, $R_2 = BEBE8$, $R_3 = 395AB$	0.95	0

4. Numerical results

In numerical experiments we analyzed output sequences of the pseudo-random generator, which is used in A5/1 protocol [7]. As a model of this generator we used the MTD model, $N = 2$, $s = 10$, the significance level $\varepsilon = 0.05$, the sample size $T = 1048576$. The results of some experiments are presented in Table 1.

References

[1] Alferov A. P., Zubov A. Yu. Kuzmin A. S., Cheremushkin A. V. Basics in cryptography. Moscow, Helios, 2001. (in Russian)

[2] Kharin Yu. S., Bernik V. I., Matveev G. V., Agievich S. V. Mathematical and computer basics in cryptology. Minsk, Novoe Znanie, 2003. (in Russian)

[3] Raftery A. E. A model for high-order Markov chains // J. R. Statist. Soc., 1985, vol. 47, no. 3, p. 528–539.

[4] Raftery A. E. A new model for discrete-valued time series: autocorrelations and extensions // Rassegna di Metodi Statistici ed Applicazioni, 1985, vol. 3–4, p. 149–162.

[5] Borovkov A. A. Probability theory. Moscow, Nauka, 1986. (in Russian)

[6] Berchtold A. Estimation of the Mixture Transition Distribution Model // J. of Time Ser. Anal., vol. 22, no. 4, 2001, p. 379–397.

[7] Ekdahl P, Johansson T. Another Attack on A5/1 // Proceeding of IEEE International Symposium Information Theory (ISIT) 2001, Washington D.C., 2001.

On the Complexity of Finding of Low Degree Annihilators for a Boolean Function

V. V. Bayev

Algebraic immunity is an important cryptographic characteristic of a Boolean function. Low algebraic immunity of a function means that this function has an annihilating multiplier of low algebraic degree. The problem of annihilator seeking was initially discussed in [1] and [2].

Let $\mathbb{F}_2$ be the field of two elements, $V_n = \mathbb{F}_2^n$ be the vector space of n-tuples over $\mathbb{F}_2$, $\mathcal{F}_n$ be the set of all functions $\mathbb{F}_2^n \to \mathbb{F}_2$. By $\deg f$ denote algebraic degree of a Boolean function $f \in \mathcal{F}_n$. A Boolean function $g \in \mathcal{F}_n$ is called an annihilator of $f \in \mathcal{F}_n$ if $f \cdot g = 0$. We shall use the following notation:

$$A_d(f) := \{g \in \mathcal{F}_n \,|\, f \cdot g = 0,\ \deg g \le d\}.$$

In [2] two algorithms for computation of $A_d(f)$ are introduced. First of them is deterministic and has complexity that bounded from above by some polynomial in 2^n. The other algorithm is probabilistic. Its time of computation has the mathematical expectation that bounded from above by some polynomial in n. But this algorithm has nonzero probability of wrong result. Besides, the algorithm assumes quick random access to input data.

In this report we introduce several deterministic algorithms such that their complexity bounded from above by some polynomial in n and in length of a function representation.

We parameterize functions from $\mathcal{F}_n$ by words of finite length in alphabet $\{0,1\}$. This means that for some set of words $Y_n \subset \{0,1\}^*$ we consider a map $\varphi_n : Y_n \to \mathcal{F}_n$. In these terms, a Boolean function is determined by some pair (n, y), where $n \in \mathbb{N},\ y \in Y_n$. We shall use only "reasonable" maps φ_n. There should exist a polynomial algorithm with input (n, y, x) (here $n \in \mathbb{N},\ y \in Y_n,\ x \in V_n$) such that this algorithm computes the value $\varphi_n(y)(x)$.

Theorem 1 ([4])**.** *Let y be a list of all monomials in polynomial representation of a Boolean function $f_y \in \mathcal{F}_n$, i. e., f_y is equal to the sum of all monomials from the list y. Then there exists an algorithm with the following features. This algorithm has input (n, d, y), it computes a basis of the vector space $A_d(f_y)$, and its time complexity is $O(M_y \cdot (S_n^d)^3)$, where M_y is the number of monomials in the list y and $S_n^d = \sum_{k=0}^{d} C_n^k$.*

Proposition 1. *For arbitrary $f_1, f_2 \in \mathcal{F}_n$ the following relations of vector subspaces of $\mathcal{F}_n$ hold:*

$$A_d(f_1) + A_d(f_2) \subset A_d(f_1 \cdot f_2),$$
$$A_d(f_1 + 1) + A_d(f_2 + 1) \subset A_d(f_1 \vee f_2 + 1),$$
$$A_d(f_1) \cap A_d(f_2) = A_d(f_1 \vee f_2),$$
$$A_d(f_1 + 1) \cap A_d(f_2 + 1) = A_d(f_1 \cdot f_2 + 1).$$

The proof is straightforward.

For $x, \alpha \in V_n,\ x = (x_1, \ldots, x_n),\ \alpha = (\alpha_1, \ldots, \alpha_n)$ we denote

$$x^\alpha := \prod_{i=1}^{n} x_i^{\alpha_i},$$

where

$$x_i^{\alpha_i} := \begin{cases} x_i, & \alpha_i = 1; \\ 1, & \alpha_i = 0. \end{cases}$$

Also, by $B_{n,d}$ denote the set $\{f \in \mathcal{F}_n \,|\, \deg f \le d\}$.

Theorem 2. *There exists an algorithm with the following features. The input of this algorithm is DNF (disjunctive normal form) that corresponds to a function $f \in \mathcal{F}_n$. The output of this algorithm is a basis of the vector space $A_d(f)$. Finally, the time complexity of this algorithm is bounded from above by some polynomial in n and in length of DNF.*

Proof. For any $\alpha \in V_n$ we can compute a basis $\mathcal{B}$ of the vector space $A_d(x^\alpha)$ using the algorithm from Theorem 1. It takes $O\left((S_n^d)^3\right)$ bit operations. Each basis vector $b \in \mathcal{B}$ is represented in the form of b's coordinates in monomial basis of $B_{n,d}$. Let $\sigma \in V_n$ be an arbitrary vector. Consider the map $\varphi_\sigma : B_{n,d} \to B_{n,d}$ that is given by the formula $\varphi_\sigma(g)(x) = g(x + \sigma)$. It is clear that for any $g \in A_d(x^\alpha)$ its image $\varphi_\sigma(g)$ belongs to $A_d((x + \sigma)^\alpha)$. Moreover, φ_σ gives isomorphism $A_d(x^\alpha) \cong A_d((x + \sigma)^\alpha)$. The linear map φ_σ has the matrix Φ_σ of size $S_n^d \times S_n^d$. It is easy to construct a polynomial algorithm that computes this matrix. Thus $\{\Phi_\sigma \cdot b \,|\, b \in \mathcal{B}\}$ is the basis of $A_d((x+\sigma)^\alpha)$. So, we can obtain polynomial algorithm that computes the basis of $A_d((x + \sigma)^\alpha)$.

Let $f \in \mathcal{F}_n$ be represented in the form of DNF:

$$f(x) = \bigvee_{k=1}^{T} (x + \sigma^k)^{\alpha^k},$$

where $\sigma^k, \alpha^k \in V_n$ $(k = 1, \ldots, T)$. Then by Proposition 1,

$$A_d(f) = \bigcap_{k=1}^{T} A_d\left((x + \sigma^k)^{\alpha^k}\right).$$

Therefore, having bases of $A_d\left((x + \sigma^k)^{\alpha^k}\right)$, we can compute a basis of $A_d(f)$ via methods of linear algebra. The time complexity of such algorithm is bounded from above by polynomial in n and in T. □

Theorem 3. *Let $f \in \mathcal{F}_n$ be represented in the form of CNF (conjunctive normal form). Consider the problem of computing of a basis of $A_d(f)$, having CNF of f. We claim that for every $d \geq 0$ this problem is NP-hard.*

Proof. It is clear that

$$f = 0 \iff A_d(f) = B_{n,d} \iff \dim A_d(f) = S_n^d.$$

Thus the problem of computing of a basis of $A_d(f)$, having CNF of f, is polynomial-time reducible to CNF-satisfiability problem, which is NP-complete. □

Now, let a Boolean function $f \in \mathcal{F}_n$ be given by a formula F such that this formula consists of symbols of variables, brackets, and the Boolean operations $\neg, \&, \vee$. We want to search for low degree annihilators recursively. Sometimes we shall replace the operation $\neg$ by "+1". Let F' be some subformula of F, f' be the Boolean function that corresponds to F'. In this notation, for every subformula F' we shall obtain a pair of vector spaces

$$G_d(f') \subset A_d(f' + 1),\ H_d(f') \subset A_d(f'), \tag{1}$$

These vector spaces are given by their basis functions. As above, each basis function is represented in the form of its coordinates in monomial basis of $B_{n,d}$.

In the leaves of recursion tree we have the functions of the form $f_i(x_1, \ldots, x_n) = x_i$. In this case, there exists an algorithm such that its time complexity is polynomial in n and this algorithm computes bases of the following vector spaces:

$$A_d(x_i + 1) = \{g \cdot x_i \,|\, g \in \mathcal{F}_n,\ g \text{ does not depend on } x_i,\ \deg g \leq d - 1\},$$
$$A_d(x_i) = \{g \cdot (x_i + 1) \,|\, g \in \mathcal{F}_n,\ g \text{ does not depend on } x_i,\ \deg g \leq d - 1\}.$$

Therefore, we can assign $G_d(f_i) := A_d(f_i + 1),\ H_d(f_i) := A_d(f_i)$.

Let a subformula be of the form $f' = f_1 + 1 = \neg f_1$. Suppose recursive condition (1) holds for the function f_1. Then, if we make the following assignments

$$H_d(f') := G_d(f_1),$$
$$G_d(f') := H_d(f_1)$$

recursive condition (1) holds for the function f'.

Let a subformula be of the form $f' = f_1 \cdot f_2$. Suppose (1) holds for the functions f_1 and f_2. By definition, put $G_d(f') := G_d(f_1) \cap G_d(f_2)$, $H_d(f') := H_d(f_1) + H_d(f_2)$. Using Proposition 1 and recursive condition (1) for f_1 and f_2, we obtain

$$G_d(f') \subset A_d(f_1 + 1) \cap A_d(f_2 + 1) = A_d(f_1 \cdot f_2 + 1) = A_d(f' + 1),$$
$$H_d(f') \subset A_d(f_1) + A_d(f_2) \subset A_d(f_1 \cdot f_2) = A_d(f').$$

Finally, let a subformula be of the form $f' = f_1 \vee f_2$. Suppose (1) holds for the functions f_1 and f_2. By definition, put $G_d(f') := G_d(f_1) + G_d(f_2)$, $H_d(f') := H_d(f_1) \cap H_d(f_2)$. Again, using Proposition 1 and recursive condition (1) for f_1 and f_2, we obtain

$$G_d(f') \subset A_d(f_1 + 1) + A_d(f_2 + 1) \subset A_d(f_1 \vee f_2 + 1) = A_d(f' + 1),$$
$$H_d(f') \subset A_d(f_1) \cap A_d(f_2) = A_d(f_1 \vee f_2) = A_d(f').$$

We can use this recursive algorithm to compute bases of the vector subspaces $G_d(f) \subset A_d(f + 1)$, $H_d(f) \subset A_d(f)$. It is easy to check that the time complexity of this algorithm is polynomial in n and in length of the formula F.

But this algorithm has a drawback. The vector subspaces $G_d(f)$, $H_d(f)$ might be equal to $\{0\}$, while $A_d(f + 1)$ and $A_d(f)$ are nontrivial. In some cases the inclusion $A_d(f_1) + A_d(f_2) \subset A_d(f_1 \cdot f_2)$ is, in fact, the equality. The remaining part of this report contains two theorems about this property.

Theorem 4. *Let $f_1, f_2 \in \mathcal{F}_n$ be nonzero affine functions such that $f_1 \neq f_2$ and $f_1 \neq f_2 + 1$. Then the vector space $A_1(f_1 \cdot f_2)$ is the following direct sum*

$$A_1(f_1 \cdot f_2) = A_1(f_1) \oplus A_1(f_2).$$

Proof. If $\ell \in \mathcal{F}_n$ is an arbitrary nonzero affine function then $A_1(\ell) = \{0, \ell+1\}$. Hence the sum of subspaces $A_1(f_1)$, $A_1(f_2)$ is direct. We have to prove that $\dim A_1(f_1 \cdot f_2) = 2$.

It is easy to prove that for the functions f_1, f_2 there exists an invertible affine map $\tau : V_n \to V_n$ such that

$$\ell_1(x_1, \ldots, x_n) := f_1 \circ \tau(x_1, \ldots, x_n) = x_1,$$
$$\ell_2(x_1, \ldots, x_n) := f_2 \circ \tau(x_1, \ldots, x_n) = x_1 + x_2.$$

Since τ is invertible, we have the following isomorphisms:

$$A_1(f_1) \cong A_1(f_1 \circ \tau) = A_1(\ell_1),$$
$$A_1(f_2) \cong A_1(f_2 \circ \tau) = A_1(\ell_2),$$
$$A_1(f_1 \cdot f_2) \cong A_1((f_1 \cdot f_2) \circ \tau) = A_1((f_1 \circ \tau) \cdot (f_2 \circ \tau)) = A_1(\ell_1 \cdot \ell_2).$$

Represent $g \in A_1(\ell_1 \cdot \ell_2)$ in the following form

$$g(x_1, \ldots, x_n) = a_0 + \sum\nolimits_{i=1}^{n} a_i x_i.$$

It is obvious that $a_i = 0$ for any $i \geq 3$. Then

$$\begin{aligned} g \in A_1(\ell_1 \cdot \ell_2) &\iff g \cdot \ell_1 \cdot \ell_2 = 0 \\ &\iff (a_0 + a_1 x_1 + a_2 x_2) \cdot x_1 \cdot (x_1 + x_2) = 0 \\ &\iff a_0 x_1 + a_1 x_1 + a_2 x_1 x_2 + a_0 x_1 x_2 + a_1 x_1 x_2 + a_2 x_1 x_2 = 0 \\ &\iff \begin{cases} a_0 + a_1 = 0 \\ a_2 + a_0 + a_1 + a_2 = 0 \end{cases} \iff a_0 + a_1 = 0. \end{aligned}$$

Thus we have three coefficients a_0, a_1, a_2 and one equation $a_0 + a_1 = 0$. Therefore $\dim A_1(f_1 \cdot f_2) = \dim A_1(\ell_1 \cdot \ell_2) = 2$. □

Theorem 5. *Let $f_1, f_2 \in \mathcal{F}_n$ be nonzero functions such that f_2 does not depend on the first m variables and f_1 does not depend on the last $n-m$ variables. Then the vector space $A_1(f_1 \cdot f_2)$ is the following direct sum*

$$A_1(f_1 \cdot f_2) = A_1(f_1) \oplus A_1(f_2).$$

Proof. It is clear that $A_1(f_1) \cap A_1(f_2) = \{0\}$. Let us show that any Boolean function $\ell \in A_1(f_1 \cdot f_2)$ can be represented in the form $\ell = \ell_1 + \ell_2$, where $\ell_1 \in A_1(f_1)$, $\ell_2 \in A_1(f_2)$. Consider $z = (z_1, \ldots, z_n) \in V_n$. By x denote $(z_1, \ldots, z_m)$, by y denote $(z_{m+1}, \ldots, z_n)$. In this notation we have $(x, y) = z$. Let $\ell \in A_1(f_1 \cdot f_2)$ be given by

$$\ell(z) = \sum_{i=1}^{n} a_i z_i + b.$$

Then ℓ can be represented in the form

$$\ell(z) = \ell'(x) + \ell''(y),$$

where

$$\ell'(x) = \sum_{i=1}^{m} a_i z_i, \quad \ell''(y) = \sum_{i=m+1}^{n} a_i z_i + b.$$

Hence

$$\begin{aligned} \ell \in A_1(f_1 \cdot f_2) &\iff \forall x \forall y \quad \ell(x, y) \cdot f_1(x) \cdot f_2(y) = 0 \\ &\iff \forall x \forall y \quad \ell'(x) \cdot f_1(x) \cdot f_2(y) + \ell''(y) \cdot f_1(x) \cdot f_2(y) = 0. \end{aligned} \tag{2}$$

There are only two possibilities:

(a) $\boxed{\forall x \quad \ell'(x) \cdot f_1(x) = 0}$ The condition $f_1 \neq 0$ means that $\exists x_0 : f_1(x_0) = 1$. Substituting x_0 for x in (2), we get

$$\forall y \quad 0 \cdot f_2(y) + \ell''(y) \cdot 1 \cdot f_2(y) = 0 \iff \forall y \quad \ell''(y) \cdot f_2(y) = 0.$$

Thus we have $\ell' \in A_1(f_1)$ and $\ell'' \in A_1(f_2)$.

(b) $\boxed{\exists x_0 : \ell'(x_0) \cdot f_1(x_0) = 1}$ In this case $f_1(x_0) = 1$. If we replace x by x_0 in (2), we obtain

$$\begin{aligned} &\forall y \quad 1 \cdot f_2(y) + \ell''(y) \cdot 1 \cdot f_2(y) = 0 \\ &\iff \forall y \quad (\ell''(y) + 1) \cdot f_2(y) = 0 \\ &\iff \forall y \quad \ell''(y) \cdot f_2(y) = f_2(y). \end{aligned}$$

If we combine the last equation with (2), we get

$$\begin{aligned} &\forall x \forall y \quad \ell'(x) \cdot f_1(x) \cdot f_2(y) + f_1(x) \cdot f_2(y) = 0 \\ &\iff \forall x \forall y \quad (\ell'(x) + 1) \cdot f_1(x) \cdot f_2(y) = 0. \end{aligned}$$

The condition $f_2 \neq 0$ means that $\exists y_0 : f_2(y_0) = 1$. Therefore

$$\forall x \quad (\ell'(x) + 1) \cdot f_1(x) = 0.$$

Finally, we obtain $\ell' + 1 \in A_1(f_1)$, $\ell'' + 1 \in A_1(f_2)$, and $(\ell' + 1) + (\ell'' + 1) = \ell$. □

References

[1] Courtois N., Meier W. Algebraic Attacks on Stream Ciphers with Linear Feedback. Proc. Eurocrypt'2003, LNCS 2656, pp. 345-359, Springer, 2003.

[2] Meier W., Pasalic E., Carlet C. Algebraic Attacks and Decomposition of Boolean Functions. Proc. Eurocrypt'2004, LNCS 3027, pp. 474–491, Springer, 2004.

[3] Armknecht F. On the Existence of low-degree Equations for Algebraic Attacks. Cryptology ePrint Archive, report 2004/185, `http://eprint.iacr.org/2004/185`.

[4] Bayev V. V. On Some Algorithms for Constructing Annihilators of Low Degree for Boolean Functions. Diskretnaya Matematika, to appear (in Russian).

The Affine Transformations Distributing Distortions and A. A. Markov's Problem

B. A. Pogorelov, M. A. Pudovkina

In 1956 A. A. Markov proved the theorem about bijective transformations of words which keep their lengths and do not distribute distortions and raised an issue of the nature of the bijective transformations distributing distortions at most k times, where $k \geq 2$. It was noted that the problem represents big difficulties even for $k = 2$ and corresponding transformations are rather various. The problem was posed to describe these transformations (see [1],[6]). In papers [4], [5], [2] injective transformations were considered not distributing distortions such as replacements, inserts and skips of letters. In [3] such transformations were completely described.

The first example of a transformations group increasing the number of distortions such as replacements of letters at most $k = 2$ times in relation to the Hamming metric was obtained in [8]. This group is a subgroup of the affine group $\mathrm{AGL}_n(2)$.

In this work Markov's problem is considered for transformations of vector spaces over a finite field, and also transformations of modules over residue rings. All affine transformations distributing distortions at most k times, $k = 2, 3, \ldots, n$, are described over these structures. Similar transformations can be built for modules over other rings. The following notation will be used throughout:

- $\mathbb{N}$ — the set of natural numbers;
- n, q — integers from $\mathbb{N}$ not equal to 1;
- R — the residue ring Z_q or the field $GF(q)$;
- $\overline{m,k} = m, m+1, \ldots, k, m < k$;
- $V_n = \{(\alpha_1, \ldots, \alpha_n) \in R^n\}$ — R-module;
- GL_n — the full linear group of degree n over R;
- e_n — the identity $(n \times n)$-matrix;
- $\alpha^{\downarrow}$–the vector - column;
- $\chi(\alpha, \beta) = |\{i : \alpha_i \neq \beta_i, i = \overline{1,n}\}|$ — the Hamming distance between vectors $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$;
- $\vec{0}, \vec{1}$– zero and identity vectors;

- $\|\alpha\|_p = \chi_p(\alpha, \vec{0})$ —the weight of $\alpha \in V_n$ in the metric χ_p, $\|\alpha\| = \|\alpha\|_1$;
- $\varepsilon_j = (0 \dots 0 \overset{j}{1} 0 \dots 0)$, $\delta_j = (\overbrace{1 \dots 1}^{j} 0 \dots 0)$–vectors from $V_n, j = \overline{1,n}$;
- M_n — the set of all $(n \times n)$-matrices over R (linear transformations of V_n in the basis $\varepsilon_1, \dots, \varepsilon_n$);
- $M_{n,p}^{(k)}$ — the set of all transformations from M_n distributing distortions at most k times with respect to the metric χ_p;
- $\widetilde{M}_{n,p}^{(k)}$ — the set of all transformations from $M_{n,p}^{(k)}$ without zero columns;
- $\mathrm{GL}_{n,p}^{(k)} = M_{n,p}^{(k)} \cap \mathrm{GL}_n$;
- $\mathrm{diag}(b_1, b_2, \dots, b_r)$ — the cellwise-diagonal matrix with blocks b_1, b_2, …, b_r;
- $V_n^{(r)} = \{\alpha \in V_n : \|\alpha\| = r\}$;
- $\langle a_1, \dots, a_m \rangle$ — the group generated by $a_1, \dots, a_m$;
- $\mathrm{Isom}\,\chi$ — the isometry group of the metric space (V_n, χ) ;
- $J_n(r)$ — the Jordan cell of order n with root r;
- $\widetilde{S}_n$ —the group of all permutational matrices from GL_n.

Let's consider transformations distributing distortions at most k times with respect to some Hamming submetrics described in [8].

Let $\chi_p(\alpha, \beta) = \left\lceil \frac{\chi(\alpha,\beta)}{p} \right\rceil$ for any $\alpha, \beta \in V_n$, $p \in \{\overline{1,n}\}$. We shall show that χ_p is a metric on V_n.

Proposition 1. *Let $p \in \{\overline{1,n}\}$. Then χ_p is a metric on V_n.*

Proof. It is obvious that $\chi_p(\alpha, \beta) = 0$ iff $\alpha = \beta$ and $\chi_p(\alpha, \beta) = \chi_p(\beta, \alpha)$ for any $\alpha, \beta \in V_n$. Let's check that $\chi_p(\alpha, \beta) \le \chi_p(\alpha, \gamma) + \chi_p(\gamma, \beta)$ for any $\alpha, \beta, \gamma \in V_n$.

We get

$$\left\lceil \frac{\chi(\alpha,\beta)}{p} \right\rceil \le \left\lceil \frac{\chi(\alpha,\gamma)}{p} + \frac{\chi(\gamma,\beta)}{p} \right\rceil \le \left\lceil \frac{\chi(\alpha,\gamma)}{p} \right\rceil + \left\lceil \frac{\chi(\gamma,\beta)}{p} \right\rceil. \quad \square$$

We say that a transformation $g\colon V_n \to V_n$ distributes distortions at most in k times with respect to χ_p if the inequality holds

$$\chi_p(\alpha^g, \beta^g) \le k \chi_p(\alpha, \beta).$$

for any $\alpha, \beta \in V_n$. Let $H_n = \{h_\alpha : \beta \to \beta + \alpha, \forall \alpha \in V_n\}$ be a shift group. By $AG(\le \mathrm{AGL}_n)$ denote $H_n G$ if it is a group.

By $\mathrm{Isom}\,\chi = A\widetilde{S}_n$ [8] it follows that H_n doesn't distribute distortions and $H_n M_{n,p}^{(k)}$ distributes distortions at most k times with respect to χ_p.

This yields that it is enough to describe only transformations $g \in M_n$ satisfying the inequality

$$\|\gamma^g\|_p \le k \|\gamma\|_p \tag{1}$$

for any $\gamma \in V_n$.

Theorem 1. *Let $k, p \in \{\overline{2,n}\}$. The transformation $g \in M_n$ distributes distortions at most k times with respect to χ_p iff $\sum\limits_{j \in J} \left\|\varepsilon_j^g\right\| \le k \cdot p$ for any set $J \in \{\overline{1,n}\}$ cardinality $r \in \{\overline{1,p}\}$.*

Proof. Necessity. Let $g \in M_{n,p}^{(k)}$, i.e. for any $\alpha \in V_n$ the inequality holds

$$\|\alpha^g\|_p = \left\lceil \frac{\|\alpha^g\|}{p} \right\rceil \le k \left\lceil \frac{\|\alpha\|}{p} \right\rceil,$$

which is equivalent $\|\alpha^g\| \in \left\{ \overline{0, k \cdot p \left\lceil \frac{\|\alpha\|}{p} \right\rceil} \right\}$. Then for any $r \in \{\overline{1,p}\}$ and any $\alpha \in V_n^{(r)}$, where $\alpha = \sum\limits_{j \in J} \alpha_j \varepsilon_j$ for some $J \in \{\overline{1,n}\}$, $|J| = r$, and $\alpha_j \ne 0$ for $j \in J$, the following relations hold

$$\|\alpha^g\| = \left\| \left(\sum_{j \in J} \alpha_j \varepsilon_j \right)^g \right\| = \left\| \sum_{j \in J} \alpha_j \varepsilon_j^g \right\| \le \sum_{j \in J} \left\|\varepsilon_j^g\right\| \le k \cdot p.$$

Sufficiency. Consider any set

$$J \subseteq \{\overline{1,n}\}, r = |J|, d \in \left\{ \overline{(t-1) \cdot p + 1, t \cdot p} \right\}, t \ge 1,$$

and any vector $\alpha \in V_n^{(r)}$, $\alpha = \sum\limits_{j \in J} \alpha_j \varepsilon_j$, $\alpha_j \ne 0$ for $j \in J$. Let $J_1, \dots, J_t$ be arbitrary t-partition J such that $|J_i| = p$ for $i \ne t$. Then the following relations hold

$$\|\alpha^g\| = \left\| \left(\sum_{j \in J} \alpha_j \varepsilon_j \right)^g \right\| = \left\| \sum_{j \in J} \alpha_j \varepsilon_j^g \right\| \le$$
$$\sum_{j \in J_1} \left\|\varepsilon_j^g\right\| + \dots + \sum_{j \in J_{t-1}} \left\|\varepsilon_j^g\right\| + \sum_{j \in J_t} \left\|\varepsilon_j^g\right\| \le k \cdot p \cdot t.$$

Hence, for any r and any $\alpha \in V_n^{(r)}$ we have $\|\alpha^g\| \in \left\{ \overline{0, k \cdot t \cdot p} \right\}$. $\square$

For $p = 1$ from theorem 1 we directly obtain the description of transformations $g \in M_n$ distributing distortions at most k times with respect to the Hamming metric.

Corollary 1. *Let $k \in \{\overline{2,n}\}$. The transformation $g \in M_n$ distributes distortions at most k times with respect to the Hamming metric iff the number of nonzero elements in each line of the matrix of the transformation g is at most k.*

In such way in theorem 1 the large class of transformations not necessarily invertible is described. Let's consider some examples of bijective linear transformations from $\mathrm{GL}_{n,1}^{(k)}$.

Corollary 2. *Let $\beta_{(i)}^{\downarrow}$ from V_n be a vector-column with zero ith coordinate, $g_{i,\beta}$ be a matrix which i-column is equal to $\beta_{(i)}^{\downarrow}$, and zero other columns, $i \in \{\overline{1,n}\}$. Then $e_n + g_{i,\beta} \in \mathrm{GL}_{n,1}^{(2)}$.*

The proof follows from corollary 1.

Note that the first example of the transvection $b = e_n + g_{1,\beta}$, $\beta_{(1)} = (0, 1, \ldots, 1)$, distributing distortions at most twice with respect to the Hamming metric was described in [8]. Moreover, $S_n^{(1)} \subset \mathrm{GL}_{n,1}^{(2)}$, where $S_n^{(1)} = \left\langle \widetilde{S}_n, b \right\rangle$ is a isometry group of χ_2.

Corollary 3. *The inclusion*

$$\{\mathrm{diag}(g_1, \ldots, g_t) | g_i \in \mathrm{GL}_{n_i}^{(k)}, i = \overline{1,t}\} \subset \mathrm{GL}_{n,1}^{(k)}$$

takes place for any $k, t \in \{\overline{1,n}\}$ and any ordered t-partition $(n_1, \ldots, n_t)$ of n.

Consider the example of transformations from $\mathrm{GL}_{n,1}^{(k)}$ generating a group for any $k \in \mathbb{N}$.

Proposition 2. *Let $t \in \{\overline{1,n}\}$, $(n_1, \ldots, n_t)$ be any ordered t -partition of n, $k = max\{n_1, \ldots, n_t\}$. Then any element of the group*

$$G(n_1, \ldots, n_t) = \{\mathrm{diag}(g_1, \ldots, g_t) | g_i \in \mathrm{GL}_{n_i}, i = \overline{1,t}\}$$

distributes distortions at most k times with respect to the Hamming metric.

It is clear that

$$\widetilde{S}_n G(n_1, \ldots, n_t) \widetilde{S}_n \subseteq \mathrm{GL}_{n,1}^{(k)}.$$

Corollary 4. *Under the condition of corollary 3 for any $r_1, \ldots, r_t$ from R and any transformation*

$$g \in H_n \widetilde{S}_n \,\mathrm{diag}(J_{n_1}(r_1), \ldots, J_{n_t}(r_t)) \widetilde{S}_n$$

the inequality holds

$$\chi(\alpha^g, \beta^g) \leq 2\chi(\alpha, \beta).$$

Let's find the number of transformations from M_n distributing distortions at most k times with respect to the Hamming metric.

Proposition 3. *Let k be arbitrary integers from $\mathbb{N}$, $k \geq 1$. Then*

$$|M_{n,1}^{(k)}| = \left((q-1) \sum_{i=1}^{k} \binom{n}{i} \right)^n,$$

$$|\widetilde{M}_{n,1}^{(k)}| = \sum_{r=0}^{n} (-1)^r \binom{n}{r} \left((q-1) \sum_{i=1}^{k} \binom{n-r}{i} \right)^n.$$

The proof follows from the inclusion-exclusion method (for example, see [7]).

It is obvious seen that $M_{n,p-1}^{(k)} \subseteq M_{n,p}^{(k)}$ for any $p \geq 2$. We shall prove that $M_{n,p}^{(k)} \setminus M_{n,p-1}^{(k)} \neq \emptyset$. In the following proposition triangular transformations from $M_{n,p}^{(k)} \setminus M_{n,p-1}^{(k)}$ will be considered.

Proposition 4. *Let k, p be arbitrary integers from $\mathbb{N}$, $p \geq 2$, $k \in \{\overline{2,n}\}$. Let also $t = (k-1) \cdot p + 1$, and a transformation g be such that*

$$\begin{aligned}
&\varepsilon_j^g - \varepsilon_j \in \langle \varepsilon_1, \ldots, \varepsilon_{j-1} \rangle, \quad j = \overline{1, t-1}, \\
&\varepsilon_t^g = \delta_t, \\
&\varepsilon_j^g - \varepsilon_j \in \langle \varepsilon_j - t + 1, \ldots, \varepsilon_{j-1} \rangle, \quad j = \overline{t+1, p-1+t}, \\
&\varepsilon_j^g = \varepsilon_j, j = \overline{p+t, n}.
\end{aligned} \tag{2}$$

Then

$$\widetilde{S}_n g \widetilde{S}_n \subseteq M_{n,p}^{(k)} \setminus M_{n,p-1}^{(k)}.$$

Proof. By theorem 1 for construction of transformations

$$g \in M_{n,p}^{(k)} \setminus M_{n,p-1}^{(k)}$$

It is enough to point n linearly independent vectors $\alpha^{(1)}, \ldots, \alpha^{(n)}$ such that $\sum\limits_{j \in J} \left\|\alpha^j\right\| \leq k \cdot p$ for any set $J \in \{\overline{1,n}\}$ cardinality $r \in \{\overline{1,p}\}$. Let $\varepsilon_j^g = \alpha^{(j)}$, $j = \overline{1,n}$.

From (2) we see that vectors $\alpha^{(1)}, \ldots, \alpha^{(n)}$ are linearly independent, thus $g \in \mathrm{GL}_n$. The most weight of sum of r lines is equal to $p - 1 + t$, $r \in \{\overline{1,p}\}$, and $k \cdot (p-1) < p - 1 + t \leq k \cdot p$. □

References

[1] Markov A. A. On transformations not distributing distortions. Selected works, v. 2, 70–93.

[2] Glukhov M. M. Injective maps not distributing distortions such as skips of letters. Discrete mathematics, v. 11, 1999, 20–39.

[3] Glukhov M. M. Injective maps of words not distributing distortions. Proceedings in Discrete mathematics, v. 4, 2001, 17–32.

[4] Babash A. V., Glukhov M. M., Shankin G. P. On transformations the set of words not distributing distortions. Discrete mathematics, v. 9, 1997, 3–19.

[5] Glukhov M. M. Injective maps of words not distributing distortions. Mathematical problems of cybernetics. 1998, v. 7, p. 349–350.

[6] Glukhov M. M., Pogorelov B. A. On some applications of groups in cryptography. Proceedings of "Mathematics and Security of Information Technologies", 2004.

[7] Sachkov V. N. Introduction to combinatorial methods of discrete mathematics. M.: Nauka, 1982.

[8] Pogorelov B. A. Metric spaces Hamming type and A. A. Markov's theorem. To appear.

Subject Session "Mathematical and software support of computer systems security"

Methodology of Dynamic Protection

P. D. Zegzhda, D. P. Zegzhda

The purpose of this report is an attempt to explain the theoretical fundamentals of a new paradigm for the so-called dynamic protection of computer systems which will make it possible to detect intrusion attempts by analyzing the channels of information exchange, to monitor the current security status, to forecast the level of security and to control it, so that the system should be secure continuously. The implementation of this technology will result in the creation of an anticipatory protection strategy, which will permit both to ensure security and prevent its violation.

1. Protection Technologies

The analysis of the existing trends prompts a conclusion about a retrospective succession of protection technologies, which can tentatively be termed as static, active, adaptive and dynamic.

The static protection implies identification of a choice of most dangerous threats whose combination determines the set of functions to be performed and the class of protection which the system should fit. The assortment of protective functions should be adequate to the threats. The principal disadvantage of this technology is the limited choice of threats, which, if becoming wider, might result in inadequate protection.

Other protection technologies under consideration incorporate a more or less developed system of control over the system status, which allows for a wider class of threats to be covered by the protection system, designed as a multi-tier protection, so that decisions can be taken on employing additional means of protection or administrative measures aimed at maintaining security.

The proposed systematic classification of the protection technologies uses two main indicators: the availability of the means of analysis of the system status and the environment where it is functioning, and the security criteria in use (see Table 1).

Table 1. Characteristics of the existing technologies of protection construction

Protection characteristic	Monitoring of objects			Methods of security assessment	Principal characteristics
	System status	Status of protection system	Exchange with the environment		
Static	None	None	Partial	Assessment in accordance with guidelines	Adequate response to threats
Active	Partial	None	Analysis of input data	Analysis of information environment	Reliability of the input data analysis
Adaptive	Partial	Partial	Analysis of input data	Control of the status of the means of protection	Resistance to threats, stable control
Dynamic	Full	Full	Analysis of input data and communication channels	Monitoring of the system security, risk assessment	Invariance of the protection, its adequacy, resistance to vulnerabilities

In order to disclose the meaning of these definitions as the authors understand it, let us introduce the following system of models.

2. The Model of the System

Let us present the system as a set of objects forming a hierarchical structure representing a semantic network. In view of the complexity of a computer system the objects can come under scrutiny at physical, signal, program and algorithmic levels. From the standpoint of information protection the network components can be either sources of information (objects O_i), or its recipients (subjects S_i), the same object being able to function both as a subject and as an object, depending on the circumstances. The subject can perform the usual set of operations on the object — reading, writing, etc. The algebra of relations installed over the sets $\{S_i\}$ and $\{O_i\}$ determines the established model of the security policy.

It is proposed that the system should be generally represented as a network model, a network of frames described by the construction

$$H = \langle I, C_1, C_2, \ldots, C_n, \Gamma \rangle,$$

where:

- I is the set of data units;
- $C_1, C_2, \ldots, C_n$ is the set of the types of links between data units;
- Γ is the map determining the links out of the established set of links between the data units.

The computer system is proposed to be simulated with a G-graph [4] of hierarchical structure with named vertices — functional links.

At that the object will be represented as a system consisting of simple objects O, having no structure and presented as a set of descriptors $\Pi = \langle p_1, \ldots, p_n \rangle$. The simple objects at each level of the hierarchy can be assigned to one of K classes. For this domain of study, for instance, such classes may include files, records, queries, database fields etc., up to the memory segments on the hard disk. The object under analysis is represented as semantic graph SG, which is a structure containing vertices of two types — object type I_n and predicate (procedural) type Y, in correspondence with declarative and procedural ways of knowledge representation, as

$$SG = \langle I, Y, G \rangle,$$

where Y are structural links between vertices.

The object vertex i is determined by three sets $\langle L, K, P \rangle$, where

- L is the type of the given object O;
- K is the name of the class it belongs to;
- P is the collection of descriptors determining the object status.

Each object vertex i is related to one entrance out of set $\{P\}$, determining the properties and the size of the entry and the multitude of links with predicates — vertices $y \in Y$, determining the structural links between simple objects.

The set of types Y is determined by a set of characteristic functions

$$g_{i_1}^{(g)}, g_{i_2}^{(g)}, \ldots, g_{i_m}^{(g)},$$

used to juxtapose the types of links between objects — vertices i_1, $i_2, \ldots, i_m$.

Function $g_{im}^{(g)}$ takes the value of 1, if the relation r, in the general case, of $\langle g \rangle$ type, is true for the set of vertices $i_1, i_2, \ldots, i_m$, belonging to classes $K_1, \ldots, K_N$, otherwise $g_1^{(g)} = 0$. If relation g_1^g is true for vertex

i, there will be a predicate vertex g in graph SG and a link to object vertices $i = 1$. The multitude of links form links $p \in \Gamma$.

The links are marked with names of relations g, i.e. they are role links. The essence of a role link is determined by the transformation function corresponding to the operation performed in the computer system. In dependence of the chosen level of the hierarchy these can be operations with files, responses to queries, transaction formation or performance of protective functions, for example, authentication or cryptographic transformation.

The performed operation can be formalized as a binary relation between the interacting objects as $O_i R O_{i+1}$. It is possible to introduce the notion of operation as a sequence of relations $O_i R_{i+1} \dots R_{k,j} O_j$. Examples of objects O are bits, files, entry fields, program entries. The choice of the object type is determined by the hierarchical level.

The relations determine both possible actions (reading, writing, deletion, assignment of rights) and protective functions (authentication, audit, encryption).

The system is governed by a security policy which determines which chains of relations are legal and which are illegal.

Besides, the system provides for access control by imposing restrictions on certain functions of the links, and for status control.

3. General Model of the Breach of Computer Security

As applied to the problem of computer security, the system status is determined by the following sets at each hierarchical level

$$\langle O, R, Rul, L_{ij} \rangle,$$

where

- O is the set of legal objects $\{O_i\}$;
- R is the set of relations constructed according to the binary principle;
- $\{R_{ij}\}$ is the type of relations as determined for the object type;
- Rul are the rules for controlling the relation chains in compliance with the security policy. The relation chains are constructed using object approach and principles of inheritance, encapsulation, polymorphism;
- L_{ij} are the functions transferred from object to object, constructed according to the type of relations R_{ij} and including reading, writing, changing etc.

The status change takes place in one of the following ways: by changing the set $\{O_i\}$; by implementation of relations R_{ij}; by constructing a chain $R \times R$ involving initialization of functions L_{ij}, which results in actions performed at object O or a change of $\{O_i\}$.

Functions L_{ij} performed with relations R_{ij}, can be classified as follows:

- functions changing the object L^{new}, which results in the change of $\{O_i\}$;
- non-changing functions (reading, copying), which keeps $\{O_i\}$ intact, but may result in violations of Rul;
- functions L_{ij}, including security functions, such as identification O_i; authentication $O_i R O_j$; filtration $O_i R O_j$ by parameters;
- cryptographic transformations of the object O_i to the object $O_i^k \colon O_i^k \xrightarrow[M_k]{} O_i^k$. $M_k(K)$ as $O^k \colon O_i$, the reverse transformation requiring key K generation.

It should be noted that some objects may be of a destructive nature — perform scanning of other objects, destroy them, engage in relations violating $\{Rul\}$.

4. Systematic Classification of Potential Mechanisms of Security Violations

Security violations can take place by using one of the following mechanisms:

Creation of a new object not included in the legal set of objects O or change of parameters (relations) inherent to the existing objects. With this mechanism the following particular cases may occur:

Assertion 1.1. Creation of a "destructive" object, which means that an *attack* was launched.

Assertion 1.2. Potentially excessive rate of creating "legal" objects in comparison with the predetermined one, which results in malfunctioning. The change in object's parameters or the parameters of the function of its transformation is called *vulnerability*.

Change (emergence) of new relations (types of relations) between the objects and violation of Rul.

Assertion 2.1. The emergence of new relations may not be in contradiction to (under control of) the SP, which will be a shortcoming of the means of control, or a sign of vulnerability, i.e. Rul does not change.

Assertion 2.2. The emergence of new relations may be a result of a temporary protection override (internal violation).

Assertion 2.3. Accidental uncontrollable relations may be formed because of implementation faults, *which does not mean an intrusion.*

Assertion 2.4. Formation of new relations following the emergence of new objects represents an *attack.*

Change in functions constructed according to the relation type, where three cases are possible: purposeful change in functions L_{ij} being a consequence of the emergence of new objects (as a rule, this is aimed at the disruption of protective functions and is an attack mechanism). A special case is copying or changing of key data as parameter L_{ij}; uncontrolled but consistent change of L_{ij} because of vulnerability; purposeful exploitation of vulnerability, which is an attack mechanism.

The proposed model makes it possible to construct a full set of mechanisms of security violations and discriminate those which are termed an *intrusion.*

The proposed systematic classification of potential mechanisms of violations can be generally presented in the table format in the following Table 2, which can serve as the basis for the systematic classification of violations, allowing, as a first approximation, to single out the signs which make it possible to distinguish an intrusion as a specific kind of security violations, different from the attack and unauthorized access as a violation of the security policy.

Table 2. The Mechanism in Terms of the Proposed Model

Attack	Creation of a new object (or a change in role links) possessing properties violating rules of access or properties of the protection system *Rul*
Intrusion	Change of parameters and functions implemented by object L_{ij} or by a new object without violation of *Rul*
Vulnerabilities	Change of L_{ij} functions or their parameters
Violation of the rules as specified by the security policy	Mismatch of L_{ij} to the set of *Rul*
User's anomalous behavior	Accidental or deliberate change of parameters of L_{ij} by users or the administrator

As follows from Table 2, under intrusion proper it is reasonable to understand the emergence of a new object in the computer system not leading to a violation of the security policy rules, or a change in the parameters of functions performed by the existing objects, which does not necessarily imply changes in the existing security policy and is not controlled by the access control system.

5. Phenomenological Approach to the Construction of Secure Information Systems

In a general case the problem of the construction of protected information processing system can be formalized in the following way:

- U — is the set of persons participating in the information process (potential users of the computer system), accessing and processing information and engaged in information exchange.
- I — is the set of information objects-containers (documents, books, folders, files etc.) where the information is kept. Information cannot exist by itself — it should be kept in a container.

From the standpoint of security information processes are patterned as the relations of information flows determined by these basic sets. The information flow is understood as an event resulting in the emergence at the destination of a flow of information which was at the point of the outflow of information prior to this event. From the standpoint of security the algorithms of information processing are not important, it is only the exchange of information between the users and the system that is of importance.

There are two kinds of flows:

- $F^W \subseteq U \times I$ — is a relation describing the flows from the users to the containers;
- $F^R \subseteq I \times U$ — is a relation describing the flows from the containers to the users.

In order to be able to make a judgment on the system security, basic theses characterizing the subject area from the standpoint of security should be provided. These theses should be formulated as the following *security axioms*:

1. For each information there should be at least one user who is its *trusted source.* The trusted sources are described by function $TrustSrc\colon I \to U$.
2. For each user there is a known set of information, of which he is an *authorized consumer.* This authority is described by function $Authority\colon U \to I$.

At each moment of time the distribution of information in the system is characterized by the following relations between the users and the information:

1. $Know \subseteq U \times I$ — the relation of knowledge, which determines which information is known to which user.
2. $Create \subseteq U \times I$ — the relation of generation, which determines which user supplies which information.

In the general case the problem of security can be formulated as follows:

The system status is secure if the following *status security criteria* are met:

- The relation of knowledge does not contradict the authority function $Know \subseteq Authority$.
- The relation of generation does not contradict the trusted source function $Create \subseteq TrustSrc$.

The system on the whole is secure if the following *system security criteria* are met:

- The current system status is secure.
- The transitive closure of the relations Know and Create does not contradict the security axioms.

5.1. The Security Model

In a computer system the users cannot process information directly and have to use mediator tools — information processing software which represents their interests within the system. In order to take this into account the following notions are introduced in the model:

- S — the set of subjects;
- O — the set of objects;
- $P, P \subseteq O$ — the set of application software which the users use to process information contained in the objects.
- Id — the relation of identification which correlates the user with at least one subject $Id \subseteq U \times P(S)$.
- Imp — the relation of impersonation determining for each program the subject whose interests it should represent $Imp \subseteq P \times S$.
  Sem — the relation of semantics which establishes a link between the objects and the information contained in them $Sem \subseteq O \times I$.

The set of operations performed by programs on the objects is designated Op and represents a set of relations of $x \subseteq P \times O$ kind, where

x is the type of operation (reading, writing, etc.). The type of operation depends on the nature of the object and the functionality of the program.

Coordination between operations and information flows is described by function $InfFlow$: $Op \rightarrow F$.

Access is described by relation $A \subseteq P \times Op \times O$, which determines the functionality of the programs in regard to operations with the objects.

The security model $SM = \{R, A^A\}$ is represented as a combination of the set of access rights R and the relation of authorized (sanctioned) access $A^A \subseteq S \times O \times P(R)$, determining the rights of the subjects to access the objects.

The means of access control are guided by the security model and prohibit operations contradicting the rules of the model. The functioning of the means of access control are described by the following relations:

- $Map \subseteq Op \times R$ — establishes the correlation between the operations and access rights;
- $A^S \subseteq P \times Op \times O$ — determines the set of program operations on the objects controlled by the means of protection.

The authors propose a new, *phenomenological* approach to the construction of protected systems and to the assessment of the degree of their protection, based on the examination of the phenomenon of vulnerability and the proposed security criteria. In accordance with this approach the security of the system is determined, on the one hand, by the absence of so-called vulnerabilities in it, which are used as a mechanism for violating security, and, on the other hand, by the ability of the means of control and access management to impose the required restrictions, i.e. their tolerance to the unauthorized access and threats in general.

The proposed criterion of security is formulated as follows:

1. The means of access control implement the security model
$$(p, op, o) \in A^{S\rightarrow}\exists(s, o, R) \notin A^A,$$
that $(s, p) \in Imp$, $(op, R) \in Map$.
2. The implementation of the security model is in compliance with the security axiom: for $\forall(s, o, R) \in A^A$ the following conditions are true:
   $\exists u$ and I, are such that $(u, s) \in Auth$ and $(I, o) \in Sem$, that $(u, i) \in Authority$, if $InfFlow(op) \in F^R$ for all op, for which $(op, R) \in Map$, and $(u, i) \in TustSrc$, if $InfFlow(op) \in F^W$ for all op, for which $(op, R) \in Map$.

3. All functionalities of the programs are under control of the means of protection: $A^S \supseteq P \times Op \times O$.

A distinction of the proposed approach to security is that it can be used for the groundwork of protected systems technology development, since its criterion can be used as an objective function in the course of design and development of the protected system.

In the space of the system model the conditions for current security formulated above will include two components:

1. Implementation of the required information process as an interaction model M^G , realized in graph G. Here the current information flow V will be a subset $M^G M^o \in M^G$.
2. Implementation for the whole M^G of the relation $R(s \times o)$ in compliance with the rules of the security policy

$$M^o \in M^G R(s \times o)|_{i=\overline{1,\ldots,n}} \in R^G.$$

5.2. The Vulnerability Model

The opportunities for access emerging because of vulnerability will not necessarily contradict the security policy, that is why we call this kind of access illegal, to distinguish it from the unauthorized access. The illegal access will always be a result of the disturbance of the balance between the functionalities of the application software and the means of protection. The illegal access is not controlled by the means of protection because it is performed either bypassing them or is ignored by them. Like the unauthorized access, the illegal access may result in the violation of the security policy (if any) or in information drain, damage to its integrity, failure of the entire system.

In accordance with the proposed general model the definition of vulnerability will be formalizes as follows: the system is vulnerable if:

- either $\exists(p, op, o) \in A^S$ for which $\exists s, r$, are such that $(p, s) \in Imp$, $(op, R) \in Map, (s, o, R) \notin A^A$
- or $\exists(p, op, o) \in A^S$ is such that $(p, op, o) \notin A^S$.

Vulnerabilities differ from common programming and design errors in that, firstly, they originate in conditions emerging as a result of deliberately created circumstances, which can hardly be accidental, and, secondly, making use of them allows actions which will not be cut short by the means of protection. Vulnerabilities can be divided into two classes — those of vulnerabilities present in the means of protection and of vulnerabilities in the application software. For the means of protection vulnerability means a property of losing the ability to perform its functions under certain conditions. For application software vulnerability means a property to acquire new functionalities under certain conditions, due to which the program acquires an ability to perform one of the following actions: reading/writing data, code execution and consumption of resources. Accordingly, vulnerabilities of the first type stem from errors in programming the means of protection and shortcomings of the security administering, and vulnerabilities of the second type — from shortcomings of the means of protection which makes it impossible to control the actions of the application software, or lack of protection altogether.

6. The Technology of Designing Secure Information Processing Systems

The proposed five fundamental principles of designing protected systems give birth to five components of technology for their construction, each determining the sequence of actions and conditions for the achievement of the required qualities of the protected system. Let us discuss the principal provisions of the developed technologies and the advantages of their use in the design of protected systems.

The new approach to the definition of the notion of a "protected system" now under development at the Specialized Center of Information Protection (SCIP) of the Public Educational Institution (PEI) "St. Petersburg State Polytechnic University" (SPSPU) formulates its main objective as that providing for adequate processing of information flows by a computer system in accordance with the rules of their management which existed before the automated means of information processing came into use. This means that the protected system should, firstly, support only the information flows that existed before it came into use without setting up new ones, and, secondly, to provide for an option of controlling the information flows in compliance with the predetermined set of rules prescribed by the security policy.

Within the framework of this approach the SCIP of PEI "SPSPU" has developed a comprehensive model of security for a protected information system which corresponds to the transit of information flows and their management, based on a generalized idea of security policies and a universal model of the interaction of the system components. Here the main object of modeling is not performance of various operations (access to files, exchange of messages etc.) but information processes behind them. In Fig. 1 main stages of building a security model for an information system using the proposed technology is shown.

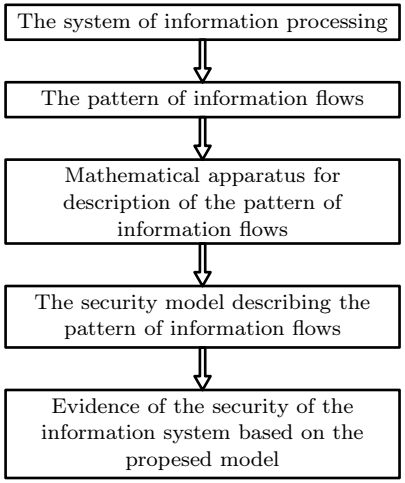

Figure 1. Modeling of information flows

Using this approach it becomes possible to control thoroughly the interaction of the subjects and the objects in line with the transit of information flows and to identify the locations for the means of access control. Besides, abstracting from the specific architecture of the computer system makes it possible to employ standardized protection methods, like means of access control and monitoring independent of the security policy, means for identification and authentication independent from the specifics of applications functioning etc.

The developed model of interactions between the subjects and the objects, reflecting the transit of information flows, can be applied to a very broad class of systems, beginning with operational systems to the distributed computer systems with complex object-oriented application software.

The schematic flow chart of control over the protection system for the proposed technology based on the dynamic principle is shown in Fig. 2.

A distinctive feature of the proposed pattern is the availability of functions listed in Table 1 and an opportunity to implement the strategy of anticipatory dynamic protection to maintain the security at a predetermined level in the dynamic mode.

The analysis of the existing technologies, like interconnecting network firewalls with the systems for intrusion detection and antiviral protection [1], [2], construction of hybrid OS [3], making it possible to prevent certain causes of vulnerabilities, as well as the systems of adaptive control

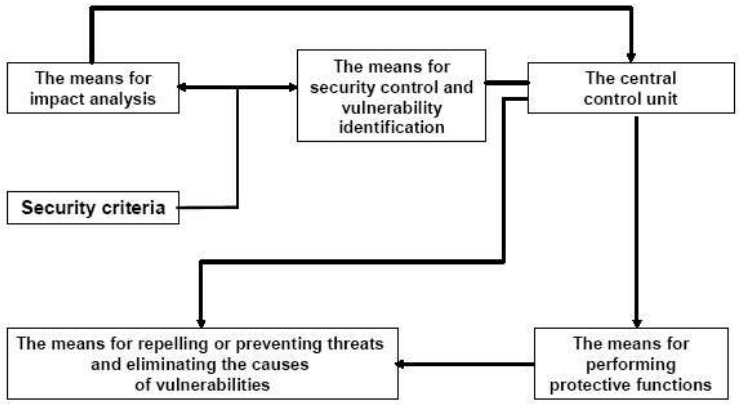

Figure 2. The schematic flow chart of the system of dynamic protection

of cryptographic protection integrating the electronic digital signature, traffic protection and the authentication system leave us with a hope that the proposed theoretical theses will find confirmation in practice.

7. Conclusions

The article describes an approach bringing together the historical development of protection technologies and cites the models and principles of protective arrangements describing various functions of the means of protection at every stage of the development of security technologies.

A possibility to construct an original and new in principle technology of dynamic protection consisting in continuous assessment of security conditions on the basis of analysis of both the system to be protected and the means of protection together with the information input into the system is demonstrated.

References

[1] Vasilyev Yu. S., Zegzhda P. D. Information Security. SPb.: Polytechnic University, 2005 (in Russian).

[2] Zegzhda D. P., Ivashko A. M. Fundamentals of Information Systems Security. Moscow, 2000 (in Russian).

[3] Zegzhda D. P., Vovk A. M. Secure Hybrid Operational System “Linux over Feniks”. Moscow, 2005 (in Russian).

[4] Pospelov D. A. (ed.) Artificial Intellect. Models and Methods. Moscow, Radio i svyaz, 1990 (in Russian).

Information Flow verification in Distributed Systems

F. M. Puchkov

The problem of computation modelling in distributed informational systems (IS) is very relevant today, especially when separate components of the IS are not trusted to each other. This means that there exists some trust function regulating interaction between different untrusted components. In such circumstances there is a need for automatic methods of control of the information flows arising between the components and prevent the untrusted component from access to the forbidden information. The existing means of access control, active audit, indentification/authentication can't provide a solution for this problem. For example, security policies based on MultiLevel Security [1] forbid information flows between higher and lower levels of security lattice hence such policies can't be used for modelling of the computation process in general case.

The main idea of the suggested method is to divide all information flows between the components into secure ones and vulnerable ones. The main problem is to define which flows are secure and which are not.

We shall consider this problem more precisely in the client-server model. Let us denote client as an untrusted component of the IS and server as a component of the IS, which stores the secure data.

We will use the following notation.

- $O_1, \ldots, O_m$ — objects, storing data of different security levels.
- $X_1, \ldots, X_m$ — sets of possible states of these objects.
- $f_1, \ldots, f_s$ — procedures forming the interface of the IS. In general case for the procedure f_k there is a collection of indices $A_k \subseteq \{1, \ldots, m\}$ and $\beta_k \in A_k$, such that

$$f_k \colon \bigotimes_{t \in A_k} X_t \to X_{\beta_k},$$

where the expression $\bigotimes_{t \in A_k} X_t$ means cartesian product of all sets whose number belongs to the set A_k.

It is also useful to define $f_k(P)$ for every subset P as a full image of P of the mapping f_k.

- For every object O_i let us define a consequent system of subsets $\mathfrak{B}_i$ of set of possible states X_i. Every $b \in \mathfrak{B}_i$ is associated with some logical property of the corresponding object. Let us suppose that the family $\mathfrak{B}_i$ is closed over the operations of union, intersection and difference (that correspond logical operations of disjunction, conjunction and negation). Such family is called the algebra of sets.

The events of the following system are either the operation of function call ($[f_k]$) or the operation of object property verification ($[read\ O_i.b]$ where $b \in \mathfrak{B}_i$). The set all events we'll denote as Ev.

The scenario of the computating of the IS is an automata-based function

$$Aut(ev, \sigma) = Aut : Ev \times \{true, false\} \to Ev.$$

Here $\sigma \in \{true, false\}$ is the result of the logical operation ev if ev is an event of property verification. Otherwise σ is arbitrary and Aut doesn't essentially depend on this argument. In any case the operation-event $Aut(ev, \sigma)$ will be executed next after ev.

Let us consider that some computational scenario is given. The client can influence on the scenario execution by means of changing states of some objects (the ones that are open for the client) But as the client is an untrusted component of the IS, we must restrict his capability to influence other (secure) objects. Let us suppose that for each object O_i in its property algebra $\mathfrak{B}_i$ there are marked out some properties holding or not holding of which shouldn't be influenced by the client. We will denote the set of these properties as $\mathcal{A}_i$ and suppose that this family is also closed over operations of union, intersection and difference. Thus for every object O_i we get the additional subalgebra of sets (properties) $\mathcal{A}_i$ of the algebra $\mathfrak{B}_i$.

The main idea of this paper is to determine whether for the given scenario described above conditions of lack of correlation are fulfilled. The negative answer to this question would mean that the IS has a potential vulnerability.

Definition. *Information flow* — is a set of vectors of states of objects in IS:

$$\mathcal{F} \subseteq \bigotimes_{t=1}^{m} X_t.$$

Let us consider the family $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_m)$ and say that the flow $\mathcal{F}$ *preserves* $\mathcal{A}$ iff for all $(\alpha_1, \ldots, \alpha_m), (\beta_1, \ldots, \beta_m) \in \mathcal{F}$, for all i and $n_i \in \mathcal{A}_i$ the condition $\alpha_i \in n_i$ implyies $\beta_i \in n_i$. In other words, the client can't enforce the object to change its state is the way that can be recognized by means of the family of properties $\mathcal{A}$.

Our primary goal in this paper is to determine whether the information flow $\mathcal{F}$ preserves the family $\mathcal{A}$ at any point of time.

Let us enumerate the basic operations on the flows.

- Merge of the flows $\mathcal{F}_1, \mathcal{F}_2$: $\mathcal{F}' = \mathcal{F}_1 \cup \mathcal{F}_2$.
- Accepting property $O_t.a$: $\mathcal{F}' = \{(a_1, \ldots, a_m) \in \mathcal{F} : a_t \in a\}$.
- Substitution. Let's suppose that procedure f has two incoming parameters: O_1, O_2 and one outcoming O_3 and the information flow before calling f was $\mathcal{F} = \{(\alpha_i, \beta_i, \gamma_i)\}$, then the information flow after calling f is generated by the substitution operation: $\mathcal{F}' = \{(\alpha_i, \beta_i, f(\alpha_i, \beta_i))\}$.

When executing different operations on the flows, their power can increase as an exponent. Taking this into account we won't try to calculate the flow at every point of the scenario. Instead we will show the method according to which in every point it is sufficient to know only the fixed number of elements in the flow.

Lemma 1. *Let $\mathcal{F}_1, \mathcal{F}_2$ — are two flows, $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathcal{F}_1$, $\beta = (\beta_1, \ldots, \beta_m) \in \mathcal{F}_2$ — their elements correspondingly. The flow $\mathcal{F}$ is generated using the merge operation on $\mathcal{F}_1, \mathcal{F}_2$. Then if $\mathcal{F}_1, \mathcal{F}_2$ preserve $\mathcal{A}$ and the flow, consisting of the two elements $\{\alpha, \beta\}$ preserves $\mathcal{A}$, then $\mathcal{F}$ preserves $\mathcal{A}$.*

Lemma 2. *Let $\mathcal{F}'$ is generated using the accepting property operation on the flow $\mathcal{F}$ and $\mathcal{F}$ preserves $\mathcal{A}$. Then $\mathcal{F}'$ preserves $\mathcal{A}$.*

The proofs of both lemmas are evident.

Let $\mathcal{D}$ is some algebra of properties with finite unit X and $\alpha \subseteq X$ is some subset. Let's consider $\mathcal{D}_\alpha = \{d \in \mathcal{D} : \alpha \subseteq d\}$ and assign

$$[\alpha] = [\alpha]_{\mathcal{D}} = \bigcap_{\beta \in \mathcal{D}_\alpha} \beta$$

to be the minimal element of the algebra that contains α.

Lemma 3. *Let $p, q \subseteq X$ are arbitrary and $\mathcal{D}$ — is some algebra of properties with finite unit X. Then*

$$[p \cap [q]] = [p] \cap [q]. \tag{1}$$

Proof. Let algebra $\mathcal{D}$ is generated by the finite collection of pairwise disjoint nonempty sets $D_1, \ldots, D_r$ (such collection always exists because X is finite). Than we can conclude that $D_i \in \mathcal{D}$ and $X = \bigsqcup_{i=1}^{r} D_r$. Let $J, K \subseteq \{1, \ldots, r\}$ — some subsets of indices such that $[p] = \bigsqcup_{j \in J} D_j$, $[q] = \bigsqcup_{k \in K} D_k$. Thus, $p \cap D_j \neq \emptyset$, $q \cap D_k \neq \emptyset$ for every $j \in J$ and for every $k \in K$. Actually, if we suppose the opposite fact, then considering $J \setminus \{j\}$ (or $K \setminus \{k\}$), and we would get some smaller element of the algebra $\mathcal{D}$, that contains p (or q). This means that if $s \in J \cap K$, then the following confirmations are true:

- $D_s \subseteq [q] \implies (p \cap [q]) \cap D_s \neq \emptyset$, and, taking into account the property of the family $\{D_i\}$, we can write that $D_s \subseteq [p \cap [q]]$;
- $(D_s \subseteq [p]) \wedge (D_s \subseteq [q]) \implies D_s \subseteq [p] \cap [q]$.

From the other hand we can confirm that $[p \cap [q]] \subseteq [[p] \cap [q]] = [p] \cap [q] \subseteq \bigsqcup_{s \in J \cap K} D_s$. Thus, $[p \cap [q]] = [p] \cap [q] = \bigsqcup_{s \in J \cap K} D_s$. □

Lemma 4. *The flow $\mathcal{F}$ preserves the family $\mathcal{A}$ iff for every $(\alpha_1, \ldots, \alpha_m)$, $(\beta_1, \ldots, \beta_m) \in \mathcal{F}$ and for every i holds:*

$$[\alpha_i]_{\mathcal{A}_i} = [\beta_i]_{\mathcal{A}_i}. \tag{2}$$

Proof. Everywhere in this proof instead of writing $[\{x\}]_{\mathcal{A}_i}$ we will write $[x]$. Let us suppose that (2) holds and we'll prove the that $\mathcal{F}$ preserves $\mathcal{A}$. It's necessary to verify that if $n_i \in \mathcal{A}_i$, then $\alpha_i \in n_i \iff \beta_i \in n_i$. Obviously we have:

$$\begin{gathered}\alpha_i \notin n_i \iff [\alpha_i \cap n_i] = \emptyset \overset{(1)}{\iff} [\alpha_i] \cap n_i = \emptyset \overset{(2)}{\iff} \\ [\beta_i] \cap n_i = \emptyset \iff [\beta_i \cap n_i] = \emptyset \iff \beta_i \notin n_i.\end{gathered} \tag{3}$$

Inversely, let $\mathcal{F}$ preserves $\mathcal{A}$. We'll show that (2) holds. Let's suppose the opposite fact: $[\alpha_i] \cap n_i = \emptyset$, but $[\beta_i] \cap n_i \neq \emptyset$. Then according to (3) we get that $\alpha_i \notin n_i$ and $\beta_i \in n_i$, and this equation contradicts that $\mathcal{F}$ preserves $\mathcal{A}$. □

Let's denote that procedure f_k *preserves* $\mathcal{A}$, iff for every flow $\mathcal{F}$, preserving $\mathcal{A}$, the result of the substitution operation on $\mathcal{F}$ by procedure f_k (the flow $\mathcal{F}'$) preserves $\mathcal{A}$.

Theorem. *Procedure f_k: $\bigotimes_{\alpha \in A_k} X_\alpha \to X_{\beta_k}$ preserves $\mathcal{A}$ iff for every $a_\alpha \in X_\alpha$, $\alpha \in A_k$ holds*

$$\left[f_k\left(\{a_\alpha\}_{\alpha \in A_k}\right)\right]_{\mathcal{A}_{\beta_k}} = \left[f_k\left(\{[a_\alpha]_{\mathcal{A}_\alpha}\}_{\alpha \in A_k}\right)\right]_{\mathcal{A}_{\beta_k}}. \tag{4}$$

Proof. Let's suppose that $\beta_k = 1$ for the purpose of simplification.

Adequacy. Let $\mathcal{F} = \{(\gamma_{1,t}, \ldots, \gamma_{m,t}) \mid t \in T\}$ — some flow preserving $\mathcal{A}$. Than

$$\mathcal{F}' = \left\{\left(f_k\left(\{\gamma_{\alpha,t}\}_{\alpha\in A_k}\right), \gamma_{2,t}, \ldots, \gamma_{m,t}\right) \mid t \in T\right\}$$

is the flow, generated by the substitution operation. If we prove that (2) holds for $\mathcal{F}'$, then according to the proved lemma we'll get that $\mathcal{F}'$ preserves $\mathcal{A}$. As (2) holds for $\mathcal{F}$ and $\mathcal{F}'$ differs from $\mathcal{F}$ only in the first component, it is sufficient to consider the case $i = 1$:

$$\left[f_k\left(\{\gamma_{\alpha,t_1}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = \left[f_k\left(\{\gamma_{\alpha,t_2}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1}.$$

As $\mathcal{F}$ preserves $\mathcal{A}$, we get

$$[\gamma_{\alpha,t_1}]_{\mathcal{A}_\alpha} = [\gamma_{\alpha,t_2}]_{\mathcal{A}_\alpha} = n_\alpha.$$

- Let's in (4) assign $a_\alpha := \gamma_{\alpha,t_1}$. Than:

$$\left[f_k\left(\{\gamma_{\alpha,t_1}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = \left[f_k\left(\{[\gamma_{\alpha,t_1}]_{\mathcal{A}_\alpha}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = \left[f_k\left(\{n_\alpha\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = N_1.$$

- Let's in (4) assing $a_\alpha := \gamma_{\alpha,t_2}$. Than:

$$\left[f_k\left(\{\gamma_{\alpha,t_2}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = \left[f_k\left(\{[\gamma_{\alpha,t_2}]_{\mathcal{A}_\alpha}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = \left[f_k\left(\{n_\alpha\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = N_1.$$

Necessity. Let's consider any collection $a_\alpha \in \mathcal{A}_\alpha$, $\alpha \in A_k$ and arbitrary append it to the full collection $a_i \in \mathcal{A}_i$, $i = 1, \ldots, m$. Let's consider the following flow consisting of the two elements:

$$\mathcal{F} := \left\{(a_1, \ldots, a_m),\ ([a_1]_{\mathcal{A}_1}, \ldots, [a_m]_{\mathcal{A}_m})\right\}.$$

According to the prooved lemma this flow preserves $\mathcal{A}$. Thus the flow

$$\mathcal{F}' = \left\{\left(f_k(\{a_\alpha\}_{\alpha\in A_k}), a_2, \ldots, a_m\right), \left(f_k(\{[a_\alpha]_{\mathcal{A}_\alpha}\}_{\alpha\in A_k}), [a_2]_{\mathcal{A}_2}, \ldots, [a_m]_{\mathcal{A}_m}\right)\right\},$$

which was generated by the substitution operation on $\mathcal{F}$ also preserves $\mathcal{A}$. In particular, this means that

$$\left[f_k\left(\{a_\alpha\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1} = \left[f_k\left(\{[a_\alpha]_{\mathcal{A}_\alpha}\}_{\alpha\in A_k}\right)\right]_{\mathcal{A}_1}.$$

If we compare the last equation with (4) and remember that $\beta_k = 1$, we conclude that the theorem has been proved. □

The theorem above gives us the means for verifying computational scenarios for some properties on noninterference of the untrusted client. In the distributed systems such means could be used when describing the security policies based on the trust relations. In this case the trust function is defined as a filter of data, declassifying from one component of the information system to another one. In this paper such filter was described by the family of properties $\mathcal{A}$ of the objects. As we haven't made any assumptions on the structure or the power of the collection, then this model is general in the sense of the mentioned interaction of client and server.

The presented model is more general than the model of MultiLevel Security, as it doesn't require such strict restrictions on the competent information flows from one hand and it is still sufficiently strict to prove some invariant properties of computation safety from the other hand.

References

[1] Denning D. *A Lattice Model of Secure Information Flow* // Communication of ASM, 19:5, May 1976. — P. 236–243.

[2] Moonen L. *Data Flow Analysis For Data Engineering.* — University of Amsterdam, 1996.

[3] Holloway G., Dimock A. *The Machine SUIF Bit-Vector Data-Flow-Analysis Library.* — Harvard University, July 2002.

[4] Mantel H., Sands D. *Controlled Declassification based on Interactive Noninterference* // APLAS 2004, LNCS 3302, 2004. — P. 129–145.

On Access Control Mechanisms in Linux Operating System when Using Role-Based Security Policies

K. A. Shapchenko

1. Introduction

Access control is one of the major problems in modern operating systems. Solving such a problem can greatly improve the assurance level of information systems, especially when functioning in non-trusted environment [1].

Many software implementations of access control mechanisms exist, but only few of them are based on proper theoretical research. The present paper is focused on analysis of functional and architectural properties of access control implementations available for Linux kernel. We examine following Linux security extensions implementing enhanced access control mechanisms:

- Security Enhanced Linux (SELinux) [2];
- Rule-Set Based Access Control (RSBAC) [3];
- GRsecurity [4].

Role-based access control in these security extensions is implemented in SELinux/Type Enforcement (TE), RSBAC/Role Compatibility (RC) [5] and grsecurity/Role-Based Access Control (RBAC) models respectively.

Role-based access control models are chosen due to the fact, that role-based access control policies can be considered more intuitive and effective when applied to separation of different processes in modern computer systems.

Software implementations of role-based access control policies may be structurally complex and may consist of large amount of source code and configuration data. Besides, access control implementation may have auto-configuration facilities, e.g., learning system in GRsecurity or tools for audit journals analysis in SELinux. Generally, automatically

configured access control policies should not be used right in place without proper analysis ensuring its secure properties.

In distributed computer systems it is a common practice to use different information security tools across the entire system, and access control tools are no exception. When different access control implementations are used in a distributed computer system, the problem of comparing different access control models arises. Practically, the problem of reviewing the entire access control policy in heterogeneous distributed system is quite important, e.g., for conducting security audit. In such case, representing different policies in terms of one unified model is a convenient solution.

Combining all of the given facts, three types of mathematical problems can be stated.

1. Providing formal models for different software implementations of access control mechanisms.

In this paper we provide formal access control models only to solve some other problems, so this sort of problems is quite auxiliary in this paper. But practically, a formal description can lead to more effective software implementation and accurate configuration of access control policy. Thus, estimated assurance level cannot be improved without proper formal models of access control.

2. Expressing one access control model in terms of another.

Solutions of such problems can provide methods to compare different software implementations of access control in terms of possibility to express some constructs of access control policy. Besides, using a unified language able to express different access models can provide means to reviewing the entire access control policy in distributed computer system. Implementation of one-to-one correspondence between access control models (with some restrictions) of different software implementations is another benefit from solving this set of problems.

3. Verification of security properties of access control models based on the configuration of their software implementations.

On this direction we need to develop a formal language for specifying security properties of access control models and methods to verify these properties in a model of system behavior with given configuration of access control tools. In this paper we use linear temporal logic as a specification language. Thus, if an automata model of system behavior is given, then ordinary model checking tools, such as NuSMV2 [6], can be used to verify properties. Similar research was conducted in [8]. In this paper we pose a slightly extended model also aimed at verifying information flow properties.

It is significant to mention that by using the approaches to solving the previous class of problems we can verify secure properties of heterogeneous distributed computer systems with different access control implementations. To achieve this, only a transformation to a unified access control model.

2. Access control models

In this section we give a short description of examined access control extensions for Linux. For all of the three role-based access control mechanisms in these security extensions we provide a short description of its basic notions and a formal set-based model.

2.1. SELinux/Type Enforcement

Security Enhanced Linux (SELinux) [2] is an officially included in Linux 2.6 access control extension implementing role-based access control using Type Enforcement (TE) policy.

Type Enforcement policy is a combination of two access control models: Domain and Type Enforcement (DTE) and Role-Based Access Control (RBAC). DTE is used to separate subjects and objects into domains and types respectively and to determine allowed accesses from domains to types. RBAC is used to assign roles to users and to determine allowed domains for a role.

Access type in TE model is determined by a pair (class of the accessed object, access right). Object classes and permissions are defined in the SELinux kernel, but classes are also described in TE configuration to provide consistency with internal definitions.

In TE configuration file new class is defined with a command

```
class <class name>
```

For example:

```
class file
class dir
```

Implemented classes include:

- common classes: `security`, `process`, `system`, `capability`;
- classes of file systems objects: `ffilesystem`, `file`, `dir`, `fd`, `lnk_file`, `chr_file`, `blk_file`, `sock_file`, `fifo_file`;
- classes of network-related objects: `socket`, `tcp_socket`, `udp_socket`, `rawip_socket`, `netlink_socket`, `packet_socket`, `key_socket`, `unix_stream_socket`, `unix_dgram_socket`, `node`, `netif`;

- classes of interprocess communication objects: `sem`, `msg`, `msgq`, `shm`, `ipc`.

Permission for a given class can be defined in configuration with two commands. For defining a set of basic permissions the following command should be used:

```
common <set id> <permissions>
```

For example:

```
common file { ioctl read write create getattr setattr lock
relabelfrom relabelto append unlink link rename execute
swapon quotaon mounton }
```

In order to define permissions for a class inheriting a set of basic permissions we should use a `class` command:

```
class <class name> [inherits <set id>] <permissions>
```

For example:

```
class dir inherits file { add_name remove_name reparent
                          search rmdir }
class file inherits file { execute_no_trans entrypoint }
```

Special permission `entrypoint` determines whether a transition from the current domain to another domain should occur.

To determine types and domains a set of types and a set of attributes are introduced. Each attribute distinguishes a subset of types, e.g., attribute `domain` is used to separate domains from other types. Attribute is defined in configuration with a following command:

```
attribute <attribute name>;
```

Types are defined with a command

```
type <type name> <set of attributes>;
```

For example:

```
attribute domain;
type sshd_t, domain;
```

The type of a newly created object is determined with a following command:

```
type_transition <source types> <target types> :
        <class> <new type>;
```

Examples of using `type_transition` command:

```
type_transition sshd_t tmp_t : file sshd_tmp_t;
type_transition sshd_t shell_exec_t : process user_t;
```

The first command determines that new type any object with class `file` created in a directory of `tmp_t` type will be `sshd_tmp_t`. The second command defines new type of processes launched from a process with domain `sshd_t` and an executable file with type `shell_exec_t` will be `user_t`.

To determine allowed accesses from source type to target type the following command is used:

```
<audit type> <source types> <target types>:
             <classes> <permissions>;
```

where `<audit type>` is one of the following options:

- `allow` — only access violations are registered;
- `auditallow` — every access attempt is registered;
- `dontaudit` — access attempt is never registered.

Example:

```
allow sshd_t shell_exec_t : file { read execute
                                   entrypoint };
```

To forbid an access a `neverallow` command is used:

```
neverallow <source types> <target types>:
           <classes> <permissions>;
```

For example:

```
neverallow domain ~domain : process transition;
```

In this example any transition from a domain to a non-domain type is forbidden when creating a new process. The `neverallow` command is applied only at policy compilation stage and removes disallowed rules from the entire policy.

To define RBAC-related notions of the policy the `user` and `role` commands are used:


```
role <role name> types <allowed types (domains)>;
user <user name> roles <allowed roles>;
```

For example:

```
role user_r types user_t;
user guest_u roles user_r;
```

Allowed role transitions are defined as follows:

```
allow <role name> <allowed roles>;
```

In Type Enforcement model a notion of *security context* is defined. Security context is a triple (user, role, type). Every access attempt in TE model is considered as an access from source security context to target security context. A special role `object_r` is defined for non-process objects. Security contexts of existing file system objects should be defined by preliminary labeling of file systems.

In order to restrict access basing on two security contexts a mechanism of constrains is introduced with following syntax:

```
constrain <classes> <permissions> <expr>
```

Conditional expression `expr` with variables `u1`, `r1`, `t1`, `u2`, `r2`, `t2` is formally defined as follows:

```
expr ::= (expr) | not expr | expr and expr | expr or expr
         | u1 op u2 | t1 op t2 | r1 op r2
         | u1 op <user name> | t1 op <type name>
         | r1 op <role name>
         | u2 op <user name> | t2 op <type name>
         | r2 op <role name>

op   ::= == | !=
```

The rules and commands described above define the policy configuration processed by policy compiler. The set-based model of TE policy follows.

Let us define following finite non-empty sets:

- C — set of classes;
- P — set of permissions;
- $\Gamma \subset C \times P$ — set of correct pairs (class, permission);
- U — set of users;

- R — set of roles, $r_o \in R$ — the `object_r` role;
- T — set of types;
- $D \subset T$ — set of domains (types with attribute `domain`).

The following predicates can be defined in TE policy.

- $\mu(u, r), u \in U, r \in R$ — role r is allowed for user u with requirement $\forall u \in U \mu(u, r_o)$.
- $\rho(r, t), r \in R, t \in T$ — type t is allowed for role r with requirement $\forall t \in T \rho(r_o, t)$.
- $\alpha_\rho(r_1, r_2), r_1 \in R, r_2 \in R$ — role transition from r_1 to r_2 is allowed.
- $\alpha(t_1, t_2, c, p), t_1 \in T, t_2 \in T, c \in C, p \in P$ — subject with type t_1 is allowed to access object with type t_2 with access type (c, p).
- $\chi_{(c,p)}(u_1, r_1, t_1; u_2, r_2, t_2)$ — relation that defines `constrain` rules in configuration:
  - if $(c, p) \notin \Gamma$ then χ is false;
  - if $(c, p) \in \Gamma$ and there are no `constrain` rules for access type (c, p) in policy, then χ is true;
  - if $(c, p) \in \Gamma$ and there are `constrain` rules for access type (c, p) then predicate χ is true iff its parameters are members of a set defined in corresponding `constrain` rule.

In given notation we can define the predicate for granting access from security context (u_1, r_1, t_1) to (u_2, r_2, t_2) with access type (c, p):

$$\Delta_{(c,p)}(u_1, r_1, t_1; u_2, r_2, t_2) = \alpha(t_1, t_2, c, p) \wedge \rho(r_1, t_1) \wedge \rho(r_2, t_2) \wedge \mu(u_1, t_1) \wedge \mu(u_2, t_2) \wedge \chi_{(c,p)}(u_1, r_1, t_1; u_2, r_2, t_2) \wedge ((c, p) = (process, transition) \implies \alpha_\rho(r_1, r_2))$$

2.2. RSBAC/Role Compatibility

Rule-Set Based Access Control (RSBAC) [3] is an open-source implementation of several access control models for Linux kernel. RSBAC is based on Generalized Framework for Access Control (GFAC).

RSBAC consists of several modules implementing different models of access control, such as Mandatory Access Control (MAC) — multilevel security policy, Role Compatibility (RC) [5] — variation of a role-based control model, Access Control Lists (ACL) — discretionary access control policy, Privacy Model (PM) — module for implementing confidentiality-aimed policies and some other modules.

RSBAC does not provide textual configuration files with strict syntax and semantics like in SELinux, but a formal set-based access control model can be built using the description of RC module and provided administration software.

In order to define a Role Compatibility model notions of users, roles and object types should be defined. Let us denote the set of users as U, the set of roles as R. For every user $u \in U$ a set of allowed roles $AR(u) \subset R$ is defined. Let us denote the default role for the user u as $DR(u) \in R$, $DR(u) \in AR(u)$.

Let us denote the set of object types as T,

$$T = FDTypes \sqcup DevTypes \sqcup ProcessTypes \sqcup IPCTypes \sqcup NetTypes \sqcup SCDTypes \sqcup UserTypes.$$

The sets in the disjunctive union contain types of ordinary files/directories, device files, processes, interprocess communication objects, network objects, system objects (e.g., timer) and user administration system objects respectively. Such separation of objects is similar to classes in SELinux/TE, but SELinux classes are more detailed.

For each set of types a set of permissions is defined:

$$FDPermissions, DevPermissions, ProcessPermissions, IPCPermissions, NetPermissions, SCDPermissions, UserPermissions$$

respectively. Similar to SELinux/TE case, it is convenient to introduce set P as union of all the permission sets.

For example, according to RSBAC documentation:

$FDPermissions = \{$*append_open, change_owner, chdir, close, create, delete, execute, get_perm_data, get_stat_data, link_hard, modify_access_data, modify_attribute, mount, read, read_attribute, read_write_open, read_open, rename, search, truncate, umount, write, write_open, map_exec*$\}$.

This set of permission slightly differs from permissions for `file` and `dir` classes in SELinux/TE.

For each role $r \in R$ a set of compatible roles $CR(r) \subset R$ is defined in RC policy similar to role transitions in SELinux/TE.

For each role $r \in R$ a set of compatible types $CT(r) \subset T \times P$ is define similar to allowed types in SELinux/TE.

Let S be a set of subjects (processes) and O — a set of objects, $S \subset O$. Each $o \in O$ has an owner $user(o) \in U$ and object type $type(o) \in T$, $type(o)$ is a member of corresponding type set, e.g., $type(s) \in ProcessTypes$. For each subject $s \in S$ its role is defined as $role(s) \in R$. For each object $o \in O$ a forced role $FR(o) \in R \sqcup \{role\_inherit\_user, role\_inherit\_process\}$ is defined.

Let us consider that $FR(o) = role\_inherit\_process$ holds for every object $o \in O \setminus \{o : type(o) \in FDTypes\}$, because forced roles have no

meaning for non-executable objects. A forced role is similar to entry points in SELinux/TE.

Each subject should be valid, that is the following predicate is true: $valid(s) = (role(s) \in AR(user(s)))$.

In given notation we can define predicate of granting access $p \in P$ of subject $s \in S$ to object $o \in O$:

$$\Delta(s,o,p) = (type(o),p) \in CT(role(s)) \wedge (p = execute \implies$$
$$\implies (FR(o) \in CR(role(s)) \vee FR(o) = role\_inherit\_process \vee$$
$$\vee (FR(o) = role\_inherit\_user \wedge DR(user(s)) \in CR(role(s)))).$$

The right part of implication denotes conditions to allow role transition with creation of a new process.

RC model has an analogue to `type_transition` rules from SELinux/TE — default types of a newly created object in role $r \in R$: $DefFDType(r) \in FDTypes$, $DefProcessType(r) \in ProcessTypes$, $DefIPCType(r) \in IPCTypes$ and so on.

Set-based model constructed above does not cover entire RC implementation in RSBAC, which also includes changing roles in setuid() system calls and separation of administrative duties, which have no analogues in SELinux.

2.3. GRsecurity/Role-Based Access Control

GRsecurity [4] is a modern security extension for Linux kernel supporting role-based access control with learning configuration system. Access control model used in GRsecurity is simpler when compared to Type Enforcement and Role Compatibility models. It lacks role entry points like in RC and TE models, roles are changed with special software tool. Besides, there is no separation of objects into classes or types, controlling access to IPC or network objects is not supported.

GRsecurity configuration and access control model can be formally constructed as follows.

The entire policy is a sequence of role definitions including a mandatory `default` role. Every role definition has following syntax:

```
role <role name> <role flags>
[role_transitions <allowed roles>]
<list of subject definitions>
```

Role flags are a set of internal GRsecurity parameters of a role which do not directly affect access of subjects to objects. E.g., `'A'` denotes an administrative role.

The `role_transitions` command defines allowed role to make a transition from current role.

Permission of each subject (process) is determined by a file which was executed to create a process. Besides, GRsecurity policy introduces notion of nested subjects to control execution paths and change permissions accordingly.

Subject is defined in configuration with following syntax:

```
subject { <object> or <nested subject> } <subject flags>
<list of object definitions>
<list of capabilities>
```

Here, nested subject denotes a string

```
<executable object>:<executable object>:...:
                                    <executable object>
```

Inheritance of permissions in a directory subtree is provided by defining object as a directory and setting `'i'` subject flag. Other subject flags control tracing of a process and using PaX — GRsecurity's subsystem of preventing execution of arbitrary code. So, these subject flags do not directly affect access to objects similar to role flags.

Each object is defined as follows:

```
<path to object> <object flags>
```

Object flags denote allowed permissions:

- `'r'` — permission to read;
- `'w'` — permission to write or append;
- `'a'` — permission to append only;
- `'c'` — permission to create files in a directory;
- `'d'` — permission to remove files in a directory;
- `'x'` — permission to execute/search in a directory;
- `'l'` — permission to create links;
- `'m'` — permission to create setuid/setgid files;
- `'i'` — inheritance flag similar to the same subject flag.

Permissions in GRsecurity access control policy are not as detailed as in TE or RC models. So, fine-grained detailed policies cannot be expressed in GRsecurity access control model.

Processes owned by user 'root' have the first administrative role by default. Ordinary users have appropriate user, group, or `default` role.

Let us construct a formal model of GRsecurity access control basing on the configuration language described above.

Let O be a finite non-empty set of objects. Objects can be considered file system objects. Set of executable objects is denoted as $E \subset O$.

Let R be a set of roles and A — a set of access rights, which are finite and non-empty. $A = \{r, w, a, c, d, x, l, m\}$ according to object flags.

Predicate $\rho(r_1, r_2)$ determines allowed role transitions. Let $\hat{\rho}(r_1, r_2)$ be a transitive closure of $\rho(r_1, r_2)$ as a relation on $R \times R$.

A set of subjects can be defined as a finite set of pairs (role, finite word in alphabet E): $S = R \times E^*$. Empty word ε in this case denotes 'empty' subject, which exists right after entering a new role.

Let U be a finite non-empty set of users. Let us denote $DR(u)$ as a default user role for user $u \in U$.

Policy configuration includes a directly specified predicate of granting access: $\Delta(s, o, a)$, $a \in A$, $s = (r, e) \in S$, $r \in R$, $o \in O$. This predicate can be extended from its domain $S \times O \times A$ to $(R \times E^*) \times O \times A$ if we assume $\Delta(s, o, a)$ false for all $s \in (R \times E^*) \setminus S$. Predicate $\Delta(s, o, a)$ is defined for all subjects, but not every subject can exist in a given role $r \in R$. Subject exists only if all components of its executable object path are sequentially executed, this can be expressed formally as $(s = \varepsilon) \vee (s = e_1 e_2 \dots e_n \in E^* \implies (\Delta((r, \varepsilon), e_1, x)) \wedge (\forall i = 1, \dots, n-1 \implies \Delta((r, e_1 \dots e_i), e_{i+1}, x)))$.

3. Comparison of access control models

In this section we give approaches to solving two model problems connected with converting access control policies from one access control model to another. The obtained results can be used to compare 'expressiveness' of access control models.

Here we understand 'expressiveness' as ability to simulate given configuration of access control mechanisms in one security subsystem or access control model in terms of another model with respect to equivalence of allowed accesses. Expressiveness is a relational notion, that is, only statements like "one model is no more expressive than the other" can be made. Such results can be used also in unification and integration of different access control mechanisms. For example, two theoretical results can be used due to its constructive proofs in providing SELinux/TE model as a unification base for all of the three examined Linux security extensions with role-based access control models.

Let us pose a general problem as follows. A policy in source model (RSBAC/RC or grsecurity/RBAC) is given with some restrictions. A policy in target model (SELinux/TE) should be constructed and the constructed policy must be equivalent to the source policy in terms of allowed accesses.

3.1. Expressing RSBAC/RC policy in SELinux/TE

Restrictions in this particular problem can be stated as follows.

- Only non-administrative RC policies are considered. This restriction is due to the mandatory origin of compared policies.
- There is a one-to-one correspondence between access types in policies, that is, policies are compared only using common set of access types.
- Users, subjects and objects are identical in compared policies.
- There are no objects in RC policies with forced role inherited from user. This is the only restriction weakening the model.

With given source policy satisfying the above restrictions we need to construct SELinux/TE policy with equivalent access granting. To separate notions in TE and RC models we will use prefixes "$RC.$" and "$TE.$" respectively.

Construction of TE model sets and predicates follows.

$TE.U := RC.U$,
$TE.R := \{u_i\_role | u_i \in RC.U\} \sqcup \{r_o\}$,
$TE.D := RC.R \times RC.ProcessTypes$,
$TE.T := TE.D \sqcup (RC.T \setminus RC.ProcessTypes)$,
$(TE.\mu(u_i, r) := (r = u_i\_role \vee r = r_o)$, $u_i \in TE.U$, $r \in TE.R$,
$(TE.\rho(u_i\_role, t) := t \in RC.ProcessTypes \wedge$
$(t, CREATE) \in RC.CT(u_i\_role))$, $u_i \in RC.U$, $t \in RC.T$,
$TE.\alpha(t_1, t_2, c, p) := (t_1 = (r_1, \tau_1) \in TE.D \wedge (t_2 \notin TE.D \implies$
$(class(t_2) = c \wedge (t_2, p) \in RC.CT(r_1))) \wedge$
$(t_2 = (r_2, \tau_2) \in TE.D \implies$
$((c = process \wedge (\tau_2, p) \in RC.CT(r_2)) \vee$
$(c = process \wedge p = transition \wedge r_2 \in RC.CR(r_1)))))$,
$t_1, t_2 \in TE.T$, $c \in TE.C$, $p \in TE.P$,

where $class : RC.T \to TE.C$ is a mapping from object types in RC model to classes of objects in TE model.

$TE.\alpha_\rho(r_1, r_2) := (r_1 = r_2)$,
$\forall c \in TE.C, p \in TE.P, u_i \in TE.U, r_i \in TE.R, t_i \in TE.T,$
$i = 1, 2 : (TE.\chi_{c,p}(u_1, r_1, t_1; u_2, r_2, t_2))$.

In order to express one policy in terms of another access control model we need to show equivalence of access granting predicates in two compared models.

In predicate $TE.\Delta$ expression

$TE.\mu(u_1, r_1) \wedge TE.\mu(u_2, r_2) \wedge ((c, p) = (process, transition) \implies TE.\alpha_\rho(r_1, r_2))$

connects users and artificial roles. Expression $TE.\rho(r_1, t_1) \wedge TE.\rho(r_2, t_2)$ connects artificial roles to domains corresponding to original roles from $RC.R$ and process types from $RC.ProcessTypes$. There are no constrains in TE model, so $TE.\chi$ is always true. Expression $TE.\alpha(t_1, t_2, c, p)$ denotes allowed role transitions in RC model.

So, the constructed policy is equivalent to the source policy in terms of allowed accesses and given restrictions.

3.2. Expressing GRsecurity/RBAC policy in SELinux/TE

The problem posed in this section slightly differs from the previous one. In this problem GRsecurity/RBAC is taken as a source policy and there are fewer restrictions compared to previous problem.

The following restrictions hold in this problem.

- There is a one-to-one correspondence between access types in policies, that is, policies are compared only using common set of access types.
- Users, subjects and objects are identical in compared policies.

Now we need to express GRsecurity/RBAC in SELinux/TE model. Let us denote sets and predicates in these models with prefixes "$GR.$" and "$TE.$" respectively.

Construction of the target SELinux/TE policy will be carried out similar to previous problem.

Let us define sets and predicates in SELinux/TE model:

$TE.U := GR.U,$
$TE.R := GR.R \sqcup \{r_o\},$
$TE.D := GR.S,$
$TE.T := TE.D \sqcup GR.O.$
$TE.\mu(u, r) := (GR.\hat{\rho}(GR.DR(u), r) \vee r = r_o),$
$TE.\rho(r, t) := (t = (r, e) \in GR.S \vee r = r_o),$
$TE.\alpha(t_1, t_2, c, p) := (t_1 = (r_1, e_1) \in TE.D \wedge ((t_2 \in TE.T \setminus TE.D \wedge GR.\Delta(t_1, t_2, access(c, p))) \vee (t_2 = (r_2, e_2) \in TE.D \wedge c = process \wedge p = transition \wedge (\exists e \in GR.E : e_2 = e_1 \cdot e)))),$

where $access(c, p) : TE.C \times TE.P \to GR.A$ maps access types from SELinux/TE policy to grsecurity/RBAC, and '$\cdot$' is a concatenation operator.

$TE.\alpha_\rho(r_1, r_2) := (r_1 = r_2),$

$\forall c \in TE.C, p \in TE.P, u_i \in TE.U, r_i \in TE.R, t_i \in TE.T,$
$i = 1, 2 : (TE.\chi_{c,p}(u_1, r_1, t_1; u_2, r_2, t_2)).$

Let us show the equivalence of access granting predicates in source and constructed models.

- Firstly, in predicate $TE.\Delta$ expression $TE.\mu(u_1, r_1) \wedge TE.\mu(u_2, r_2) \wedge ((c, p) = (process, transition) \implies TE.\alpha_\rho(r_1, r_2))$ connects users and roles similar to their connection in original policy.
- Secondly, expression $TE.\rho(r_1, t_1) \wedge TE.\rho(r_2, t_2)$ connects roles to subjects.
- There are no constrains in TE model, so $TE.\chi$ is always true.
- Finally, expression $TE.\alpha(t_1, t_2, c, p)$ directly simulates allowed accesses in GRsecurity/RBAC and domain transitions to implement nested subjects from grsecurity/RBAC.

So, constructed SELinux/TE policy successfully simulates the original GRsecurity/RBAC policy.

4. Specifying and verifying information flow properties

In order to examine security properties of TE policy we provide state machine-based model with linear temporal logic (LTL) formulae as a specification language. Similar research was carried out in [8]. We provide a slightly extended version of the model described in [8].

4.1. Model description

Let us simplify TE model as follows. Instead of the set $\Gamma \subset C \times P$ we used simplified set of access types

$$A = \{read, write, transition\} \sqcup \{read\_by, written\_by\},$$

where read_by, written_by are additional 'reverse' accesses from types to domains, which are allowed only if the corresponding direct accesses are allowed from domains to types.

The set of states is defined as

$$S = U \times R \times T \times U \times R \times T \times A \times \{true, false\},$$

the last component of the Cartesian product denotes that all states in some path satisfies the access granting predicate. A state denotes an allowed access from one security context to another.

Let us define the transition relation $\tau \subset S \times S$:

$$((u_1, r_1, t_1, u_2, r_2, t_2, a, f), (u_1', r_1', t_1', u_2', r_2', t_2', a', f')) \in \tau \iff$$
$$(\Delta_a(u_1, r_1, t_1, u_2, r_2, t_2) \wedge \Delta_{a'}(u_1', r_1', t_1', u_2', r_2', t_2') \wedge$$
$$u_1' = u_2 \wedge r_1' = r_2 \wedge t_1' = t_2 \wedge f' = f) \vee f' = false.$$

Initial states set $\iota \subset S$ is defined as follows:

$$\iota = \{(u_1, r_1, t_1, u_2, r_2, t_2, a, f) \in S : \Delta_a(u_1, r_1, t_1, u_2, r_2, t_2 \wedge f = true\}.$$

LTL formulae include:

- all propositional formulae with state variables or other LTL formulae;
- "`X` f" formulae, where f is a LTL formula, `X` is a LTL operator, such construction implies that f is true in all states next to current state;
- "`G` f" formulae, where f is a LTL formula, `G` is a LTL operator, f is true in all states on every path starting in current state;
- "`F` f" formulae, where f is a LTL formula, `F` is a LTL operator, there exists a state on every path from current state in which f is true;
- "f `U` g" formulae, where f,g are LTL formulae, `U` is a LTL operator, f is true in all paths from current state until a state in which g is true and such state always exists;
- "f `V` g" formulae, where f,g are LTL formulae, `V` is a LTL operator, on every path starting in the current state g is true in all states before a state in which f is true is such state exists.

Described language of LTL formulae can be used to express security properties, three examples of such properties in SELinux/TE policy follow.

- Allowed information flow from type (`src_t`) to type `dst_t`:

$$(t_1 = \texttt{src\_t}) \wedge ((a = read\_by \vee a = write)\ \texttt{U}$$
$$(t_2 = \texttt{dst\_t} \wedge a = write \wedge f)).$$

- Access to `domain2` is allowed only from `domain1` and only through `domain3`:

$$\neg(t_1 = \texttt{domain1} \wedge (a = transition\ \texttt{U}\ (a = transition \wedge f \wedge t_2 = \texttt{domain2})) \wedge \neg(t_1 = \texttt{domain1} \wedge a = transition \wedge t_2 = \texttt{domain3} \wedge \texttt{X}\ (t_1 = \texttt{domain3} \wedge a = transition \wedge t_2 = \texttt{domain2} \wedge f))).$$

- No user acting in a role `user_r` is allowed to access data of type `system_data_t` for writing after domain changes

$$(r_1 = \texttt{user\_r}) \wedge ((a = transition)\ \texttt{U}$$
$$(t_2 = \texttt{system\_data\_t} \wedge a = write \wedge f)).$$

4.2. Example of verifying a security property

Let us examine a simple example of using model checking tool NuSMV2 ([6], [7]) for verifying security properties of SELinux/TE policy.

The example access control configuration in simplified SELinux/TE configuration language follows.

```
user system_u
role system_r
user_roles system_u system_r
role_types system_r server1_process_t server2_process_t
type server1_process_t
type server2_process_t
type server1_data_t
type server2_data_t
type shared_data_t
allow server1_process_t server1_data_t read
allow server1_process_t server1_data_t write
allow server1_process_t shared_data_t read
allow server2_process_t server2_data_t read
allow server2_process_t server2_data_t write
allow server2_process_t shared_data_t read
```

In this example configuration a system with two processes of domains `server1_process_t` and `server2_process_t` respectively. Each of them has full read/write access to data of types `server1_data_t` and `server2_data_t` respectively. Also data of type `shared_data_t` are accessible to processes of both types.

Disallowed information flow from data of type `server1_data_t` to data of type `server2_data_t` can be specified with a formula

$$\neg((t_1 = \texttt{server1\_data\_t}) \wedge ((a = read\_by \vee a = write)\ \texttt{U}$$
$$(t_2 = \texttt{server2\_data\_t} \wedge a = write \wedge f))).$$

In the example above it is obvious that this formula is true.

Description of the corresponding state machine and specification in NuSMV syntax follows.

```
MODULE main

VAR
 _u1 : { system_u };
 _r1 : { _object_role, system_r };
 _t1 : { server2_data_t, server1_data_t, server2_process_t,
          server1_process_t, shared_data_t };
 _u2 : { system_u };
 _r2 : { _object_role, system_r };
 _t2 : { server2_data_t, server1_data_t, server2_process_t,
         server1_process_t, shared_data_t };
 _a  : { _read, _write, _read_by, _written_by,
         _transition };
 _flow : boolean;

DEFINE
    _user_role_1 := case
        _u1=system_u : _r1 in {system_r};
        TRUE : FALSE;
    esac;

    _user_role_2 := case
        _u2=system_u : _r2 in {system_r};
        TRUE : FALSE;
    esac;

    _role_type_1 := case
        _r1=system_r : _t1 in {server1_process_t,
                               server2_process_t};
        TRUE : FALSE;
    esac;

    _role_type_2 := case
        _r2=system_r : _t2 in {server1_process_t,
                               server2_process_t};
        TRUE : FALSE;
    esac;

    _role_transition := case
        _r1=system_r : FALSE;
        TRUE : FALSE;
```



```
    esac;

    _allow := case
        _t1=shared_data_t & _t2=server2_process_t &
            _a=_read_by : TRUE;
        _t1=server1_process_t & _t2=shared_data_t &
            _a=_read : TRUE;
        _t1=server2_process_t & _t2=shared_data_t &
            _a=_read : TRUE;
        _t1=server1_data_t & _t2=server1_process_t &
            _a=_read_by : TRUE;
        _t1=server1_data_t & _t2=server1_process_t &
            _a=_written_by : TRUE;
        _t1=server2_data_t & _t2=server2_process_t &
            _a=_written_by : TRUE;
        _t1=server2_data_t & _t2=server2_process_t &
            _a=_read_by : TRUE;
        _t1=server1_process_t & _t2=server1_data_t &
            _a=_write : TRUE;
        _t1=server1_process_t & _t2=server1_data_t &
            _a=_read : TRUE;
        _t1=shared_data_t & _t2=server1_process_t &
            _a=_read_by : TRUE;
        _t1=server2_process_t & _t2=server2_data_t &
            _a=_read : TRUE;
        _t1=server2_process_t & _t2=server2_data_t &
            _a=_write : TRUE;
        TRUE : FALSE;
    esac;

    _constrain := case
        TRUE : TRUE;
    esac;

    _trans :=
        (_role_type_1 | _r1=_object_role) &
        (_role_type_2 | _r2=_object_role) &
        (_user_role_1 | _r1=_object_role) &
        (_user_role_2 | _r2=_object_role) &
        _allow & _constrain &
        (_a=_transition -> _role_transition);
```

```
INIT
    _trans & _flow;

TRANS
    (_trans & next(_trans) &
    next(_u1) = _u2 & next(_r1) = _r2 & next(_t1) = _t2 &
    next(_flow) = _flow) | (next(_flow) = FALSE);

LTLSPEC
    ! ((_t1=server1_data_t) & ((_a=_read_by | _a=_write)
    U (_t2=server2_data_t & _a=_write & _flow)) );
```

In this example in `VAR` section a set of states is defined, in sections `INIT` and `TRANS` set of initial states and a transition relation are defined. In `LTLSPEC` section the specification is defined.

In the given example applying NuSMV confirms that specification holds in the model:

```
-- specification !(_t1 = server1_data_t & ((_a = _read_by |
_a = _write) U ((_t2 = server2_data_t & _a = _write) &
_flow))) is true
```

Let us change the access control configuration and add a following string:

```
allow server1_process_t shared_data_t write
```

In this case the specification is no longer valid and a counterexample can be built automatically with NuSMV. The counterexample includes a path of state transitions which does not satisfy the specification.

```
-- specification !(_t1 = server1_data_t & ((_a = _read_by |
_a = _write) U ((_t2 = server2_data_t & _a = _write) &
_flow))) is false
-- as demonstrated by the following execution sequence
Trace Description: LTL Counterexample
Trace Type: Counterexample
-- Loop starts here
-> State: 1.1 <-
  _r1 = _object_role
  _t1 = server1_data_t
  _r2 = system_r
  _t2 = server1_process_t
  _a = _read_by
  _flow = 1
  _u1 = system_u
  _u2 = system_u
-> State: 1.2 <-
  _r1 = system_r
  _t1 = server1_process_t
  _r2 = _object_role
  _t2 = shared_data_t
  _a = _write
  _flow = 1
  _u1 = system_u
  _u2 = system_u
-> State: 1.3 <-
  _r1 = _object_role
  _t1 = shared_data_t
  _r2 = system_r
  _t2 = server2_process_t
  _a = _read_by
  _flow = 1
  _u1 = system_u
  _u2 = system_u
-> State: 1.4 <-
  _r1 = system_r
  _t1 = server2_process_t
  _r2 = _object_role
  _t2 = server2_data_t
  _a = _write
  _flow = 1
  _u1 = system_u
  _u2 = system_u
```

In this execution trace there is a path implementing an information flow from data of `server1_data_t` type to data of `server2_data_t` type via data of `shared_data_t` type. The examined example is trivial but it shows how security properties can be verified with ordinary model checking tools.

5. Conclusion

The results on formal set-based models of access control obtained in this paper and on simulating one model in terms of another provide methods of comparing access control implementations to choose more appropriate implementations for specific tasks.

A state machine and LTL-based method of verifying security properties of access control implementations is provided by using ordinary model checking tools. This method can be used for finding weak places in access control implementation, e.g., in constructing distributives of secure operating systems where role-based access control can be used for isolation of critical processes, but large amount of configuration data make verification of security properties by hand nearly impossible.

References

[1] Vasenin V. A. Problems of mathematical, algorhithmic and software means for enforcement of information security in the Internet. Materials of MaBIT-03, 2004.

[2] National Security Agency. Security-enhanced Linux. `http://www.nsa.gov/selinux`.

[3] Rule-Set Based Access Control. `http://www.rsbac.org`.

[4] grsecurity. An Innovative Approach to Security. `http://www.grsecurity.net`.

[5] Amon Ott, The Role Compatibility Security Model, Nordic Workshop on Secure IT Systems (NordSec) 2002.

[6] NuSMV2 — A New Symbolic Model Checker. `http://nusmv.irst.itc.it`.

[7] Cimatti A., and Clarke E., Giunchiglia E., Giunchiglia F., Pistore M., Roveri M., Sebastiani R., Tacchella A. NuSMV Version 2: An OpenSource Tool for Symbolic Model Checking, Proc. International Conference on Computer-Aided Verification (CAV 2002), July 2002, LNSC, vol. 2404.

[8] Guttman Joshua D., Herzog Amy L., Ramsdell John D., Skorupka Clement W. Verifying information flow goals in Security-Enhanced Linux. Journal of Computer Security, vol. 13, num. 1, 2005, p. 115–134.

Architecture and Models for Security Policy Verification

I. V. Kotenko, A. V. Tishkov, O. V. Chervatuk

1. Introduction

Policy-based security management of computer networks is one of the most actual directions of research in information security area. At present the IETF [1] recommendations are commonly accepted standard for the architecture of policy-based management systems. According to these recommendations such architecture should contain the centralized repository of policy rules for entire system, thus making the policy available for analysis and verification. This paper elaborates the architecture and models of security policy verification system — SEcurity Checker (SEC) — originally suggested in [2] and implemented corresponding to the IETF recommendations.

In the paper the improved architecture of Security Checker is considered and the mechanisms of operating with the policies of three levels are described: (1) upper-level, that is approximated to the user requirement language, (2) intermediate level, classifying rules according to several categories, and (3) low-level, describing the policy in the format of Common Information Model (CIM). The approach to the design and implementation of SEC kernel is given. An example of authorization policy conflicts emulation and detection is suggested. The relevant works are analyzed.

2. Improved architecture of Security Checker

In SEC the policy description language has three levels: upper, intermediate, and low (Fig. 1).

The research is supported by the "Fund for support of national science", grant of Russian Foundation of Basic Research (No. 04-01-00167), grant of the Department for Informational Technologies and Computation Systems of the Russian Academy of Sciences (contract No. 3.2/03) and partly funded by the EC as part of the POSITIF project (contract IST-2002-002314).

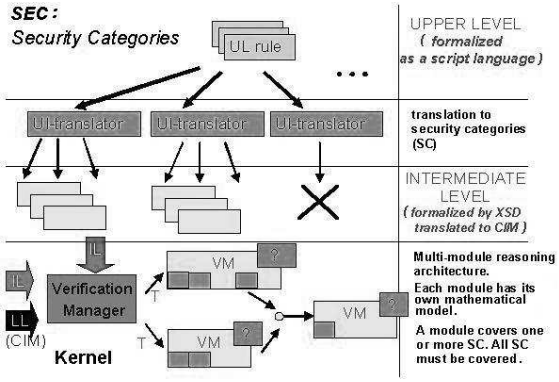

Figure 1. Generalized SEC architecture

Upper-level language (UL) describes the problem from generalized point of view. Formulations allow mentioning the groups of devices and the types of applications ("subnet S should not be accessible from host H by protocol P"). For specification of upper-level policies a scripting language is used as well as the set of translators from upper (U) level to the intermediate (I) one (UI-translators).

Upper-level rules are translated to the intermediate level (specified in *intermediate level language (IL)*) into the one of six categories of policy rules: authentication, authorization, filtering, confidentiality, operation rules, and vulnerability assessment rules. For each mentioned category an UI-translator is created. UI-translator receives upper-level rule as input and gives XML documents as output. These XML documents are valid according to XML-schema of corresponding category.

One of non-trivial translators from upper level to the intermediate one is UI-translator that defines filtering policy. In this task context the nodes of computer network are divided into two types: filtering and non-filtering (see Fig. 2).

When a policy specifying non-filtering node is set, a task of using filtering nodes for granting or denying access to protected node is solved on the graph representing the network topology. This task is solved as a one about minimal graph cut. Fig. 2 gives an example of creating four filtering rules by UI-translator, when upper level policy requires prohibiting the access between non-filtering nodes.

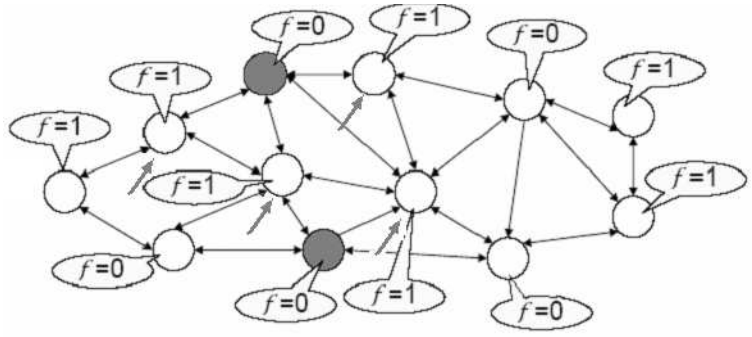

Figure 2. Filtering ($f = 1$) and non-filtering ($f = 0$) nodes

At extending of upper-level language with new constructions, a set of new UI-translators should be uploaded to the system for each category involved into such extending. Only those extensions are allowed, that do not change existing sublanguage. Thus, the SEC architecture is implemented as open for interpreting rules of other languages, such as Ponder [3] and other user-defined languages.

Finally, *low-level language (LL)* is a translation of intermediate level rules to object-oriented format of Common Information Model (CIM).

Structure of SEC kernel contains two types of basic elements: verification manager and verification module (VM).

Each verification module has its own knowledge base (as axiomatics, temporal logics formulae, action semi-lattices and others) and implements its own algorithm for checking policies consistency and applicability to given system description. Besides that, each module declares security categories with which it works.

Verification manager, getting intermediate and low-level policies as input, calls verification modules in parallel or subsequently. Parallel verification is possible only for modules that do not change the set of rules. Modules, that delete, change or add rules, are launched subsequently, getting at input a policy that is potentially changed by preceding modules. Such algorithm of kernel processing implies iterative calling of modules sequence. Iterations continue until the set of rules stops changing or until stop condition executes, in simplest case — by the explicit limitation of iterations number.

3. Security categories

As it was mentioned above, the intermediate level language is based on XML schemas for six categories of rules.

Authentication rule contains subjects (roles and users), objects (services defined on system description language [1]), actions that can be performed on the services, authentication method and security level, which the rule is associated with. The authentication method is defined by classes which are derived from CIM-class AuthenticationCondition. These classes are as follows: SharedSecretAuthentication, AccountAuthentication, BiometricAuthentication, NetworkingIDAuthentication, PublicPrivateKeyAuthentication, KerberosAuthentication, DocumentAuthentication, PhysicalCredentialAuthentication. All rules are accompanied with security level label. Security system can switch from one security level to another if, for example, the attack is detected.

Authorization rule is formulated as if-then rule. The conditional part contains quantifier-free predicate formula using NOT, AND, and OR logical operations. Atoms are the definitions of subject, object, action, security level, and the condition of system state. System state is described by the current state of services (run, stopped, waiting, busy), the results of authorization and authentication rules enforcement (the subject is authorized/authenticated for performing the action on the object), and user-defined system state conditions. The main used CIM-classes are as follows: Policy, AuthorizedSubject, AuthorizedObject, AuthorizedPriviledge, ComputerSystem, Role, and Identity.

Filtering rule represents commonly used access control list, each row of which consists of source address and port, destination address and port, deny/allow privilege and, additionally, security level. The used CIM-classes are Policy, FilterList, and FilterEntry.

Confidentiality rules are currently considered only for communication security, and define security protocols for data channels. The corresponding XML schema supports SSL or IPSec protocol. The main used CIM-classes are Policy, IPSecRule, and SSLRule.

Operational rules are specified by system state condition and actions, which should be performed on objects when system state matches the condition of a rule. The corresponding XML schema contains components for definition of network services installed on hosts, and actions which can be performed on those services. The CIM-class Policy is used, and three classes are added to CIM policy class hierarchy. They are OperationalRule, StatusCondition, and OperationalAction.

Vulnerability assessment rules are created by use of vulnerability database [4]. The rule contains vulnerability ID, reference to exploit, name and version of vulnerable software, information about patch/update that eliminates the vulnerability, and some additional information [5].

4. Kernel implementation

Basic SEC kernel classes are verification manager and verification module.

Verification manager (VerificationManager) gives to verification modules the system specification (in system description language) and fragments of policy specifications, according to security categories, for which the verification module is responsible. Besides that in suggested representation the manager gives out information about verification results, information about contradictions, if they appeared, and achieved security level. This class implements design pattern “singleton” [6], because verification manager should be only one in the system.

UML-representation for verification manager is given in Fig. 3. For each public field the existence of set value and get value functions is supposed.

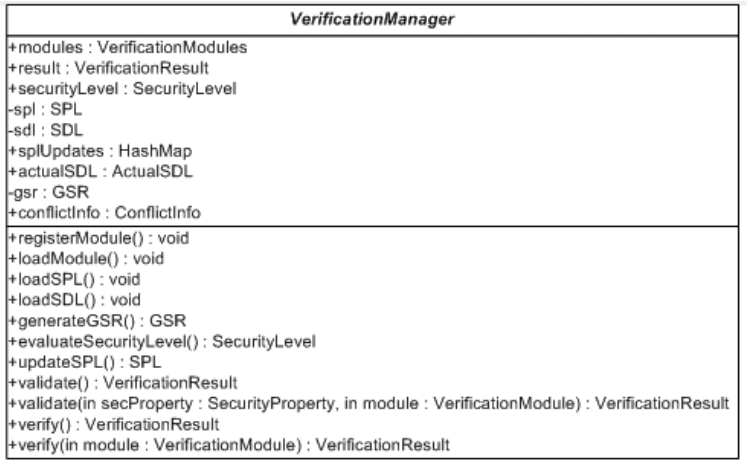

Figure 3. VerificationManager class

In this paper let us consider only several main fields and methods of class VerificationManager:

- Field *HashMap splUpdates* contains references to objects SPLUpdates created in each module. Objects SPLUpdates store list of changes that are necessary to be applied to rules set of policy for resolution of conflicts that were revealed during verification.
- Field *ActualSDL actualSDL* is revised network topology, in which some services are blocked by policies. ActualSDL contains list of blocked services.
- Field *ConflictInfo conflictInfo* contains information about conflicts revealed in the process of validation and verification.
- Method *updateSPL*() implements rules sets changing, proposed by modules.
- Method *validate*() without parameters checks rules for each security category using all registered and loaded modules that are responsible for this security category.
- Method *validate*() with parameters performs detecting and resolving rules conflicts within one security category. Security category and module that performs checking are passed as method parameters.
- Method *verify*() checks consistency of entire rules set and their applicability to the given systems description using special module.

Verification module VerificationModule (Fig. 4) performs validation and verification of categories rules SecurityProperty, for which it is responsible and which are listed in corresponding field.

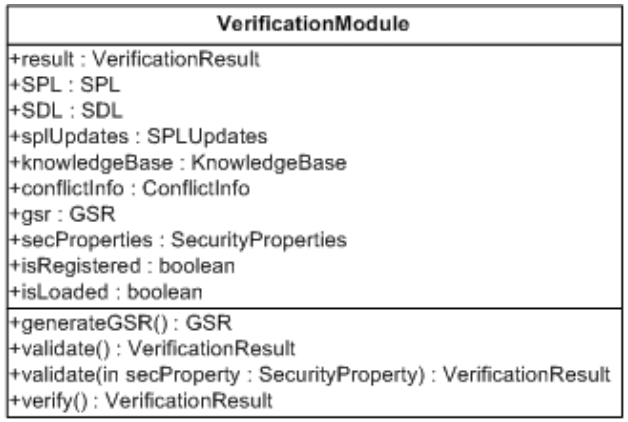

Figure 4. VerificationModule class

Main methods of class VerificationModule are validate() and verify(). Through these methods class VerificationModule delegates corresponding functionality to class VerificationModule.

5. Example of conflict detection

At current implementation of three verification modules is being done: (1) based on Event Calculus [7], (2) based on Model Checking [8], and (3) by creating the semi-lattices of actions.

Let us describe a simple example of modeling and detection of authorization conflict implemented by SPIN models checker [9].

Authorization conflict appears in the case when one user is attached to two roles R1 and R2 that have contradictory privileges for the same action: for one role there is permission, and for the other there is prohibition.

Key blocks of the program are two processes. The first process appoints and deletes belonging of a user to one of two roles (R1 or R2) at random. The following code corresponds to assigning a user to a role:

```
active proctype userRoleAssignment()
{
...
    :: (r.q<max_q_roles-1)->
        atomic {
            r.q++;
            if
            ::r.ar[r.q]=R1;
            ::r.ar[r.q]=R2;
            fi
            }
...
```

The second process models print requests, sent by user at random moments. Procedure IsAssigned checks user's belonging to the given role. The following code, receiving print request, assigns true value to variable deny (if the user at current belongs to role R1), or variable allow (if it belongs to role R2):

```
::printer_in?action,rr-> atomic
            {
            deny=false;
            allow=false
```

```
            IsAssigned(rr,R1,R1Res);
            IsAssigned (rr,R2,R2Res);
            if
            ::R1Res->deny=true
            ::else
            fi
            if
            ::R2->allow=true
            ::else
            fi
...
```

Conflict appearance is in non-fulfillment of the following system state correctness condition: allow and deny cannot be performed simultaneously:

```
assert((allow && !(deny)) || (!(allow) && deny))
```

6. Relevant works

Many contemporary policy-based security systems are well-matured, but do not involve all the security categories that are presented in this paper and have differing architectures.

Extensible markup language for access control XACML [10] corresponds to SEC authorization policy. Three-level structure of policy specification (rule — policy as set of rules — set of policies) allows to build flexible resolution system using the formalized notion of decision algorithm on the levels of policy and policy set. Unlike the suggested approach, XACML does not have special system specification language, and the specification of network nodes is a part of rules description.

Language Ponder [3] contains the rules of positive and negative authorization, the rules of obligation and delegation. The authors of Ponder suggested several interesting approaches for conflict resolution strategies [11, 12], which are nevertheless too specific to the policies formalism introduced.

Flexible Authorization Framework (FAF) [13, 14] corresponds to access control systems. The FAF advantages are the detailed considering the hierarchies of objects, subjects and privileges on access estimation. The formalism used allows specifying of positive and negative authorization, involves terms of privileges propagation through hierarchies, algorithms and strategies of conflicts resolving on authorization.

There are other approaches representing different techniques for conflict detection and resolution in security policies. Here we mark the deontic logics approach [15], dynamic conflict detection with temporal logics [16, 17], as well as one of basic papers on classification of security policy conflicts [18].

7. Conclusion

This paper proposes the Security Checker architecture for policy-based security management system. Three-level structure for policies definition language is defined: from nearly natural upper-level language to object-oriented policy representation in CIM format. Security categories are specified, into which policy rules are separated. UML representation of principal classes of SEC kernel is given, the idea for implementation of conflict detecting in authorization policy is demonstrated.

Further work deals with the enhancement of techniques and algorithms of security policy verification and the design of SEC prototype basing on web-services technology.

References

[1] IETF Policy Framework (policy) Working Group. `http://www.ietf.org/html.charters/policy-charter.html`.

[2] I. V. Kotenko, A. V. Tishkov. Events calculus for specification and verification of security policies for protected computer network. 3rd Russian Conference “Mathematics and Security of Information Technologies”. Moscow, MSU, 2004.

[3] Ponder: A Policy Language for Distributed Systems Management. Department of Computing, Imperial College. `http://www-dse.doc.ic.ac.uk/Research/policies/ponder.shtml`.

[4] OSVDB: The Open Source Vulnerability Database. `http://www.osvdb.org/`.

[5] M. Rohse. Vulnerability naming schemes and description languages: CVE, Bugtraq, AVDL and VulnXML. SANS GSEC PRACTICAL, 2003.

[6] M. Grand. Patterns in Java, Volume 1, A Catalog of Reusable Design Patterns Illustrated with UML. John Wiley & Sons, 1998.

[7] R. A. Kowalski, M. J. Sergot. A Logic-Based Calculus of Events. New Generation Computing, No. 4, 1986.

[8] E. M. Clarke, O. Grumberg, D. A. Peled. Model Checking. MIT Press, 1999.

[9] G. J. Holzmann. The Spin Model Checker. IEEE Trans. on Software Engineering, Vol. 23, No. 5, 1997.

[10] OASIS: eXtensible Access Control Markup Language (XACML). http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml.

[11] L. Lymberopoulos, E. Lupu, M. Sloman. Ponder Policy Implementation and Validation in a CIM and Differentiated Services Framework. IFIP/IEEE Network Operations and Management Symposium (NOMS 2004), Seoul, Korea, 2004.

[12] A. Bandara, E. Lupu, A. Russo. Using Event Calculus to Formalize Policy Specifications and Analysis. IEEE 4th International Workshop on Policies for Distributed Systems and Networks, 2003.

[13] S. Jajodia, P. Samarati, M. L. Sapino, V. S. Subrahmanian. Flexible support for multiple access control policies. ACM Trans. Database Systems, Vol. 26, No. 2, 2001.

[14] S. Jajodia, P. Samarati, V. S. Subrahmanian. A Logical Language for Expressing Authorizations. IEEE Symposium on Security and Privacy, 1997.

[15] L. Cholvy, F. Cuppens. Analysing consistency of security policies. Proceedings of IEEE Symposium on Security and Privacy, 1997.

[16] N. Dunlop, J. Indulska, K. Raymond. Methods for Conflict Resolution in Policy-Based Management Systems. Proceedings of the Seventh IEEE International Enterprise Distributed Object Computing Conference (EDOC'03), 2003.

[17] N. Dunlop, J. Indulska, K. Raymond. Dynamic Conflict Detection in Policy-Based Management Systems. Proceedings of the Sixth IEEE International Enterprise Distributed Object Computing Conference (EDOC'02), 2002.

[18] E. Lupu, M. Sloman. Conflict Analysis for Management Policies. Fifth IFIP/IEEE International Symposium on Integrated Network Management IM'97, San-Diego, 1997.

Telematic Information Security Systems Based on Network Processors Functioning in the Stealth Filtration Mode

V. S. Zaborovsky

1. Introduction

Present-day telematic networks encompass a broad range of technical systems employed for providing packet traffic, processing of measurement data, analysis, and distribution of navigation information and telemetry data. In all these applications, data exchange is executed by forwarding and receiving network packets. The packet is a specific logical sequential/recursive structure formed at network nodes to perform data exchange. The sequential part of this structure consists of two, header and data, fields. The recursiveness of a packet is accounted for by the fact that the data itself may be another packet with its specific structure. The rules of telematic application interaction are governed by data on the destination and source addresses, and the actual packet transmission path is defined by routing protocols at network nodes. No packet processing is executed in the communication line. If the packet handling procedure produced a decision to prevent packet sending into the network, it is assumed that the packet either has reached a given network node or will be dropped. The basic router functionality is determined by two consecutive stages in the processing of packets following their arrival from the communication line, namely, "store — forward". The paper considers a new approach to choosing the architecture of telematic control devices with due account of information security issues, by which extension of functional demands on the various packet processing stages is executed by distributing the procedures of their realization among network processors (NP) and communication channels. The principal feature of the proposed approach lies in that application of NPs of a specific type functioning in the stealth regime does not interfere with the existing address connections among the network nodes and does not require any change in the routing policy. The address invariance per-

mits one to integrate information security (IS) systems into a network infrastructure without replacement of the already installed equipment, because scaled up network performance does not corrupt the throughput of the communication channels used.

2. Modern trends in the development of telematic systems

The rise of information transmission speed over communication lines and extension of the available protocol spectrum gives rise to an ever increasing demand on the performance of processors employed in packet processing at network nodes. The architecture and specific features of functioning of such devices, the network processors, has recently become a subject of intense interest [1, 2, 3]. The available solutions can be divided into two types. Solutions of the first type are aimed at boosting the router performance. The main parameters governing router operation are the packet destination addresses, therefore, these solutions are intended to accelerate data search in the router lookup tables. Solutions of the second type make use of the various packet classification procedures to improve the QoS through priority bandwidth allocation. While this separation of the processing stages increases the integrated network node throughput, part of the data can nevertheless be lost or will have to be sent once more. The packet handling efficiency and network performance can be improved by implementing a new telematic paradigm, more specifically, *process — store — forward*.

3. Information security systems

Application of the new paradigm reduces essentially to using the open-system interaction model (OSIM), in which, in contrast to classical solutions based on the TCP/IP protocol, an additional control level channel is introduced. As a result, the system controls special data structures representing actually network packets containing a set of address information. Consider the process of packet generation in the course of data transmission through a telematic network. The packet generation procedure can be divided into the following stages:

1. Collection of a data package to be transmitted through the network.
2. Configuring a structure to quantitatively determine the data volume to be transmitted.

3. Attaching to the data a special header containing a set of parameters by which the packet will be handled at network nodes.
4. Generation of a frame with a structure meeting the requirements of the communication line hardware.
5. Frame transmission over the communication line connecting two network nodes.

Several types of network nodes can be involved in packet transmission, more specifically, generation nodes which handle only packet headers; and nodes processing both headers and data. Routing or selection of the interface where the packet has to be forwarded after handling is a procedure of a local character, i.e., it is executed at each network node crossed by the packet. Routing is based on the packet destination node address specified in the corresponding header space and the lookup table connecting the network node addresses with the router interface numbers. The above process is prone to malicious actions capable of interfering with the standard packet transmission procedure or even of substituting packets on their way from the generation to destination nodes. One can conceive of the following main protection measures:

1. Tracing a packet transmission trajectory through special network nodes which execute specifically designed processing rules denying passage of packets with certain preset addresses and header parameters;
2. Operating in the tunneling mode, in which the packet to be protected is transmitted in the data space of another network packet;
3. Special packet transmission regimes, in which the header parameters are protected by cryptographic means.

These protective measures can be implemented in several ways, which can be divided into methods of filtration and of cryptographic data processing. Methods of network address space protection belonging to the first group are executed by means of special devices called firewall processors, which are installed in the network segments crossed by packet flows. These segments separate customarily the network to be protected from the interface of the router connected with this network. Protective measures of the second type require designing special network gateways supporting the tunneling regime, which in this particular case may or may not employ packet encoding. If such gateways support the routing function, using the IPSec network protocol may prove to be a promising direction for their development, thus offering the possibility of authenticating the headers of all network packets and warranting integrity of

transmitted data through application of specific cryptographic means, in particular, of electronic digital signature. Having at its disposal physical communication channels with excessive throughput, modern telecommunication industry experiences an ever increasing demand for high-efficiency methods of packet processing for routing and information protection. This demand has stimulated a broad range of studies aimed at development of special-purpose computers for use in packet processing. Development of modern NPs should take into account the present trend of growth of communication line throughput benefiting from the wide use of optical media and wave multiplexing technologies. General solutions to the problem of boosting the computer performance can be judiciously divided into the following groups: NPs based on parallel processors with a shared RAM; development of pipeline NPs with RAM resources distributed among different processing phases; and hybrid pipeline/pooled architectures in which the sequential and parallel processing stages are matched with the actual number of independent data flows. The efficiency of these solutions is dominated by the specific algorithmic features of the problems to be solved and the way by which the relevant data is supplied. Of particular significance for network packet processing are the following factors: the flow character of data, in which the number of simultaneously handled connections depends on that of nodes with different network addresses; and the sequential approach to packet transmission in the flow. Because packet transmission is executed asynchronously, i.e., it is initiated independently by each node, the number of logical sessions passing through the routers is a random quantity obeying a fractal distribution function. The packet switching processes being of a complex character, the nominal number of parallel processors in the NP architecture does not fully determine its performance, and the optimum number of pipeline processing stages depends essentially on the actual character of the problem to be solved and, thus, can vary. It is these factors that underlie the demand for development of new approaches to refinement of network packet handling organization.

4. Application of distributed NPs in security systems

Development of NPs for use in security systems can be based on separation of the packet handling procedure into basic and additional operations. Belonging in the basic operations is packet routing, while the other operations associated with extension of the NP functionality, for instance, packet filtration, should be classed with additional stages.

The proposed separation permits one to consider a network node as a part of a special packet processing network. The processing devices should be topologically connected in such a way that packet transmission among them shall not involve addresses of the nodes included into the routing lookup table. Application of this approach to information security issues offers a possibility of using the network control technology, which is based on the principle of "security through protecting the protection devices". This principle places the significance of the two key aspects of information security underlying State Standard GOST 15408, namely, functionality and confidence, on equal footing. Adhering to this principle implies that the means employed to protect information in computer networks should incorporate efficient mechanisms to ensure their own security, both in the stage of development (verifying the absence of undeclared features, UDF) and in the course of operation. To reach this goal, one should undertake at several levels of the open-system interaction model (OSIM) measures that would make localization of protection devices in the network by remote monitoring impossible. This concealment of functioning calls for modification of the protection model itself, because most of the presently employed methods of network attacks and destructive unauthorized access is based on remote incapacitation of the devices used to protect information resources in a network. Development of security means based on distributed NPs and operation in the stealth regime has become possible because the protective devices do not act in most functioning regimes as sources or destinations of network packets. Therefore, network interfaces of these devices may not have physical of logical addresses altogether; hence, IP packets or MAC frames pass through them in a way similar to how they pass through a HUB or segments of cable lines used in internetwork exchange. On the one hand, this method based on concealing the network addresses of information protection devices provides the conditions necessary for reliable operation of the security systems, while on the other, it does not require introducing any changes in network connection topology and the earlier accepted packet routing concept because the network interfaces of packet processing devices do not have addresses. Security devices based on the stealth technology have a number of assets associated not only with their concealed character of functioning but with the inherent possibility of throughput scalability and boosting the performance of operation as well. The enhancement of throughput ensues from the use of sequential/parallel network traffic regime employed, in which independent logical connections form through pipeline transmission of packets labeled by definite addresses of message sources and destinations. This offers a

possibility of cutting packet delays in the "packet filtering, processing, transfer to network" and "packet reception from network, filtering and processing" operation sequence by integrating the NPs into a specialized computer cluster. Operation with network devices based on IEEE 802.3 Ethernet technologies in the stealth regime permits packet processing in the kernel of the built-in operating system without using the TCP/IP protocol stack. This method of processing reduces the packet delay fluctuation level in buffering, which makes concealment of the protection device location still more reliable.

5. Conclusion

Application of network processors with distributed architecture broadens substantially the range of use of information security systems in telematic networks. The concealed character of operation of the protection devices offers a possibility of integrating additional packet processing procedures into the standard switching process without changing in any way the routing policy. Application of the stealth technology cuts the costs of network upgrading, because its implementation distributes the required computer power among various network devices. NP clusterization technology provides a way to scaling up the performance of network nodes and improving the reliability of new technical solutions.

References

[1] Vladimir Zaborovskii. Multiscale Network Processes: Fractal and p-Adic Analysis. Proceedings of 10th International Conference on Telecommunications (ICT'2003), University of Haute Alsace, Colmar, France, 2003.

[2] Vilchevskii N. O., Zaborovskii V. S., Klavdiev V. E., and Shemanin Yu. E. Methods of Estimation of Control Efficiency and Protection of Transport Connections in High-Speed Computer Networks. Proceedings of the Conference on "Mathematics and Security of Information Technologies" (MaBIT-03), M. V. Lomonosov Moscow State University, October 23–24, 2003.

[3] Vladimir Zaborovskii, Yuri Shemanin, Jim A. McCombs, and Alex Sigalov. Firewall Network Processors: Concept, Model, and Platform, Proceedings of International Conference on Networking (ICN'04). Guadeloupe, 2004.

Access Control Model Description Language and its Implementation in Linux Operating System Kernel

O. O. Andreev

Introduction

Growing interest to information security policies and access control models in particular exists lately due to the increasing usage of information systems in execution of tasks of practical importance. Access control models take one of the central places among other security policy components, including identification/authentication, cryptographic mechanisms and others, many of which are embedded into operating systems and are distributed with them [4].

Modern operating systems, such as Linux and Windows, use access control mechanisms based on models that were developed back in 70s, like discretionary [1] and mandatory [2]. Discretionary model is a rather primitive, though simple for description and enforcement access control model, and policies based on it, turn out to be cumbersome, hard to verify and uneasy to control. Mandatory access control model based on ordered labels is, on the other hand, very easy to verify, configure and control, but at the same time, imposed constraints are too hard, and appropriate only for a small subset of real information systems. Enforcement of security policies, using modern and more complex access control policies, such as role-based [6] is entailed by difficulties of integration additional security mechanisms into operating systems' kernels and software. There are also additional problems in UNIX operating systems, which are connected to the existence of *root* user. Root access is required in these systems for performing large volumes of administrative tasks. At the same time, this access gives its posessor uncontrolled abilities over system, which forms very rough model "everything or nothing". Additional security means were created to solve this problem, the most well known of them is `sudo` — the command, giving unprivileged user ability to execute a specified set of commands with higher privileges.

But these means do not solve the whole problem, as they have too low granularity — in bounds of one program (process, created by running an executable file and all its child processes).

The highlighted, and an amount of other, deficiencies of currently used systems stimulate work on new language means of access control description, such as eXtended Access Control Language [5], Enterprise Privacy Authorisation Language [7]. These languages give the user, responsible for information security, who will be called later *security officer* an ability to define for himself access control model, specifically tailored for needs of protected system, or to select one from the list of models, created by third-party developers or distributors. Also, work on creation and integration of additional access control models and mechanisms into popular operating system kernels, such as RSBAC Linux [8] is actively performed now.

The aim of presented work is the study of approaches, and based on them development of perspective language for access control models description and its implementation in Linux. Its primary results are the language for description of access contol models developed by the author and additional software for Linux, which allows enforcement of access control to local files and devices according to models specified in terms of the presented language. It should be noted that the language is universal and may be used in development and enforcement of security policies in distributed systems and environments based on Linux.

1. Formal description

The formal description of class of access control models that can be specified by the offered language will be presented in the current section.

Security policy of information complex is understood as a restriction imposed on functioning of this complex according to a set of informally defined rules. One of aspects of security policy is control on accesses of users of complex to its resources. This control is performed by access control subsystems of each component of complex and is based on one of access control models. The model along with control mechanisms configuration is defined in security policy. One of the most important resources in information systems are local filesystems. Access to them is controlled by operating systems. Presented language allows to describe a family of models and defines the capabilities for configuration of access control mechanisms that are based on them.

As most of the others access control models, the ones which can be described with the presented language are based on three cardinal notions:

subject, *object* and *access*. The main task of access control subsystem is giving answers to the requests "can the subject in question be granted access to specific object". Additional tasks which are supported by the subsystem may include logging and auditing of successful or unsuccessful access attempts or supplying data to intrusion detection systems.

Access control subsystem can base its decision on granting or not granting access on different data. In case of discretionary policies the decisions are based on subject, object and access identifiers. In case of multilevel security policy they are based on security labels (so-called secrecy levels in Bell — LaPadula model or integrity levels in Biba model) of subject and object and type of access. Presented language defines the class of models that are based in some sense on advanced principles laid into multilevel security.

Cardinal notion which lies in the base of models, defined by the presented language is the notion of *attribute*. Each object and subject can have attributes, which are defined by user of by system. Access is granted or denied depending on the result of applying the boolean-valued function, which is defined in the model, to attributes of subject and object in question, requested access type and system variables. That is, to defined access control model, the set of attributes, the boolean-valued function, so called *access function*, from that attributes should be defined, and configuration of access control system reduces to setting attribute values. The examples of attributes are user position in the company, secrecy level or time of last access to the object. That is, attribute can correspond a property of subject or object, which, according to the security policy, influences the decision on granting access, and access function is formalization of security policy rules, regulating access.

Traditional access control models are static, they do not evolve in time. This prohibits security officer from defining complex models and security policies, based on them, which base access decisions on the history of previous accesses of user or to object. Presented language allows to describe an access control models with such abilities. For this purpose a new notion of *post-action* is introduced. These are actions which should be executed in case of any access attempt. In the current language post-actions can change values of attributes subject and object involved in access attempt. An example post-action is incrementing attribute "number of denied access requests" at each unsuccessful access attempt. This attribute can be used later to block any access when it reaches predefined level.

Let's present the formal definition of class models described earlier.

Definition. Let's define:

- S, set of subjects;
- O, set of objects;
- A, set of accesses;
- $Attr = Attr_S \sqcup Attr_O$, set of subject ($Attr_S$) and object ($Attr_O$) attributes;
- $Value$, set of possible attribute values;
- $V_S\colon S \times Attr_S \to Value$, $V_O\colon O \times Attr_O \to Values$, functions, giving attribute values for specific subject or object;
- $P\colon Value^{|Attr|} \times A \to \{True, False\}$, access function;
- $Success\colon Attr \times Value \to (Attr_S \sqcup Attr_O) \times Value$, function, setting attribute values in case of granted access;
- $Fail\colon Attr \times Value \to (Attr_S \sqcup Attr_O) \times Value$, function, setting attribute values in case of denied access;

If subject s requests access a to object o, then access control subsystem calculates

$$P(V_S(s, attr_S^1), \ldots, V_O(o, attr_O^1), \ldots, V_A(a, attr_A^1), \ldots)$$

and if it is $True$, then access is granted, otherwise the access is denied. If access is granted attributes of s and o are set into

$$Success(V_S(s, attr_S^1), \ldots, V_O(o, attr_O^1), \ldots, a),$$

otherwise they are set into

$$Fail(V_S(s, attr_S^1), \ldots, V_O(o, attr_O^1), \ldots, a).$$

Presented language doesn't allow to define arbitrary functions P, $Success$ and $Fail$. Instead of that for defining P the class of boolean-valued functions like "not equal" or "greater" is presented to developer or security officer and access function can be defined as a superposition of function from that class, applied to attributes and arbitrary boolean function. Another class of functions is presented for defining $Success$ and $Fail$. Post-actions are sequence of steps, each of steps is setting some attribute to other attribute or value of function from the class, applied to the attribute. Each step of post-action is executed depending on the result of evaluation of condition, which is analogic to P.

This restriction on class of access functions and post-actions not only makes language syntax easier, but also allows development of automatic tools for analysis of access control model, described in the presented language. Such mechanisms are especially needed when creating dynamic models, when answering the question "will the given subject have the ability to access the specific object *in any time in future*" isn't trivial. Development of such tools will reduce the amount of work security officer has to spend on checking consistency between the given security policy and models used. Also, the need in such tools may arise when integrating multiple security policies, and, as a consequence, access control models. In these cases models should be ensured not to contradict each other. While such tasks can be performed manually, manual checks are error-prone, so automatic analysis tools should be used at maximum.

2. Language semantics

The ways to implement the described above class of access control models using the proposed language will be showed in the current section.

Description of model in the presented language is composed of structural elements — *submodels*. Each of submodels is an access control model, which can be applied to some part or all access requests. Decision, which the access control subsystem makes is based on the results of application of all submodels to the request. This composition of several submodels into one makes installation and configuration of system, where the proposed language is used, easier, allowing to add or remove submodels, when needed, making the whole access control model more hard or lax. System distributors can supply different submodels with it, allowing the security officer to construct model instead of writing his own.

Each submodel is composed of *rules*. This division, among other advantages, eases development of submodel and increases its readability. Each rule consists of *target*, *condition* and four post-actions. Target defines the requests, to which the rule is applicable, condition defines the result of such application and post-actions define the actions which will be executed if the rule is applicated. Target and condition define a pair of functions of the same type (with the same arguments and returned value) as P, post-actions are of the same type as $Success$ and $Fail$.

Calculation of access function by control subsystem, which implements the presented language, goes on the following way.

1. For each of submodels in the subsystem steps 2–4 are executed.
2. For each rule in submodel target is calculated.
3. If target is true, then condition is calculated, else go to the step 2.

4. If condition is true, add the rule into true rules list, else add it into the false rules list.

5. If after steps 1–4 list of false rules is not empty, access is denied. If list of false rules is empty and list of true rules is not, access is granted. If both lists are empty, the default decision is given.

Default decision should be defined by the implementation of subsystem. For example, it can be specified by configuration file. The appoach, taken in the presented Linux implementation is explained in section 4.

Execution of post-actions happens after all submodels are applied. The four post-actions correspond to the following cases.

1. Rule is true and the access is granted.

2. Rule is false and the access is granted.

3. Rule is true and the access is denied.

4. Rule is false and the access is denied.

Implementation should guarantee, that post-actions should be executed in the order they were specified in the submodel, but no order is guaranteed about the execution of post-actions from different submodels.

Each object and subject has a standard, predefined by an implementation set of attributes. For example, it can be filename or file owner. Values of such attributes cannot be modified by user or security officer. Also, user and security officer are given the abilities to create additional attributes with arbitrary names. Name is a character string, value can be boolean, integer, float or string. A special *nil* value is used to indicate that the attribute is not set. Security officer can allow the ordinary unprivileged users modify attributes, set by him.

System variables are defined by implementation and cannot be modified.

Language semantics gives user the ability to define complex access control models, at the same time leaving the definitions readable and easily understandable. Division into submodels allows to move the large amount of work on development of models to the operating system distributors, allowing the security officer to use the supplied submodels, describing his own model only when such need arises and limiting it to the specific area with advanced access control.

3. Syntax

The current section contains the language syntax and description of representation of the constructs described above using it.

Proposed in the current paper language is a subset of XML. One XML document defines one access control submodel. This separation of united model into several files allows to emphasize its division into submodels.

The root element should be `<policy>`. It can contain the following elements:

- `<rule>`,
- `<description>`.

Submodel is defined in the following way:

```
<policy>
    <description>
        ...
    </description>
    <rule>
        ...
    </rule>
        ...
    <rule>
        ...
    </rule>
</policy>
```

`<description>` element plays the role of comment and is used for description of submodel or one of it's rules. It's ignored by the access control subsystem.

`<rule>` element describes one submodel rule. It can contain the following elements:

- `<description>`,
- `<target>` (no more than one element),
- `<condition>` (no more than one element),
- `<rule-success-policy-success-action>` (no more than one element),
- `<rule-fail-policy-success-action>` (no more than one element),
- `<rule-success-policy-fail-action>` (no more than one element),
- `<rule-fail-policy-fail-action>` (no more than one element).

If <condition> element is absent, condition is considered to be always true. If <target> element is absent, target is considered to be always true. If one of <*-action> elements is absent, the corresponding action does not perform anything.

Generic view of the rule:

```
<rule>
    <description>
        ...
    </description>
    <target>
        ...
    </target>
    <condition>
        ...
    </condition>
    <rule-success-policy-success-action>
        ...
    </rule-success-policy-success-action>
    <rule-fail-policy-success-action>
        ...
    </rule-fail-policy-success-action>
    <rule-success-policy-fail-action>
        ...
    </rule-success-policy-fail-action>
    <rule-fail-policy-fail-action>
        ...
    </rule-fail-policy-fail-action>
</rule>
```

<target> and <condition> elements should contain boolean-valued expressions. These expressions can be composed of the following elements:

- <and>, defining boolean "and",
- <or>, defining boolean "or",
- <not>, defining boolean "not",
- <less>, defining "less",
- <greater>, defining "greater",
- <lequal>, defining "less or equal",
- <gequal>, defining "greater or equal",
- <equal>, defining "equal",

- <nequal>, defining "not equal",
- <substr>, defining "is substring",
- <strprefix>, defining "is string prefix",
- <strpostfix>, defining "is string postfix",
- <exists>,
- <access-is>,

<and>, <or> and <not> elements, located inside <target> or <condition>, are logical conjunctions and should contain two (in case of <and> and <or>) or one (in case of <not>) boolean-valued expression.

Let's give an example of describing target:

```
<target>
    <and>
        <not>
            ...
        </not>
        <or>
            <not>
                ...
            </not>
            <or>
                ...
            </or>
        </or>
    </and>
</target>
```

<less>, <greater>, <lequal>, <gequal>, <equal>, <nequal>, <substr>, <strprefix>, <strpostfix> elements define boolean-valued functions from two arguments, which may be defined by <subject>, <object>, <system>, <bool>, <integer>, <string>, <float> elements.

Functions "less", "greater", "less or equal", "greater or equal", "equal" and "not equal" are defined on the whole set of attribute values, functions "is substring", "is string prefix" and "is string postfix" are defined on string values.

If one or both values do not belong to the class of possible values, functions' result is "false". If one or both values are *nil*, all functions', except "not equal", result if "false", "not equal" results as "true".

The following expressions checks if the subject attribute e-mail ends in @molvania.com

```
<strpostfix>
    <subject>e-mail</subject>
    <string>@molvania.com</string>
</strpostfix>
```

`<access-is>` element defines function, checking if access type is equal to the given. It should contain string.

The following expression check whether the `execute` access is requested:

```
<access-is>execute</access-is>
```

`exists` elements defines function, checking the existence of attribute. It should contain one element `<subject>` or `<object>`.

The following expression checks, if subject has `e-mail` attribute:

```
<exists>
    <subject>e-mail</subject>
</exists>
```

`<subject>`, `<object>` and `<system>` elements define subject and object attributes and system variables. They should contain a string.

The following element defines `domainname` attribute of subject:

```
<system>domainname</system>
```

`bool`, `integer`, `string`, `float` define boolean, integer, string and float constants. They should contain a string.

The following expression defines *True* boolean constant:

```
<bool>true</bool>
```

`<*-action>` elements define post-actions. They should contain no less than one `<set>` element.

The generic post-action should look the following way:

```
<success-action>
    <set>
        ...
    </set>
    ...
    <set>
        ...
    </set>
</success-action>
```


`<set>` element defines that the system should set described attribute value. Attributes are defined by `<subject>` and `<object>` elements, values are defined by the following elements:

- `<subject>`,
- `<object>`,
- `<system>`,
- `<bool>`,
- `<string>`,
- `<float>`,
- `<and>`,
- `<or>`,
- `<xor>`,
- `<not>`,
- `<add>`, defining function “add”,
- `<subtract>`, defining function “subtract”,
- `<concat>`, defining function “concatenate strings”,

The following expression sets subject attribute `last-access` to the value of object attribute `id`:

```
<set>
    <subject>last-access</subject>
    <object>id</object>
</set>
```

`<and>`, `<or>`, `<not>` (located inside `<*-action>` elements), `<xor>`, `<add>`, `<subtract>` and `<concat>` define functions from two arguments, they should contain two elements `<subject>`, `<object>`, `<system>`, `<bool>`, `<string>`, `<float>`, `<and>`, `<or>`, `<xor>`, `<not>`, `<add>`, `<subtract>` or `<concat>`.

If a function receives *nil* value as it's input, all action step, defined by `<set>` element is not executed.

In the following example `num-accesses` subject attribute is incremented:

```
<set>
    <subject>num-accesses</subject>
    <add>
        <subject>num-accesses</subject>
        <float>1</float>
    </add>
</set>
```

The following example is given to prove language broad descriptive abilities. It gives the definition of simplest version of multilevel security model. Only two operations are allowed: read and write. Read access is granted if subject level is not lower than object, write is granted in the opposite case.

```
<policy>
    <rule>
        <description>
            ss-property
        </description>
        <target>
            <and>
                <access-is>read</action-is>
                <or>
                    <equal>
                        <object>type</object>
                        <string>dir</string>
                    </equal>
                    <equal>
                        <object>type</object>
                        <string>file</string>
                    </equal>
                </or>
            </and>
        </target>
        <condition>
            <gequal>
                <subject>level</subject>
                <object>level</object>
            </gequal>
        </condition>
    </rule>
    <rule>
        <description>
            *-property
        </description>
        <target>
            <and>
                <access-is>write</action-is>
                <or>
```



```
                    <equal>
                        <object>type</object>
                        <string>dir</string>
                    </equal>
                    <equal>
                        <object>type</object>
                        <string>file</string>
                    </equal>
                </or>
            </and>
        </target>
        <condition>
            <lequal>
                <subject>level</subject>
                <object>level</object>
            </lequal>
        </condition>
    </rule>
</policy>
```

As the presented example show, language syntax is relatively simple and readable even to the person which is unaware of detailed language description. Its expressive abilities are high and at the same time it is not overloaded with extensive details.

4. Implementation details

The developed language is implemented on the top of RSBAC subsystem which is a patch to Linux kernel and number of additional commands. This subsystem is included in several Linux distributions, such as Mandrake/Mandriva Linux, ALT Linux Castle. This subsystems let the developer control each access to objects under its control, including files, directories, devices, interprocess communication, processes, network channels, system data. Means for momental policy changes are developed (that means, that rights are checked at every read and write operation, not only during the opening of the file). This lessens the possibilities of race conditions. The subsystems give the developer ability to install his own access decision modules in addition to ones, present in the kernel.

The binary format was designed to improve system load times. First, policies are transformed from their XML representation into binary one

with the help `xmlplc` command. Binary representation is a memory region written to file with information about relocations. No parsing is done, only several addresses are updated according to the supplied relocations information.

After conversion compiled policy can be loaded into kernel using `xmlplset` utility. This command also let's user and system officer set subject and object attributes. These attributes are saved between system reloads, unlike loaded policies.

One of standard object attributes is it's type, value of this attribute can be "file", "directory" or "device". Standard file attributes include filename, user and group owners of the file, device attributes include minor and major device numbers. Standard subject attributes include executable filename and owner and effective owner username. System variables include system time, hostname.

Any attribute can be tagged with a flag `inherited`. If this flag is set, the value of tagged attribute will be inherited by children. Setting this flag has sense only for directories. Any user can change the value of this flag, if he has access to change value of corresponding attribute. Another flag is `user-modified`. It defines, whether ordinary user can modify this flag. It is set by security officer only.

Section 2.mentioned the case, when no rules are applicable to the requested access. In this case the implementation returns `DO_NOT_CARE` to RSBAC subsystem, which will let the other RSBAC modules decide about granting the access. It should be noted, that in case this module is only one enabled and standard UNIX access control is disabled, access will be granted.

Conclusion

The proposed language is an effective mechanism for description of rather complex access control models. Structural division eases model's development and allows to add or remove unneeded components to the model. Ability of setting attribute values delivers user from requesting security officer each time for configuration of access to his own data. For example, user can set attribute on his objects, defining the importance of these objects to him, setting the important documents as the top priority and music or video — the lowest.

Implementation of the presented language in Linux allows the process of translation existing systems and their security policies from the old mechanisms to the new ones be as easy and obstacle-free as possible, allowing at the same time enforcing old UNIX permissions security for

most of files and language-defined models for specific part of important files.

Considering the directions of development of the presented work several perspective aspects should be highlighted. As has already been mentioned in section 1, one of primary tasks for language developers is the creation of automatic tools for checking models properties. These tools can be created by the intentionable constraints imposed on the language. The author thinks that this decrease of expressiveness is compensated by the possibility that such tools can exist. As has already been mentioned, such tools will greately ease and improve security officer work and will allow him to avoid information security risks, connected with the misconfiguration of access control mechanisms.

Other directions of development of this work may be enrichment of language by adding additional types, such as bag of values, and additional functions and actions. Also, implementation of this language into other access control subsystems, for example in Web server is intended.

References

[1] A guide to understanding discretionary access control in trusted systems. NCSC-TG-003. National Computer Security Center, 1987.

[2] Bell D., LaPadula L. Secure computer systems: Unified exposition and multics interpretation. Technical Report MTR-2997. Mitre Corporation, 1976.

[3] Andreev O. O. Comparison of role-based and discretionary access control models. Materials of MaBIT-04, 2005, p. 284–291.

[4] Vasenin V. A. Problems of mathematical, algorhithmic and software means for enforcement of information security in the Internet. Materials of MaBIT-03, 2004, p. 111–143.

[5] eXtensible Access Control Markup Language (XACML) Commitee Specification. OASIS Open, 2003.

[6] Ferrariolo D., Kuhn R. Role-Based Access Controls. 15th National Computer Security Conference, 1992.

[7] Enterprise Privacy Authorization Language. IBM Research Report, 2003.

[8] `http://www.rsbac.org`.

Proactive Security and Self-Correcting Environments

O. V. Kazarin

1. Introduction

Proactive security of the computer systems (CS) is internal structural property of CS, which allows them to save functionality and security of the information resources from cyberattacks, both on a development cycle, and on an operation phase.

Moreover, if such hypothetical proactive secure CS is created, the potential user (maintaining organization) in most cases can "simply" not care whether or not it is exposed to CS cyberattacks. This CS, in such case, "will simply swallow" cyberattack, thus saving functionality.

The basic elements of methodology of creating secure CS are considered in [1], including in it proactive part. In [2] as the basic algorithmic tool for creation proactive secure distributed CS it is offered to use the distributed algorithms (protocols) of confidential calculations [3, 4] or, in more general case, protocols, which realize the criterion function, even if some of the participants of the protocol (some of processors of the distributed computing system) deviate the actions, ordered by the protocol, (for example, fault-tolerant schemes, Byzantian agreements, consensus protocols etc.).

In the given paper it is offered to use for creation proactive secure CS algorithmic toolkit, using methodology of self-testing and self-correcting programs [3, 4], which alongside with the self-correcting circuits (for example, see definition from [5, 6]), can become one of fundamental "ingredients" for creation of such CS.

2. Basic elements of methodology of creation self-testing/correcting programs

2.1. General statement of a task

Let us assume it is necessary to write the program P for function evaluation f so that $P(x) = f(x)$ for all x. The traditional methods of testing of the programs do not allow us to be convinced with probability 1 of a correctness of result of performance of the program, even because a test set of the input data, as a rule, do not cover all their possible spectrum. One of methods of the decision of the given problem consists in creation so-called self-testing/correcting programs, which allow to estimate probability of an incorrectness of result of performance of the program, that is, that $P(x) \neq f(x)$ and it is correct to calculate $f(x)$ for any x, in the event that program P on the majority sets of the input (but not all) works correctly.

To achieve correct result of performance of the program P, calculating function f, it is necessary to write such program T_f, which would allow to estimate the probability that $P(x) \neq f(x)$ for any x. Such probability will be called *error probability* of performance of the program P. Thus T_f can address to P as to the subroutine.

Obligatory condition for the program T_f is its basic *temporary difference* from any correct program of function evaluation f, in the sense that the time of performance of the program T_f, not taking into account time of calls of the program P, should be much less, than time of performance of any correct program for calculation f. In this case, calculation agrees T_f by some quantitative image should to differ from functions evaluations f and *self-testing program* can be considered as an independent step at verification of the program P, which presumably calculates function f. Besides the desirable property for T_f should consist in, that its code was as far as it probably more simple, that is T_f should be *effective* in the sense that the time of performance T_f even in view of time spent on calls P should make the constant multiplicative factor from time of performance P. Thus, the self-testing should only insignificant to slow down time of performance of the program P.

Let π means some computing task or some task of search of the decision. For x, considered as an input of a task, let $\pi(x)$ designates result of the decision of a task π. Let P be the program (presumably intended) for the decision of a task π, which do not stops (for example, has no cyclings) on all inputs of a task π. Let's speak, that P *has defect*, if for some input x of task π takes place $P(x) \neq \pi(x)$.

Let's define (*effective*) *program checker* C_π for a task π as follows. Checker $C_\pi^P(I, k)$ is any probabilistic Turing machine, satisfying to the following conditions. For any program P (presumably deciding a task π), carried out on all inputs π, for any element I tasks π and for any positive k (security parameter) takes place:

- if the program P has no defects, i.e. $P(x) = \pi(x)$ for all inputs x of task π, then $C_\pi^P(I, k)$ will give out on an output the answer "Norm" with probability not less $1 - 1/2^k$;

- if the program P has defects, i.e. $P(x) \neq \pi(x)$ for all inputs x of task π, then $C_\pi^P(I,k)$ will give out on an output the answer "Failure" with probability not less $1 - 1/2^k$.

Self-correcting program is probabilistic program C_f, which helps the program P to correct itself, if only P gives out correct result with low error probability. The given estimation means, that for any x, C_f call program P for correct calculation $f(x)$, while itself P has low error probability.

Self-testing/correcting program pair named pair of programs of a kind (T_f, C_f). Let's assume, that the user can take any program P, which purposefully calculates f and tests itself through the program T_f. If P passes such tests, then on any x, the user can call the program C_f, which, in turn, call P for correct calculation $f(x)$. Even if the program P, which calculates function f incorrectly for some small share of inputs, it in this case all the same can be used for correct calculation $f(x)$ for any x. Besides if it will be possible in the future to write the program P' for calculation f, then some pair (T_f, C_f) can be used for self-testing and self-correcting P' without its any updating. Thus, it is meaningful to spend the certain amount of time for development of self-testing/correcting program pair for applied computing functions.

Before that how to proceed to more formal description of definitions self-testing and self-correcting programs it is necessary to give definition of probabilistic oracle program (by analogy with probabilistic oracle Turing machine). Probabilistic program M is *probabilistic oracle program*, if it can call other program, which is executable during performance M. The designation M^A means, that M can make calls of the program A.

Let P be a program, which presumably calculates function f. Let I is union of subsets I_n, where $n \in \mathbf{N}$ and let $D^p = \{D_n \,|\, n \in N\}$ is a set of distributions of probabilities D_n above I_n. Further, let $err(P, f, D_n)$ is a probability that $P(x) \neq f(x)$, where x is chosen random according to distribution D_n from a subset I_n. Let β be some security parameter. Then (ϵ_1, ϵ_2)-self-testing program for function f in the relation D_p with parameters $0 \leq \epsilon_1 < \epsilon_2 \leq 1$ — the program T_f is called probabilistic oracle program which for security parameter β and any program P on an input n has the following properties:

- if $err(P, f, D_n) \leq \epsilon_1$, then the program T_f^P will give out on an output the answer "Norm" with probability not less $1 - \beta$;
- if $err(P, f, D_n) \geq \epsilon_2$, then the program T_f^P will ive out on an output "Failure" with probability not less $1 - \beta$.

Oracle program C_f with parameter $0 \leq \epsilon < 1$ is called *ϵ-self-correcting program* for function f in relation set of distributions D^p, which has the following property on an input n, $x \in I_n$ and β. If $err(P, f, D_n) \leq \epsilon$, then $C_f^P = f(x)$ with probability not less $1 - \beta$.

$(\epsilon_1, \epsilon_2, \epsilon)$-self-testing/correcting program pair for function f named pair of probabilistic programs (T_f, C_f), that there are constants $0 \leq \epsilon_1 \leq \epsilon_2 \leq \epsilon < 1$ and the set of distributions D^p at which T_f — is (ϵ_1, ϵ_2)-self-testing program for function f in relation D^p and C_f- is ϵ-self-correcting program for function f in relation distribution D^p.

Property of random self-reducibility. Let $x \in I_n$ and let $c > 1$ is an integer. The property of random self-reducibility consists that there is an algorithm A_1, working in time proportional $n^{O(1)}$, by means of which the function $f(x)$ can be expressed through computable function F from x, $a_1 \ldots, a_c$ and $f(a_1) \ldots, f(a_c)$ and algorithm A_2, working in time proportional $n^{O(1)}$, by means of which on given x it is possible to calculate $a_1 \ldots, a_c$, where everyone a_i is random distributed above I_n according to D^p.

2.2. Stability, linear and unity consistency

Let property I is expressed by the equation $I(x_1, \ldots, x_k) = 0$, where $\langle x_1 \ldots, x_k \rangle$-tuple gets out with distribution E of space D_k. Pair (I, E) characterizes family of functions F, where $f \in F$ iff when for all $(x_1 \ldots, x_k)$ with no-zero sample of elements of a tuple from E, $I^f(x_1 \ldots, x_k) = 0$. Base engineering of self-testing is the identification of property of stability for family of functions F. Informally (D, D')-stability of pair (I, E) for family of functions G realize, that if for the program $P \in G$, property $I^P(x_1 \ldots, x_k) = 0$ is satisfied with high probability, when $\langle x_1 \ldots, x_k \rangle$ is chosen with distribution E from D^k, then there is a function $g \in F \cap G$, which will be coordinated with P on the most part of inputs from D'.

Let's consider some property of linearity (I, E), where the property $I^f(x_1, x_2, x_3)$ is identical $f(x_1) + f(x_2) = f(x_3)$ and E means

$$x_1 \in_{\mathrm{R}} Z_p, x_2 \in_{\mathrm{R}} Z_p, x_1 + x_2.$$

Pair (I, E) characterizes $F = \{f(x) = cx \,|\, c \in Z_p\}$ — set of all linear functions above Z_p. In this example G is the trivial set of all functions and pair (I, E) is stable for G.

As soon as it was possible to convinced by means of random attempts, that the program P satisfies to property of stability, it is possible to pass to testing the program to a linear and unity consistency.

There are two base tests for self-testing of the programs — *test of a linear consistency* and *test of an unity consistency* [8]. Let's show construction of such tests on an example next trivial modular function.

Let x, R be positive integers, then $f_R(x) \equiv x \pmod R$, where R is fixed.

Let x_1 and x_2 is randomly, equiprobably and independently of other events are chosen from Z_{R2^n} and x accepts: $x :\equiv x_1 + x_2 \pmod{R2^n}$. It is necessary to note, that $f_R(x) \equiv [f_R(x_1) + f_R(x_2)] \equiv \pmod R$ is a linear function on all inputs (arguments). Then the test of a linear consistency consists in performance or not performance of equality: $P_R(x) \equiv [P_R(x_1) + P_R(x_2)] \pmod R$, and *error of a linear consistency* is probability that the given test was not executed.

Let z is randomly chosen from Z_{R2^n} according to distribution and z accepts: $z' :\equiv z+1 \pmod{R2^n}$. Let's note also, that $f_R(z') \equiv [f_R(z)+1] \pmod R$. Then the test of an unity consistency consists in performance or not performance of equality: $P_R(z') \equiv [P_R(z) + 1] \pmod R$, and *error of an unity consistency* is probability that the given test was not executed.

3. Applied results

3.1. Short remark

In the early 90s, during the creation of library of basic cryptographic functions "CRYPTOOLS 1.0" [11, 10] the author of the given paper one of the developers of library, without realizing, used ideas of self-testing and self-correction (in a natural manner) for debugging codes of the programs for calculation of theory-numerical functions used in interests of cryptography. Precisely at this time, formation of methodology of creation self-testing and self-correcting programs and their combinations also began [7, 8, 9, 10]. Later, at creation of newer versions of library the authors had, in essence, already good mathematical tools, affirming similar process of debugging [10].

3.2. Method of the verification of the calculating programs on a basis ST-pair functions

As the calculating program any program solving a task of reception of value of some computable function is considered. Thus verification of the calculating program is understood as process of the proof that the program will receive on some input true values of function researched. In other words, the verification of the calculating program is directed at proving the absence of deliberate and/or inadvertent program defects in a verified program.

Method of creating self-testing programs for verification of calculating program modules in this case is offered [10]. The given method does not require calculation of reference values and is independent from the

programming language used at a spelling of the reference program that essentially raises efficiency of research of the program and accuracy of an estimation of probability of absence of program defects. It is necessary at the same time to note, that the offered method can be used for the program's calculating function of the special kind, namely functions having property of random self-reducibility.

Let's assume for function $Y = f(X)$ there is pair of functions $(g_c, h_c)^Y$ such, that: $Y = g_c\big(f(a_1), \ldots, f(a_c)\big)$ and $X = h_c(a_1, \ldots, a_c)$.

It is easy to see, that if the values a_i are chosen from I_n according to distribution D^p, then pair of functions $(g_c, h_c)^Y$ provides performance for function $Y = f(X)$ property of random self-reducibility. A pair of functions $(g_c, h_c)^Y$ we shall name ST-pair of functions for function $Y = f(X)$.

Let's assume, that on ST-pair of functions can be imposed some set of restrictions on complexity of program realization and time of performance. In this case, let length of a code of the programs realizing function g_c and h_c, and the time of their performance makes the constant multiplicative factor from length of a code and time of performance of the program P.

The offered method of verification of the random calculating program P on a basis ST-pair functions for some input value of a vector X^* consists in performance of the following algorithm. Everywhere further, if the random choice of values is carried out, this choice is carried out according to distribution of probabilities D^p.

Algorithm ST

1. To determine set $A^* = \{a_1^*, \ldots, a_c^*\}$ such, that
$$X^* = h_c\{a_1^*, \ldots, a_c^*\},$$
where $a_1^*, \ldots, a_c^*$ are chosen randomly from an input subset I_n.
2. To call the program P for calculation of value $Y_0^* = f(X^*)$.
3. To call c times program P for calculation of set of values $\{f(a_1^*), \ldots, f(a_c^*)\}$.
4. To determine values $Y_1^* = g_c(f(a_1^*), \ldots, f(a_c^*))$.
5. If $Y_0^* = Y_1^*$, is made a decision, that the program P is correct on set of values of input parameters $\{X^*, a_1^*, \ldots, a_c^*\}$ otherwise given program is incorrect.

Thus, the given method does not require calculation of reference values and for one iteration allows verify correctness of the program P on $(n+1)$ values of input parameters. Thus the time of verification can be

estimated as $T = \sum_{i=1}^{c} t_i + t_x + t_g + t_{h^{-1}}$, where t_i and t_x — time of performance of the program P at input values a_i, $i = 1, \ldots, c$ and X^* accordingly; t_g and $t_{h^{-1}}$ — time of definition of meaning(importance) of function g_c and set A^* accordingly: $T_P(X)$ — time (not asymptotical) complexity of performance of the program P; $K_{gh}(X, c)$ — factor of time complexity of program realization of function g_c and definition A^* in relation to time complexity of the program P (under the assumption he makes the constant multiplicative factor from $T_P(X)$, and its value is less 1). For the traditional above-stated method of testing the time of performance and comparison of the received result with reference values makes:

$$T_0 = \sum_{i=1}^{c} t_i + t_x + \sum_{i=1}^{c} t_i^e + t_x^e > 2T_P(X)(1 + c),$$

where t_i^e and t_x^e — the time of definition of reference values of function $Y = f(X)$ at values a_i and X^* accordingly (generally, can not be less time of performance of the program).

Hence, relative prize from the point of view of efficiency of the offered method of verification (in relation to a method of testing of the programs on the basis of its reference values):

$$T_0 = \frac{T}{T_0} = \frac{\sum_{i=1}^{c} t_i + t_x + t_g + t_{h^{-1}}}{\sum_{i=1}^{c} t_i + t_x + \sum_{i=1}^{c} t_i^e + t_x^e} < \frac{1 + c + K_{gh}}{2(1 + c)} = \frac{1}{2} + \frac{K_{gh}}{2(1 + c)}.$$

As, factor $K_{gh} < 1$, and $c \geq 2$, is received a relative prize on efficiency of test of the calculating programs of the specified type (having property of random self-reducibility) more than in 1.5 times.

3.3. Researches of process of verification of the computing programs

As an example of serviceability of the offered method we shall consider verification of the program of function evaluation of discrete exponentiation: $y = f_{AM}(x) = A^x \bmod M$.

The choice of the given function is caused by that it is one of the basic functions in various theory-numerical designs, for example, in the schemes of the digital signature and authentication, public-key systems etc. Marked fact, in turn, demonstrates an opportunity of application of the offered method at research of the calculating programs deciding concrete applied tasks. Besides it is obvious, that the given function has property of random self-reducibility and proceeding from results of [8] it is possible to show, that for the given function there is $(1/288, 1/8)$-self-testing program.

For experimental researches program EXP from library of basic cryptographic functions CRYPTOOLS [11] are choice. This program realize function of discrete exponentiation (dimension of variable and constants does not exceed 128 bytes). The experimental researches consist of definition of the temporary characteristics of process of verification on the basis of use ST-pair functions and definition of an opportunity of detection by the offered method of the purposely brought in program defects.

For this purpose the following ST-pairs functions were determined:

$$g_2(a_1, a_2) = [f_{AM}(a_1) \cdot f_{AM}(1)] \pmod M,$$
$$h_2(a_1, a_2) = a_1 + 1;$$
$$g_3^1(a_1, a_2, a_3) = [f_{AM}(a_1) \cdot f_{AM}(a_2) \cdot f_{AM}(a_3)] \pmod M,$$
$$h_3^1(a_1, a_2, a_3) = \sum_{i=1}^{3} a_i;$$
$$g_3^2(a_1, a_2, a_3) = [f_{f_{AM(a_1)}}(a_2) \cdot f_{AM}(a_3)] \pmod M,$$
$$h_3^2(a_1, a_2, a_3) = a_1 \cdot a_2 + a_3;$$

During researches varied used ST-pair functions and the dimension of parameters A, M and argument X varied. The results of experiments completely have confirmed above mentioned temporary dependences (technical results of researches the author in the given paper omit).

The research of an opportunity of detection by the offered method of the purposely brought in changes consist in a spelling of the program EXPZ. The specification for the programs EXP and EXPZ same, difference consists that the program EXPZ contains a program bug of destructive character. Intentional introduction of a bug into research guaranteed failure of performance of the program of calculation of value of function $y = f_{AM}(x) = A^x$ modulo M (that is provided reception of an incorrect value of function) on each eighth part of input values of exponent x.

All input values, on which there was a failure of the program, were found out, that further has proved to be true by the verifying tests based on use of the small Ferma theorem and the theorem Euler. This fact, in turn, experimentally has shown, that the program realizing algorithm ST, is $(1/8, 1/288)$-self-testing program.

Thus, the offered method allows substantially to reduce time of test of the calculating programs for revealing inadvertent and deliberate program defects. Thus by results of tests it is possible to receive experimental estimations of probability of presence of program defects in verifiable calculating program.

Moreover, the experiment was reduced also developing algorithm SK, which allowed effectively to calculate $y = A^x$ the module M, not looking on available program bugs. Thus, in a result $(1/8, 1/288, 1/8)$-self-testing/correcting program pair for function discrete exponentiation was received.

4. Conclusion

As a result, a certain illustrative example would show fundamental essence proactive secure CS. Such figurative sketches at early stages of research are more informative, than formal language means.

Let's try to compare CS with live organisms, and proactive secure CS with genetically designed live organism (the moral and social party of a problem we omit). In general, use of a medical terminology by consideration of problems of information security is not new. It is enough to recollect computer virusology and terms used at it: "vaccination", "inoculation", "incubation period", "virus epidemic" and line others [3, 4].

The figurative analogies proactive secure CS with genetic protected alive organisms are given in the following table.

NN	**Objects of analysis**	**Genetically protected organisms**	**Proactive secure systems**
1	*Objects*	Live organisms	Computer systems
2	*Founders*	Father, mother, primogenitors on the parental lines (can have genetic anomalies; at creation not have abuse purposes);	Developer, collective of developers (some from them can be careless or can have abuse purposes)
3	*Property of organism*	Physical and intellectual properties (force, endurance, longevity, intelligence, stability to pathologies and illnesses)	Quality, reliability and security of CS
4	*Concentration on properties*	Stability to pathologies, illnesses, external traumas etc.	Protection of information and functional resources CS
5	*Subjects of attack at a creation stage*	Inherent genetic pathology (genetic defects of an exchange of substances, malignant new forms etc.)	Apriority program bugs, hardware bugs
6	*Subjects of attack on a life stage*	External caused factors (burns, wounds, etc.), virus infections	External destructive influences, destroying software (a posterity program bugs, computer viruses)

7	*Purpose of protection*	Initialization of protecting and compensating processes of organism	Initialization of processes of protection of information and functional resources of CS
8	*Basic task of protection*	Purposeful inclusion of new information in cells high organisms (this information to be directed on protecting of organisms, also on the decision of medical and biological problems, connected with correction genetic defects exchange of substances, treatment of virus diseases and malignant new forms)	Purposeful inclusion of means of protection with good spatial and time characteristics in created CS, that allows operatively "to correct" behavior of CS in a case abuse influence
9	*Tools*	Genetic engineering — designing material substance of heredity, i.e. recombinant DNA (DNA with given combination genes)	Mathematical (including cryptographic), engineering-technical
10	*Methods of protection*	Finding vector DNA, in which molecules it is possible easily to build a piece of another's DNA or to replace by a fragment of another's DNA, not breaking thus ability itself vector DNA to duplication (replicating) in cell of "owner"	Fault-tolerant schemes, (n, t)-threshold schemes, verifiable secret sharing schemes, confidential calculation, self-correcting environment etc.
11	*Traditional actions, usual scenario (reactive security)*	"Man with a heap of tablets and vaccines, with bulky medical equipment"	"Including bulky means of protection, attraction of additional system functions, software, hardware and "human" resources"
12	*Received effect*	Stable functioning of alive organism, not looking on negative external influence of destructive influances	Normal functioning CS, which "ignores" actions and afteractions of destructive means

On the basis of above-stated, it is possible to ascertain, that proactive secure CS (if they will be created) look extremely attractively for

the users of CS, which, by and large, can "not care at all" about protection of the information and functional resources from cyberattacks. In the given paper one of potentially effective tools for creation of similar systems is self-correcting hardware-software environments is offered only. Besides, as it is visible from the given paper and [3, 4], the self-correcting programs already have today a wide enough sphere of application.

References

[1] Vasenin V. A. *Problemy mathematicheskogo, algorithmicheskogo i programmnogo obespecheniya computernoy bezopasnosty v Internet* // Mathematics and security of information technologies (MaBIT-03). Materials of a conference in MSU. M.: MCCME, 2004. P. 111–141.

[2] Kazarin O. V. *Proactivnaya bezopasnost' computernych sistem* // Mathematics and security of information technologies (MaBIT-04). Materials of a conference in MSU. M.: MCCME, 2005. P. 306–320.

[3] Kazarin O. V. *Bezopasnost' programmnogo obespecheniya computernych sistems*. M.: MGUL, 2003, 212 pp.

[4] Kazarin O. V. *Teoriya i praktika zacshity programm*. 2004, 450 pp. `http://www.cryptography.ru`.

[5] Redkin N. P. *Asimptoticheski minimalniye samokorrektiruyucshiyesya schemy dlya odnoy posledovatelnosty boulevich funkciy* // Discrete analysis research of operations. Series 1. 1996. V. 3. N 2. P. 62–79.

[6] Redkin N. P. *O samokorrektiruyucshichsya schemach i o testach dlya inversnych neispraivnostey elementov* // Materials of VIII International Seminar "Discrete mathematics and its application", MSU, Moscow, 2004. P. 4–8.

[7] Blum M., Kannan S. *Designing of the programs, which check their work* // Proc. 21st ACM Symposium on Theory of Computing (STOC'89). P. 86–97.

[8] Blum M., Luby M., Rubinfield R. *Self-testing/correcting with applications to numerical problems* // Proc. 22nd ACM Symposium on Theory of Computing (STOC'90). P. 73–83.

[9] Gemmel P., Lipton R., Rubinfield R., Sudan M., Wigderson A. *Self-testing/correcting for polynomials and for approximate functions* // Proc. 23rd ACM Symposium on Theory of Computing (STOC'91). P. 32–42.

[10] Kumar R. S., Sivakumar D. *Effectient self-testing/self-correcting of linear recurrences* // Proceedings of 37th IEEE Symposium on Foundation of Computer Science (FOCS'96). P. 602–611.

[11] *Library of basic cryptographic functions CRYPTOOLS* // The Copyright Certificate RosAPO N940518 from 16.12.94.

[12] Kazarin O. V., Skiba V. Yu. *Ob odnom metode verifikacii raschetnych programm* // Security of information technologies. 1997. N 3. P. 40–43.

Improvement of Access Control Mechanizms for Distributed Computer Systems

A. A. Itkes, V. B. Savkin

Introduction

Lately the field of usage of the distributed computer systems expands very fast. The detriment that can be caused by their misuse also grows fast. The distributed computer systems can be misused with malicious actions of external or internal violators. The violators differs by their abilities, from kids to multi-national terrorist organizations. Malfunction of critical informational systems can cause ecological catastrophes, human casualities and other high losses, and current tendencies of development of modern societies and telecommunication technology are so that risks of such high-profile incidents are growing with time.

Informational security specialists often say that a system is as secure as its weakest link. This so-called "rule of weakest link" is mostly true but we should also design complex systems so that they are not completely lost even when some part of them are controlled by an adversary. This article deals with practical problem of developing design methods for complex geographically-distibuted informational system so that, on one hand, minimize inter-component dependencies and mitigate possible damage from break-in into one the of the components, and, on the other hand, leave all the links between components necessary for all principal functionality of the system. The question is also raised of designing secure heterogeneous systems in which separate components have their own incoherent security policies.

This article contains two major parts. The first part describes a possible approach to inventing formal concept of trust between nodes of a distributed informational system, giving a method to derive a security policy of distributed system when policies for its components are ready. The second part introduces an ongoing work by the authors which is

devoted to building software tools in order to define and control security policy for distributed systems in the operating system's kernel.

Here and below we define "security policy" as definition of allowed access only and do not touch upon such aspects of information security as authentication, audit, cryptography etc. Such a policy can be based on one of well-known models of access control systems. For example, when using the term "role-based security policy" we mean a policy where access control is designed after role-based access control model.

1. Trusting Relations

This chapter deals with one of the formal concepts of the confidence between different nodes of a distributed computer system. The central concept of the following reasoning is the *trusting relation*.

Before giving new defititions it is necessary to fix some designations. If A is an information system, let $O(A)$ be the set of objects of the system A, and $S(A)$ will designate the set of subjects of A.

Definition 1. The trusting relation between computer systems A and B is a subset $T_{A,B} \subset S(A) \times S(B)$. Let us say that S_B trusts S_A if $(S_A, S_B) \in T_{A,B}$. It means that, if S_B trusts S_A, then S_A can access the objects of system B through subject S_B.

One of the most widespread example of the trusting relation is that WEB-server trusts WEB-client. It means that WEB-client can read remote objects thanks to the WEB-server. Note that WEB-client, except for special cases, can not write remote objects, even of the WEB-server can write them. Such restrictions are not reflected in the models based on simple trusting relations. There is a concept of *restricted trusting relation* described below that can reflect the partial confidence between subjects of different computer systems. Models based on the restricted trusting relations have their own disadvantages that are also described below.

Sometimes it is necessary to deal with *local trusting relations*, i.e., subsets $T_{A,A} \subset S(A) \times S(A)$. Usually the confidence between different subjects of a single computer system is needed when the local security policy can not be supported by the operation system.

Let us give an exaple of a local trusting relation. The default security model of UNIX-like operations systems can not let a certain user to append data to a certain file, but deny this user to modify existing data in this file. That is why write access to system log files is usually denied for all users except for `root`, and there is a special daemon process named

`syslogd` that cat append data to the log files when other applications requests for that. So, an application can access to the system journal files through the `syslogd` subject, so, `syslogd` trusts other applications.

There is another example of the local trusting relation associated with the `bindd` daemon. In Linux, only `root's` applications can bind to ports from 1 to 1023. But, network servers are being attacked more often then other applications, so such services must have minimum privileges necessary for their primary functions. So, some of network services can run without `root's` privileges and access to a port using `bindd`, that receives a socket descriptor through a local socket and binds it to a given port, if local security policy allows it. So, `bindd` also trusts some other applications.

The `bindd` example means that the trusting relations can be used to make the local security model more expressive. But, the trusting relations can also be used to merge different security policies of different computer systems.

1.1. Merging Security Policies Using Trusting Relations

This section describes some examples of constructing the distributed security policies based on the trusting relations. All statements are given for case of a distributed computer system that consists of two systems A and B. If system C is a union of systems A and B, let us say that security policy on C is based on the trusting relations $T_{A,B}$ and $T_{B,A}$, if each subject S_A has these and only these permissions on system B, that belongs to some subject S_B that trusts S_A. The permissions of S_B on system A are defined by the similar way.

The new subjects are usually added to systems A and B to unite them, and this is a naturally action, because we allways have to install some new software to a computer when we need to connect it to a network.

1.1.1. Merging Role-Based Access Control Models

We start with description of how to merge two most simple Role-Based Access Control policies using the trusting relations. It means that two sets of elemental privileges $P(A)$ and $P(B)$ are given and each subject S_A has its role $R_{S_A} \subset P(A)$ (and, of course, $R_{S_B} \subset P(B)$). Many documents gives different versions of Role-Based Access Control Model, but we will deal with this one. More complex variations of Role-Based models can be found it [5].

Definition 2. Let computer systems A and B have Role-Based security policies. Let C be a union of A and B if $P(C) = P(A) \sqcup P(B)$ and

$S(C) = S(A) \sqcup S(B)$. Let us say that (C) is a correct union of A and B if each subject S_A has same set privileges on A, with its set of privileges according to the local security policiy of A, and if same is true for subject S_B of B.

Theorem 1. *Let computer systems A and B have Role-Based security models, with fixed sets $P(A)$ and $P(B)$. Then, any correct Role-Based union of A and B can be defined using a pair of trusting relations $T_{A,B}$ and $T_{B,A}$.*

Proof. Let $r_{A,1}, r_{A,2}, \dots r_{A,N_A}$ be elemental privileges of system A. Let us add roles $R_{A,1} = \{r_{A,1}\}, \dots R_{A,N_A} = \{r_{A,N_A}\}$ and subjects $S_{A,1}$, $S_{A,2}, \dots S_{A,N_A}$ to A. Each subject $S_{A,j}$ will have role $R_{A,j}$ for $j = 1$, $2, \dots N_A$. Furthermore, let each subject $S_{A,j}$ trust these and only these subjects of system B that must have privilege $r_{A,j}$ on A. This is how $T_{B,A}$ must be built.

Similarly, we can build the set $T_{A,B}$ to grant subjects of A necessary access rights to objects of B. □

1.1.2. Merging Multi-Level Security Models

So, each correct union of Role-Based Access Control models can be defined using a pair of trusting relations. Unfortunately, this is not true for Multi-Level Security Policies, i.e., policies that assign security levels to each object and set all the access rights to allow all information flows to higher security levels and deny all information flows to lower security levels. It means that subject S can read object O if and only if security level of S is higher then security level of O and S can write to O if and only iff level of S is lower then level of O.

Definition 3. Let A and B be computer systems with Multi-Level Security policies. The system C is a union of systems A and B if $O(C) = O(A) \sqcup O(B)$ and $S(C) = S(A) \sqcup S(B)$. Let us say that C is a correct union of A and B if C has a Multi-Level Security policy so that:

- Each subject of A has those and only those permissions to objects of A that it has according to security policy of A.
- Each subject of B has those and only those permissions to objects of B that it has according to security policy of B.
- Subjects of A can not transmit information to lower security levels of A using objects of B.
- Subjects of B can not transmit information to lower security levels of B using objects of A.

Theorem 2. *Let computer systems A and B have Multi-Level Security policies and each system has at least one subject on each security level. If a correct union of systems A and B can be defined using trusting relations, then level-sets of A, B and C are isomorphic.*

We will need some lemmas to proof that.

Lemma 1. *Let S_1 and S_2 be any subjects of A. Then, if S_1 and S_2 hase same security level in C, then S_1 and S_2 has same security level in A.*

Proof. Let S_1 and S_2 belong to different security levels in A and to single level in C. If levels of S_1 and S_2 are comparible, we will suppose that S_1 is higher then S_2. It means that there is an object O in A, so that S_1 can read O, but S_2 can not do it. That is also true if security levels of S_1 and S_2 are not comparible.

But object O also belongs to C, so, in system C, S_1 can read O, but S_2 can not do it. So, S_1 and S_2 have different security levels in C. This contradiction proves the theorem. □

Lemma 2. *Let subjects S_1 and S_2 belong to a single security level in system A. Then, according to security policy of C, S_1 and S_2 has same permissions to objects of B.*

Proof. Let O_B be an object of B, so that S_1 can read O_B, but S_2 can not read it. Let O_A be an object of A that belongs to same security level with S_1 and S_2. The subject S_1 can read O_B, so, in system C, S_1 is higher then O_B. Next, S_1 can write to O_A, so O_A is higher then S_1. And, S_2 can read O_A, so S_2 is higher then O_A. It means, that

$$L(O_B) \le L(S_1) \le L(O_A) \le L(S_2)$$

so $L(O_B) \le L(S_2)$. So, S_2 can read O_B.

The case that S_1 can write to O_B, but S_2 can not do the same, can be considered similarly. □

Lemma 3. *Let S_1 and S_2 be subjects of A. Then, S_1 and S_2 has same security level in A if and only if the have same security level in C.*

Proof. According to lemma 1, if subjects S_1 and S_2 have same security level in C, then S_1 and S_2 have same security level in A. We only have to prove the inverse statement. Let S_1 and S_2 lay on the same security level in A. So, S_1 and S_2 has exactly same access rights to objects of A. But, according to lemma 2, S_1 and S_2 has also same access rights to objects of B. So, subjects S_1 and S_2 has same permissions in system C, and belongs to same security levels in C. □

Lemma 4. *Level-sets of systems A and B are subsets of level-set of C.*

Proof. We will prove that level-set of A is isomorphic to a certain subset of the level-set of system C. Let $l_1 \dots l_n$ be security levels of A, and $L_1 \dots L_m$ be security levels of C. Let $F : \{l_1 \dots l_n\} \to \{L_1 \dots L_m\}$ be a function with the following properties: if subject S_1 belongs to level l_j of A, then that subject S_1 belongs to level $L_k = F(l_j)$ of C. According to lemma 3, F is a well-defined reflection and F is injective. We still have to prove that, if $l_1 \geq l_2$, then $F(l_1) \geq F(l_2)$, and if l_1 and l_2 are not comparible, then $F(l_1)$ and $F(l_2)$ are incomparible, too.

If $l_1 \geq l_2$, then let subject S_1 belong to l_1 and subject S_2 belong to l_2. In system A there is an object O such that S_1 can read from O and S_2 can write to O. It is also true according to security policy of C, so $F(l_1) \geq F(l_2)$.

Now, let l_1 and l_2 be incomparible. Then, let an object O belong to security level l_1. Let subject S_1 belong to l_1 and subject S_2 belong to level l_2 in system A. Then, S_1 can both read and write O, so, in C, object O belongs to level $F(l_1)$, just like subject S_1. If $F(l_2)$ was comparible with $F(l_1)$, then, in system C, S_2 could read or write object O, but that is not true in system A, hence, and in system C, too. So, if l_1 and l_2 are not comparible, then $F(l_1)$ and $F(l_2)$ are incomparible, too. □

Theorem proof. Let Multi-Level systems A and B be merged using the trusting relations. Let L_A be the level-set of system A, L_B be the level-set of system B and L_C be the level-set of system C. According to lemma 4, $L_A \subset L_C$. We must prove that $L_A = L_C$. Let l_0 be a level of C that does not contain any objects of system A. Then l_0 corresponds to a certain level of system B, because otherwise it would not contain either objects of A or B. Let a subject S_B belong to l_0.

Let level l_0 be comparible with some security level of A. Then, subject S_B has some permissions on an object O of system A, i.e., there is an object S_A in system A, such that S_A trusts S_B. Then, S_A lays on the level l_0 in system C. To prove that, we will first suppose that S_A and O lays on a single secuirty level in A. Then, S_A can read or write to object O, and subject S_B can do the same, so S_A, S_B and O lays on one security level l_0. On the other hand, if S_A can read object O or only write to it, we may use object O' instead of O, where O' lays on the same security level with S_A in system A.

We still have to prove the theorem in case when level l_0 is not comparible with any level of system A. It means that all subjects of system C that lays higher then l_0 can not write to any object of A. This is also true to level $l_{\max}$ that is the highest level of system C. It means that $l_{\max}$ does not correspond any security level of system A, so subjects of C, that lay on $l_{\max}$ has no permissions to objects of A. So, no one subject of system C can read any object of A, but subjects of A are also subjects of C, so such case is impossible. □

We just proved that Multi-Level Security policies of computer systems A and B can be merged using the trust relations only if security grids of A and B are isomorphic. But, other policy merge methods can merge Multi-Level policies without such condition. For example, we allways can allow all subjects of A to write to any object of B, and allow all subjects of B to read all subjects of A. In that case, the united security policy will also be Multi-Level, without dependency to structure of security grid of A and B.

Note that there is an important condition that each security level of each system contains at least one subject. If it is not true, this is sometimes possible to merge the Multi-Level Security polcies with non-isomorphic security grids using the trusting relations.

Let us give an example. Let computer system A have two security levels H_A and L_A, such that $H_A > L_A$. Let system B have threee security levels $L_B < M_B < H_B$, and there are no subjects at level M_B. Let us build the trust relation $T_{A,B} = (L_A \times L_B) \sqcup (H_A \times H_B)$ and $T_{B,A} = T^T_{A,B}$. It means that each subject of level L_A trusts each subject of L_B, and each subject of H_A trusts each subject of H_B, and $(S_B, S_A) \in T_{B,A}$ if and only if $(S_A, S_B) \in T_{A,B}$. It that case, system C that is union of A and B has a Multi-Level Security policy of three security levels: $H_C = H_A \sqcup H_B$, $M_C = M_B$ and $L_C = L_A \sqcup L_B$.

1.2. The Limited Trusting Relations

The described method of uniting the security policies has some disadvantages. One of such disadvantages is that the model of trusting relation does not reflect the nature of the subjects. For example, the `syslogd` program trusts all other subjects and has a permission to clear the registration files, but really it would not do it. If we use the simple trusting relations model, we get that any subject can clear the system log, but that would not happen.

The second disadvantage of trusting relations model is that not any pair of security policies can be merged using it. For example, the Multi-Level Security models can be united only in an individual case.

To remove these disadvantages we can use the concept of *limited trust relation*. This trusting model minds to subject S_B can particularly trust subject S_A and execute not all requests of S_A that it can execute according to its permissions.

Definition 4. Let M be a certain grid, called the trusting grid. The limited trusting relation between systems A and B is the pair of $(T_{A,B}, F)$ where $T_{A,B}$ is a normal trusting relation and F is a reflection $F : T_{A,B} \to M$ and the value of $F(S_A, S_B)$ is called the trust level between S_B and S_A.

In such conditions, each operation that can be executed by S_B has a minimum trust level $m_0 \in M$. Then, S_A can execute an operation on objects using S_B if and only if $(S_A, S_B) \in T_{A,B}$ and $m_0 \leq F(S_A, S_B)$.

Note that the grid M may depend on a particular subject S_B of system B. In that case, $F(S_A, S_B)$ must belong to $M(S_B)$.

Let us give an example. Let S_2 be the service `bindd`. Then, if the system considered has only one local IP address, $M = 2^P$ is the grid of all subsets of the set of ports $P = 1, 2, \ldots, 65535$. Furthermore, for each subject S_2 we can declare trust level $F(S_1, S_2)$ as the set of all ports that can be used by S_1. If the system has several local addresses, the set M will consist of all subsets of set $P \times A$ of all ordered pairs (port, address).

1.2.1. Merging Multi-Level Security Models Using the Limited Trust Relations

The limited trust relations lets us avoid another disadvantage of the simple trust relations. It is that not all the Multi-Level Security policies can be united using the simple trust relations.

Theorem 3. *Any correct union of the Multi-Level Security systems A and B can be expressed using limited trust relations between A and B.*

Proof. Let us build a limited trust relation $(T_{A,B}, F)$. Let

$$L_1(B) \ldots L_N(B)$$

be security levels of system B and

$$L_1(C) \ldots L_M(C)$$

be security levels of system C that is the union of A and B.

Let us add subjects $S_1 \ldots S_N$ such that for each $j \in \{1 \ldots N\}$ S_j lays on $L_j(B)$, to system B. Let each subject S_B have trust grid $L = 2^P$, where P is the access set $P = \{read, write\}$.

So, the trust level between S_j and subject S_A of A is a subset $P' \subset P$. It means that if $read \in P'$, then S_A can read objects at level $L_j(B)$ using subject S_j. Similarly, if $write \in P'$, then S_A can write to objects of level $L_j(B)$ using S_j.

Let subject S_A lay on level $L_S(A)$ of system A, that corresponds the level $L_S(C)$ of system C. For each j, assume that level $L_j(B)$ corresponds to level $L_{f(j)}(C)$ of system C, where

$$f : \{1, \ldots, N\} \to \{1, \ldots, M\}.$$

Then, for each $j \in \{1, \ldots, N\}$, let us define the trust level of S_j to S_A using the following rules:

- $F(S_A, S_j) = P$, if $L_S(C) = L_{f(j)}(C)$.
- $F(S_A, S_j) = \{read\}$, if $L_S(C) > L_{f(j)}(C)$.
- $F(S_A, S_j) = \{write\}$, if $L_S(C) < L_{f(j)}(C)$.
- $F(S_A, S_j) = \emptyset$, if $L_S(C)$ is not comparible with $L_{f(j)}(C)$.

So, the Multi-Level Security systems allways can be merged using a pair of limited trust relations. □

1.2.2. Merging Role-Based Access Control Models Using the Limited Trust Relations

It is already proved that any correct union of Role-Based Access Control models of computer systems A and B can be represented using a pair of trust relations $T_{A,B}$ and $T_{B,A}$. Note that if the systems mentionet have too much privileges, such uniting may require too much extra subjects. But, using the limited trust relations, we can unite any pair of Role-Based Access Control policies with adding anly one extra subject to each of these systems.

Theorem 4. *Any correct union of Role-Based Access Control policies of computer systems A and B can be represented using a pair of limited trust relations between subjects of A and B.*

Proof. Let S_B^A be a subject of system B such that the role of S_B^A is the full set $E(B)$ of the privileges of system B. Let the trust grid of S_B^A be the set of $R(B) = 2^{E(B)}$. Then, for each subject S_A of system A the trust level between S_B^A and S_A can be calculated by formula $m = r_C(S_A) \cap E(B)$, where $r_C(S_A)$ is the role of subject S_A in the system C. This is how to build the limited trust relation $(T_{A,B}, F_{A,B})$.

The limited trust relation $(T_{B,A}, F_{B,A})$ can be defined similarly. □

The considered method of merging the Role-Based Access Control Models has some disadvantages when practically realized. The systems

A and B have now the single points of vulnerability — subjects S_B^A and S_A^B. These subjects have all possible privileges, so successfull attack on such subject will grant all the privileges on the system to attacker. We think that the best way is the combinde way, i.e., add some particallytrusting subjects to system A (B), each with different set of privileges.

1.2.3. Some Disadvantages of Security Models Based on Limited Trust Relations

Complexity is one of the most valuable disadvantages of the security models based on limited trust relations. It is usually difficult to describe the trust grid for a subject formally.

The trusting grid for a certain subject can be built using the set of all requests that may be executed by that subject. If R is such set of requests, the set $M = 2^R$ may be used as a trust grid. If a subject S_1 has trust level m at subject S_2, then, S_2 will execute request r of S_1 if and only if $r \in m$.

This approach to the problem of constructing trust grids has a disadvantage that the request set is usually very large.

Another disadvantage of models based on the limited trust relations is that applications are oftenly vulnerable for buffer overflow attacks and other attack types that can be used to force a remote subject to do execute some evil code. So, if subject S_B partially trusts S_A, buffer overflow attack may allow S_A to execute any operation on system B with rights of S_B, regardless to the trust level. So, limited trust relations may "transform to simple trust relations" in some cases. According to this fact, it is needed to continue work on analyzing risks of compromising some components of distributed computer systems.

1.3. Resume

There are two methods of building distributed security policies. Neither method is perfect, each has its own disadvantages. Note that simple trust relations can be quite easily controlled by the operating system, instead of the limited trust relations. So, it seems reasonable to modify operating system kernel to support the simple trust relations, and let applications control trust level of the limited trust relations without any help from the kernel.

2. Constructing the Distributed System Based on Trusting Relations

This section deals with one approach to a practical implementation of distributed system whose security policy is based on trusting relations.

The authors are working on a software module for Linux kernel that will be used for defining and controlling security policies of distributed systems. This software package is based on SELinux — an NSA project which has a support from several major Linux distributions, e.g., Red Hat.

2.1. Security Enhanced Linux

SELinux is an extension of the Linux kernel for improving of Linux access control mechanisms. It includes support of Type Enforcement model which is briefly described below. It should be noted that traditional UNIX security mechanisms based on simplified discretional model continue to work when using SELinux.

2.2. Type Enforcement Security Model

SELinux accepts security policy definitions in terms of Type Enforcement (TE) model. This model assoctiates each object with a label called a *type*. Type is also called a *domain* when one or more subject can have this type. Besides that every object belongs to some *class* which describes its "nature", e.g., plain files, device files, FIFOs etc. are classes. A function $F : T \times T \times C \times A \to \{allow, deny\}$, where T is the set of all types, C — the set of all classes of the system, and A is the set of kinds of access describes all allowed accesses of subjects to objects. So, access can be permitted based on a type of subject, a type of object, a class of object and a kind of access. Permitted accesses depends on class of object in such a way that e.g., it is possible to send a signal to a process but not to a file.

Policies based on Type Enforcement models can always be combined with the aid of trusting relations. To prove this statement we will first make some new definitions. If A is a component of distributed system, let's call $T(A)$ a set of all types of A, and $T'(A) = T(A) \sqcup \{none\}$ — an extended set of types, where *none* is a special type that has no access rights.

Definition 5. Let A, B, and C be informational systems and C be constructed using the following rules: $O(C) = O(A) \sqcup O(B)$, $S(C) = S(A) \sqcup S(B)$, $T(C) \subset T'(A) \times T'(B)$, where access rights of subject S of type $t = (t_A, t_B)$ to the objects of A are defined by t_A, and to objects of B — by t_B. In this case C is called a correct union of A and B iff every subject from $S(A)$ is allowed that and only that accesses to objects from $O(A)$, which are allowed to this object by the security policy of A, and,

similarly, every subject from $S(B)$ is allowed that and only that accesses to $O(B)$, which are allowed by the security policy of B.

Theorem 5. *Every correct union of TE-systems A and B can be defined with a trusting relation between subjects of A and B.*

Proof. Let subject S_A of A with a type of t_A trust a subject S_B of B having a type of t_B if and only if S_B has a type of $t = (t_A, t_B)$ in the system C. Relation $T_{B,A}$, constructed by this rule, gives all necessary access rights to subjects of A. Relation $T_{A,B}$ to give all necessary rights to $S(A)$ can be constructed in the same way. □

2.3. The Mathematical Model for the System

Currently the authors are developing an extension to the Linux kernel for allowing the specification of trusting relations between nodes of a distributed system. Let us assume that a trusting relation between hosts corresponds to the dissection of objects to types, namely, if a pair $(S_A, S_B) \in T_{A,B}$ and pairs of subjects S_A, S'_A and S_B, S'_B have the same types of TE model, then $(S'_A, S'_B) \in T_{A,B}$. In the other words, one should specify trust not between subjects (the set of which can grossly change with time) but between types of A and B (which is more convinient because sets of types do not change without policy reloads in SELinux).

When different hosts interact over a network, each of them will identify types of remote interacting subject by the labels on IP packets which are carried in IP options field. It is better to exchange not symbolic type names but specially defined numeric identifiers. Also to lessen possible interdependencies between security policies of the hosts, we will allow several mapping tables between SELinux types and their numeric identifiers on each host. Thus a node can simultaneously join several subnetworks, each having its own policy of assigning type labels.

Let's give a formal model for the proposed system.

Let A be a component of distributed informational system, $T(A)$ beeing the set of types of A and M — the set of nodes with which A interacts. Let M be divided into disjoint subsets $M = M_1 \cup \ldots \cup M_N$. For each $k \in \{1 \ldots N\}$ then will defined a mapping $F_k : T(A) \to I_k(A)$, where $I_k(A)$ is a set of natural numbers, so that image of F_k will cover full set $I_k(A)$, and each non-zero element of $I_k(A)$ will have only one prototype. We will call such mappings *type correspondence tables*. Value of zero will mean that a trust level for the type is not defined.

Let for each node $B \in M_k$ be defined a set $T'_{A,B} \subset I_k(A) \times I_j(B)$, where j is such that A belongs to the jth component of the set of nodes which are exterior to B (in a same way as B belongs to kth component of the set of nodes which are exterior to A). Assume that subject S_B of type t_B allowed to interact with S_A of type t_A iff $\big(F_k(t_A), F_j(t_B)\big) \in T'_{A,B}$. This condition define a trusting relation between A and B.

2.4. Current State of Development and Further Research Areas

A Netlink family `NETLINK_TRUST` is defined for configuration of type identifiers and trusting relations between them. Netlink interfaces is very good for this tasks for several reasons, one of them beeing that an application can easily be notified when configuration change is occured.

Currently the development work is still in its early phase but even today we can name some merits of our approach.

- Independently developing security policies can be combined. When a policy of one node changes, it's not necessary to reconfigure other nodes presuming that numeric type identifiers preserve their meanings.
- A node can join several distributing systems each having its own numbering policy. Different type correspondence tables can be used for it.
- This approach does not require a central authentication server which gives an ability to design systems of complex structure without a single point of failure.

Conclusion

As the value of distributed informational systems for many aspects of society rises, so does the importance of the adequate protection of these systems againt malicious acts. While the goal of absolute protection is unreachable, resistance of a system as a whole to compromises of its components can be significantly improved by designing security policies in such a way as to lessen interdependencies between components, allowing only necessary interactions between them.

In this article we considered one approach to make a formal model of a notion of trust between components of a distributed system. We showed that this approach can describe merging of components with different security policies into one distributed system. Also we described some technical decisions we made when designing specific software tools for building distributed systems with this approach.

References

[1] A. A. Grusho, E. E. Timonina. Teoreticheskie osnovy zaschity informatsii (In Russian). Moscow, Yachtsmen, 1996.

[2] V. A. Vasenin. Informatsionnaya bezopasnosty i computerniy terrorizm. Nauchnye i metodologicheskie problemy informatsionnoy bezopasnosti (In Russian). Moscow, MCCME, 2005.

[3] V. A. Vasenin, A. V. Galatenko. Matematicheskie modeli raspredelennych informatsionnyh sistem (In Russian). MaITS — 2004. Moscow, MCCME, 2005.

[4] O. O. Andreev. Sravnenie rolevoy i diskretsionnoy modeley razgranicheniya dostupa (In Russian). MaITS — 2004. Moscow, MCCME, 2005.

[5] Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, Charles E. Youman. Role-Based Access Control Models. IEEE Computer, 29, 2, 38–47, 1996.

[6] Joel McNamara. Secrets of Computer Espionage: Tactics and Countermeasures. Moscow., BINOM, 2004.

[7] Gowri Dhandapani. Netlink Sockets — Overview. The University of Kansas, 1999.

On Analysis of Approaches to Cyberattack Taxonomy

A. A. Klimovsky

Introduction

Constantly increasing number of computer security incidents leads [1] to the necessity of creation of organized (or self-organizing) structures, which are targeted at providing actual information about exposed vulnerabilities, fixing these vulnerabilities, creating intrusion detection systems and other measures, etc [2]. Due to the above reasons there is a sufficient amount of information about the latest vulnerabilities and computer attacks. However such information (especially concerning security incidents) is heterogeneous, unstructured and hardly suitable for further analysis. As the result there emerges the necessity of creating a model and a toolkit for knowledge accumulation ans systematization. Taxonomy could be used as such a model. Besides essential and systematic computer attack description in practice taxonomy can be also used to perform further attack analysis, computer system risks estimation, attacker model creation and security policy formulation. Another application is intrusion detection systems [3].

1. Posing the problem

First let us define the term attack [4]:

Attack — a series of steps taken by an attacker to achieve an unauthorized result.

A subject, performing these actions, is called *an attacker*, and a system, which is attacked, is called *a victim*.

Formally, the problem of attack classification is to create the categorisation scheme — the method of referring the attack to one of categories using its distinguishing features.

To avoid confusion in terms the classification, the classification scheme and others, further we will use one term — the taxonomy.

The word taxonomy origins from Greek: $\tau\alpha\xi\iota\nu o\mu\iota\alpha$ (taxinomia) origins from "taxis" — order and "nomos" — law [5], [6]. Formal modern definition can de found in [7]: *"A taxonomy is a classification scheme that partitions a body of knowledge and defines the relationship of the pieces"*. Most famous example of taxonomy is Carolus Linnaeus's taxonomy of animals and plants [8]. The other less famous taxonomies, directly concerning with computer attack classification, are described below. Evolution of ideas, approaches and methods of solving this problem is presented by the example of some of them.

Before describing them, let us consider approaches to formulating criteria of taxonomie's value. Taxonomy should satisfy some rational characteristics ([4], [9], [10], [12], [13], [14], [15], [16], [17]). Obviously, it is almost impossible to satisfy all of them and in practice taxonomy is often a rational compromise between them. Nevertheless, these characteristics help us to point out some advantages and disadvantages of the taxonomies. So it is useful to cite them here.

- *Mutually exclusive* ([4], [9], [10], [12], [14], [15]).
  This means that categories of taxonomy, as attributes/identificators of sets of attacks do not overlap.
- *Exhaustive* ([9], [10], [12], [13], [14], [15]).
  Categories of taxonomy should cover all computer attacks and taxonomy is able to classify all of them.
- *Deterministic* ([9], [10], [14], [16]).
  Classifying procedure should be clearly defined and explained.
- *Terms well defined* ([9], [10], [17]).
  All terms, which are used in taxonomy, should be clearly defined and explained, so there is no confusion in their values.
- *Objectivity* ([9], [14], [16]).
  All properties of attacks could be impartially identified.
- *Useful* ([4], [10], [9], [12], [13], [14], [15]).
  Taxonomy should provide information about field of research.
- *Comprehensible* ([9], [14], [15]).
  A comprehensible taxonomy will be able to be understood by those who are in the security field, as well as those who only have an interest in it.
- *Unambiguous* ([4], [9], [12], [13], [14], [15]).
  Each category of the taxonomy must be clearly defined so that there is no ambiguity as to where an attack should be classified.
- *Conforming* ([9], [10], [14], [15]).
  Existing terminology should be used in the taxonomy so as to avoid confusion and to build on previous knowledge.

- *Repeatable* ([4], [9], [10], [12], [13], [14], [16]).
  Repeated applications result in the same classification, regardless of who is classifying.

However, some of these characteristics are close in meaning to other, or are a sequent of them. For this reason we can combine these characteristics or, respectively, delete them. Let us divide characteristics into two groups: main characteristics, connected with the content of the taxonomy and secondary characteristics, concerning the form of description of the taxonomy. Main characteristics: mutually exclusive, exhaustive, useful, deterministic, objective, extensible. Secondary characteristics: term well defined, comprehensible, conforming. Extensibility is new characteristic on this list, but, in the author's opinion, it is rather important. Extensibility characterizes ability of taxonomy of being changed without global changes of main structure of the taxonomy.

2. Analysis of previous works

Traditionally, security incidents were broken into categories of disclosure of confidential information (loss of confidentiality), loss of integrity, and denial-of-service (DoS) attacks (loss of availability) [12], [19]. Main disadvantage of such division is that category of the attack doesn't provide much information about its character. However, of course, the effect of the attack is its important property and it is used in many taxonomies ([4], [10], [14]).

Another approach is classification of vulnerabilities of hardware and software. One of the first works in this area was [20] by Attanasio, Markstein, and Phillips. Separation of attacks using type of vulnerability was partially used by Howard and Longstaff in [4]. Then this approach was developed in [21], but it wouldn't be described here in detail, because it is rather specific and usually is applied in special areas (for example in problems concerning software testing).

One of possible variants is to separate attacks by using authorisation abilities of the attacker. An example of this approach is, so-called, the Anderson's matrix [18]. James P. Anderson developed a four cell matrix that covers the types of penetrators, based on whether they are authorized to use the computer and the data/program source. This matrix is shown in Table 1.

Category B is divided by Anderson into 3 subclasses:

A) External Penetrator.
B) Internal Penetration.
   (i) the masquerader (defeats procedural controls);

Table 1. Anderson's matrix

	Attacker not authorized to use data/program resource	Attacker authorized to use data/program resource
Attacker not authorized use of computer	*Case A* External Penetration	—
Attacker authorized use of computer	*Case B* Internal Penetrator	*Case C* Misfeasance

(ii) the legitimate user;

(iii) the clandestine user (defeats logical controls).

C) Misfeasance.

Another way is to make complete list of attack types. Probably, the most famous work in this field is articles of Peter Neumann and Donn Parker ([22], [23], [24], [25]) Similar ideas were developed by Simon Hansman in [10]. Main advantage of such approach is its appropriateness, because type of attack usually gives more information than other attributes of attack. However, the area in which it could be applied is not large, because it is very hard to make a list of types which is complete, exhaustive and mutually exclusive at the same time.

In Neumann's and Parker's work there are 9 computer misuse techniques (see Table 2). Later Neumann [23] expanded the nine categories of Neumann and Parkers list into twenty six types of attacks, shown in Table 3.

Table 2. Categories of Computer Misuse

1	External
2	Hardware misuse
3	Masquerading
4	Pest programs
5	Bypasses
6	Active misuse
7	Passive misuse
8	Inactive misuse
9	Indirect misuse

Table 3. Types of attacks

External	
Visual spying	Observing of keystrokes or screens
Misrepresentation	Deceiving operators and users
Physical scavenging	Dumpster-diving for printout
Hardware	
Logical scavenging	Examining discarded/stolen media
Eavesdropping	Intercepting electronic or other data
Interference	Jamming, electronic or otherwise
Physical attack	Damaging or modifying equipment, power
Physical removal	Removing equipment and storage media
Masquerading	
Impersonation	Using false identities external to computer systems
Piggybacking attacks	Usurping communication lines, workstations
Spoofing attacks	Using playback, creating bogus nodes and systems
Network weaving	Masking physical whereabouts or routing
Pest programs	*Setting up opportunities for further misuse*
Trojan horse attacks	Implanting malicious code, sending letter bombs
Logic bombs	Setting time or event bombs (a form of Trojan horse)
Malevolent worms	Acquiring distributed resources
Virus attacks	Attaching to programs and replicating
Bypasses	*Avoiding authentication and authority*
Trapdoor attacks	Utilizing existing flaws
Authorization attacks	Password cracking, hacking tokens
Active misuse	*Writing, using, with apparent authorization*
Basic active misuse	Creating, modifying, using, denying service, entering false or misleading data
Incremental attacks	Using salami attacks
Denials of service	Perpetrating saturation attacks
Passive misuse	*Reading, with apparent authorization*
Browsing	Making random or selective searches
Inference, aggregation	Exploiting database inferences and traffic analysis
Covert channels	Exploiting covert channels or other data leakage
Inactive misuse	*Willfully failing to perform expected duties, or committing errors of omission*

Indirect misuse	*Preparing for subsequent misuses, as in off-line preencryptive matching, factoring large numbers to obtain private keys, autodialer scanning*

In view of all reasons mentioned above, complex approach seems to be the most perspective one. [4], [10], [14] are the examples of using this approach. However, let us remark that the ways of combining could be different.

The first way is to consider analysed attributes as independent entities. This idea was realized in [10]. In his work Simon Hansman developed the so-called concept of dimensions. The taxonomy proposes four dimensions for attack classification. The first, or base, dimension (see Table 4) is used to categorise the attack into an attack class that is based on the attack vector, or if there is no attack vector, the attack is classified into the closest category. The attack target is covered in the second dimension (see Table 5). The target can be classified down to very specific targets, such as Sendmail 8.12.10 or can cover a class of targets, such as Unix based systems. The third dimension covers the vulnerabilities and exploits, if they exist, that the attack uses. The vulnerabilities and exploits do not have a structured classification due to the possible infinite number of vulnerabilities and exploits. Instead the list defined by the Common Vulnerabilities Exposures project [11] is used as a starting point. The fourth dimension takes into account the possibility for an attack to have a payload or effect beyond itself. In many cases an attack will be clearly a certain kind of attack, but yet it will have a payload or cause an effect that is different. For example, a virus that installs a Trojan Horse, is still clearly a virus, but has a Trojan as a payload.

Table 4. First dimension

Viruses:	File Infectors	
	System/Boot Record Infectors	
	Macro	
Worms:	Mass Mailing	
	Network Aware	
Trojans:	Logic Bombs	

Buffer Overflows:	Stack	
	Heap	
Denial of Service Attacks:	Host Based:	Resource Hogs
		Crashers
	Network Based:	TCP Flooding
		UDP Flooding
		ICMP Flooding
	Distributed	
Network Attacks:	Spoofing	
	Session Hijacking	
	Wireless Attacks:	WEP Cracking
	Web Application Attacks:	Cross Site Scripting
		Parameter Tampering
		Cookie Poisoning
		Database Attacks
		Hidden Field Manipulation
Physical Attacks:	Basic	
	EnergyWeapon:	HERF
		LERF
		EMP
	Van Eck	
Password Attacks:	Guessing:	Brute Force
		Dictionary Attack
	Exploiting Implementation	
Information Gathering Attacks:	Sniffing:	Packet Sniffing
	Mapping	
	Security Scanning	

Table 5. Second dimension

Hardware:	Computer:	Hard-disks			
		...			
	Network Equipment:	Hub			

		Cabling			
		...			
	Peripheral Devices:	Monitor			
		Keyboard			
		...			
Software:	Operating System:	Windows Family:	Windows XP		
			Windows 2003 Server		
			...		
		Unix Family:	Linux:	2.2	
				2.4	
				...	
			FreeBSD:	4.8	
				5.1	
				...	
			...		
		MacOS Family:	MacOS X:	10.1	
				10.2	
				...	
			...		
		...			
	Application:	Server:	Database		
			Email		
			Web:	IIS:	4.0
					5.0
			...		
		User:	Word procesor:	MS Word	2000
					XP
					...
			Email Client	...	
			...		
		...			
Network:	Protocols:	Transport-Layer:	IP		
			...		
		Network-Layer:	TCP		
			...		
		...			

The second way is based on the similar idea but is more flexible, because of tree-like structure of the categories. Let us consider the example of this approach ([14]) more detailed.

In grafic representation (see Fig. 1) taxonomy developed by Jeffrey Undercoffer and John Pinkston is a tree. The root of the tree is an intrusion. Dotted line denotes relation "to be a subclass of" and this caption is omitted. Let us remark that this way gives rather detailed description of the attack but it could not show some structural features, attack scenario. This fact is valuable disadvantage because of modern tendency to more and more complex and compound attacks [26], [27], [28].

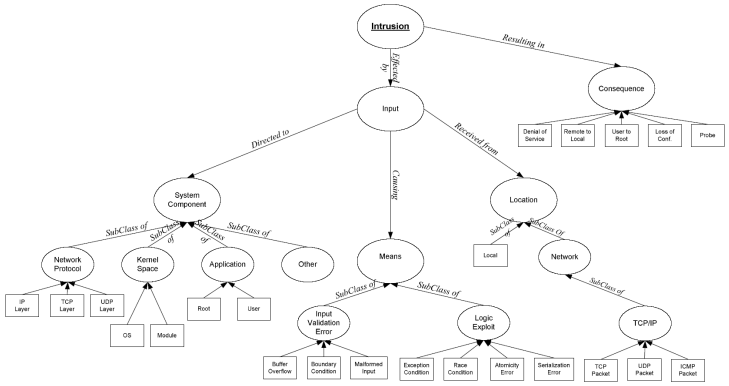

Figure 1. Intrusion

The third way is realized in Howard's and Longstaff's [4]. Its main idea is to introduce the hierarchy of notions (see Fig. 2). The main notion is incident, incident includes notion attack and attack includes notion event. Incident could consist of several attack and attack — of several events.

John Howard and Thomas Longstaff also developed common language for computer security incidents. As said in introduction: "The Common Language Project was not an effort to develop a comprehensive dictionary of terms used in the field of computer security. Instead, our intention was to develop a minimum set of "high-level" terms, along with a structure indicating their relationship (a taxonomy), which can be used to classify and understand computer security incident and vulnerability information."

Taxonomy, developed by these authors is shown at diagram on Fig. 2.

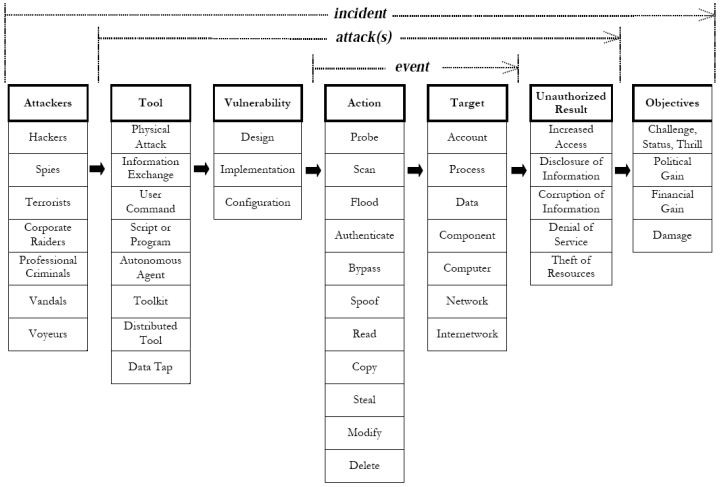

Figure 2. Incident

This taxonomy differs from the others by structural elements: incident, attack, event and by possibility to combine this elements. Such property allows us to describe nonatomic (compound) attacks and consider their scenarios.

3. Proposed taxonomy

The last way of solution described in previous part is developed here. However, unlike the other works, hierarchical structure of relations in tree-like categories is used.

We introduce notion “the stage of attack”, that helps us to describe multistage attacks in a natural way. Common scheme of “attack” is shown on Fig. 3.

Attack consists of several stages. Stage, respectively, consist of several actions. Action consists of events. For example, breaking into computer through pro-ftpd vulnerability ([29], [30]) can be a part of some stage of attack and consists of four events.

All of these four notions expand with tree of subcategories.

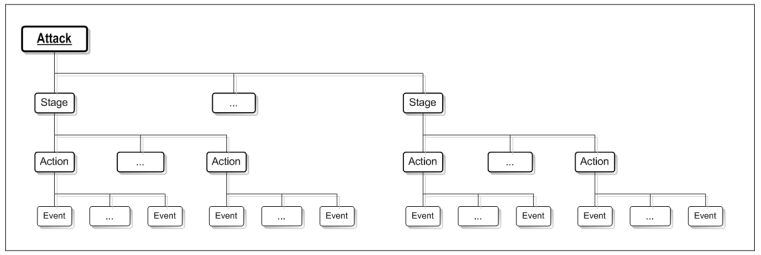

Figure 3. Common structure of attack

3.1. Attack

Attack is a most high-level notion in hierarchy. Attributes of this notion are the global goal, the properties of attack, the object of the attack and the attacker. All of these attributes have their own attributes. All of them comprise the whole tree of *attack*.

Global goal has two components: informational and social. Informational component reflects informational aspect of consequences of impact of the attack on the system. Social component, in contrast to informational, reflects social aspects of consequences of the attack.

Another important attribute is the object of the attack. Subattributes of this notion are the type of the victim system, its physical equipment, the type of the security facilities and external communications of the system.

The next attribute is the attacker. Its main properties are location about the system, rights and privileges at the beginning. If there are several attackers, we can speak about multiagent system ([31], [32], [26]). That’s why number and way of communication between them are included in the list of attributes.

Using the location of the attacker we can pick out four types of attack. First type is the local attacks. Example of the attack of this type is increasing of privileges on local machine with the buffer overflow exploit. Second type is intersegmental attacks. It means that at the beginning of the attack attacker and the victim are inside the same perimeter of the defense (for example, in the same network segment). Third type is external attacks (for example, see [33], [34]).

Fourth type is mixed type of attacks. Usually these attacks are realized by the group of the attackers. For example, attack described in [36].

Last attribute on the diagram 4 is “properties of the attack”.

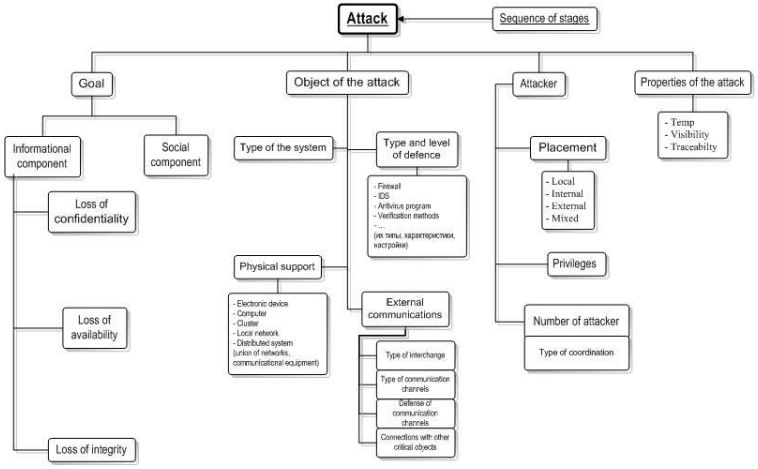

Figure 4. Attack

3.2. Stage

As mentioned above, attack consists of the stages. Stages have their own attributes too (see Fig. 5). For example, one of the stages could be reconnaissance stage, when attacker scans network segments trying to get some information about it. Goal of this stage is to explore internal structure of the victim and to find vulnerabilities in the victim's defense system.

3.3. Action

The stage constists of the actions. Action is, in a sense, an "atomic" attack. It is a minimal step of the attack scenario, in fact. Attributes of the action are: type of the action, subject, consequences of the action and result.

Let us consider attributes of this notion in more detail. Attribute "type of the action" describes directly the action. "Object/group of objects" is an entity, which an action is directed on. "Subject/group of subjects" is an entity, which execute this action. Attribute "consequences" characterises consequences of the action: rights and privileges, obtained by the attacker, information, gained during the action, etc.

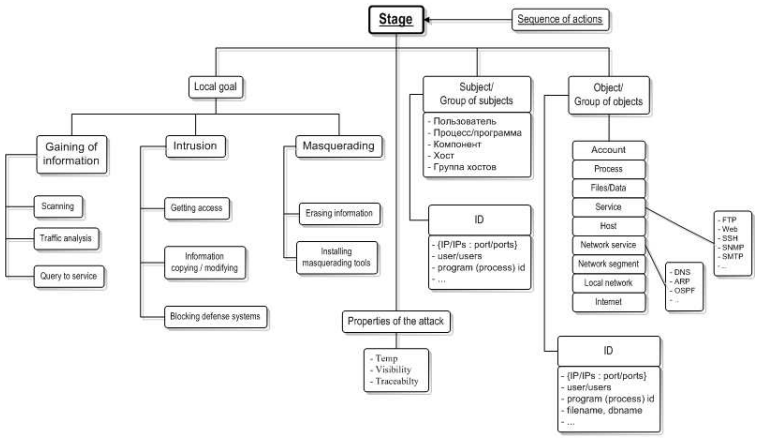

Figure 5. Stage

3.4. Event

Let us remark that from system's viewpoint, action is not an atomic entity. For example, port-scanning is a succession of events, and the sequence of the steps could be various [39], [38]. That's why sometimes it is useful to introduce a new level of abstraction — level of events.

Let us define event as minimal (at a given level of detalization) step of the attack from the system viewpoint. However, this level in most cases is rather detailed and often increases the volume of the description.

Conclusion

Existent approaches to the problem of attack classification are considered. In the introduction we based the urgency of solving problem, discussed possible application areas of the taxonomy and formulated list of rational taxonomie's characteristics.

In section 2.some famous taxonomies are described and analyzed. On the basis of described methods, new approach to the problem is proposed (section 3). Common scheme of the taxonomy and all its components (attack, stage, action, event) is described.

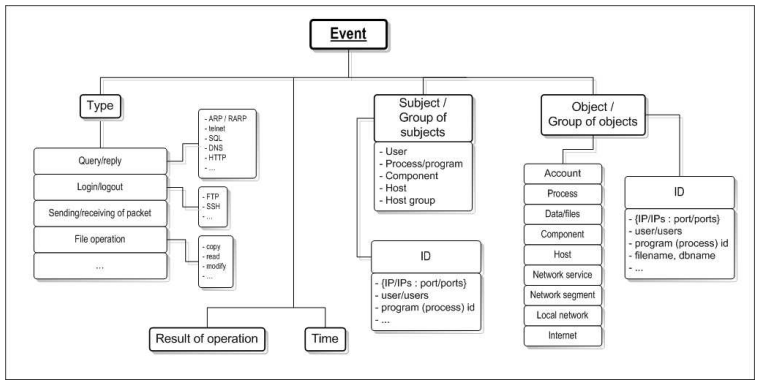

Figure 6. Event

References

[1] CERT/CC Statistics 1988–2005. `http://www.cert.org/stats/cert_stats.html`.

[2] V. A. Vasenin. *Mathematical, algorithmic and software problems of informational security if the internet*. Materials of conference "Mathematics and security of informational technologies-2003", Moscow, 2003, p. 111–142.

[3] V. A. Vasenin, A. V. Galatenko, V. V. Korneev, A. A. Makarov. *Mathematics and software in intrusion detection in large distributed systems*. Materials of conference "Mathematics and security of informational technologies-2004", Moscow, 2004, p. 99–117.

[4] John D. Howard, Thomas A. Longstaff. *A common language for computer security incidents*. Sandia Report, Sandia National Laboratories, 1998.

[5] Wikipedia, free internet encyclopedia.

[6] *Big Soviet encyclopedia*, 1969–1978.

[7] *The IEEE standard dictionary of electrical and electronics terms*. Sixth edition. John Radatz, editor. Institute of Electrical and Electronics Engineers, New York, 1996.

[8] Carolus Linnaeus. *Systema Naturae per Regna Tria Naturae, Secundum Classes, Ordines, Genera, Species, cum Characteribus, Differentis, Synonymis, Locis*. n/a, editio duodecima, reformata edition, 1766. Tomus I, Regnum Animale, 1766; Tomus II, Regnum Vegetabile, 1767; Tomus III, Regnum Lapideum, 1768.

[9] Daniel L. Lough. *A taxonomy of computer attacks with applications to wireless networks*. Ph. D. dissertation, 2001.

[10] Simon Hansman. *A taxonomy of network and computer attacks methodologies*. University of Canterbury, New Zealand, November 2003.

[11] Common Vulnerabilities and Exposures. 2003. `http://www.cve.mitre.org/`.

[12] Edward Amoroso. *Fundamentals of Computer Security Technology*. P T R Prentice Hall, New Jersey, 1994.

[13] John D. Howard. *An analysis of security incidents on the internet 1989–1995*. Ph. D. thesis, Carnegie Mellon University, 1997.

[14] Jeffrey Undercoffer, John Pinkston. *Modeling computer attacks: a target-centric ontology for intrusion detection*. University of Maryland Baltimore Country.

[15] Ulf Lindqvist, Erland Jonsson. *How to systematically classify computer security intrusions*. Chalmers University of Technology, Sweden, 1997.

[16] Ivan Victor Krsul. *Software vulnerability analysis*. PhD thesis, Purdue University, 1998.

[17] Matt Bishop, David Bailey. *A critical analysis of vulnerability taxonomies*. University of California, Davis, September 1996.

[18] James P. Anderson. *Computer security threat monitoring and surveillance*. Technical Report Contract 79F296400, Washington, April 1980.

[19] V. A. Vasenin, A. V. Galatenko. *Mathematical models of large distributed computer systems*. Materials of conference "Mathematics and security of informational technologies-2004", Moscow, 2004, p. 91–98.

[20] C. R. Attanasio, P. W. Markstein, R. J. Phillips. *Penetrating an operating system: a study of VM/370 integrity*. IBM System Journal, 15(1):102–116, 1976.

[21] Giri Vijayaraghavan, Cem Kaner. *Bug Taxonomies*. STAR EAST 2003, Orlando, FL, May, 2003.

[22] Peter Neumann, Donald Parker. *A summary of computer misuse techniques*. In 12th National Computer Security Conference, 1989.

[23] Peter G. Neumann. *Computer-Related Risks*. ACM Press/Addison Wesley, 1995.

[24] Donald B. Parker. *COMPUTER CRIME Criminal Justice Resource Manual*. U. S. Department of Justice National Institute of Justice Office of Justice Programs, August 1989.

[25] Donald B. Parker. *Computer Security Reference Book*, chapter 34, Computer Crime, p. 437–476. CRC Press, K M. Jackson and J. Hruskh, U.S. AssociateEditor Donn B. Parker, Boca Raton, Florida, 1992.

[26] A. A. Grusho, E. E. Timonina. *Role of covered channels in building of defence of distibuted computer channels*. Materials of conference "Mathematics and security of informational technologies-2003", Moscow, 2003, p. 276–283.

[27] A. V. Galatenko, A. A. Naumov, A. F. Slepuhin. *Realization of access control system as pluggable authentication modules*. Materials of conference "Mathematics and security of informational techonologies", Moscow, 2003, p. 237–240.

[28] V. B. Betelin, V. A. Galatenko, A. N. Godunov, A.I. Gruntal. *Providing the informational security of systems on the base of OS2000*. Materials of conference "Mathematics and security of informational techonologies", Moscow, 2003, p. 254–267.

[29] Shai Rubin, Somesh Jha, Barton P. Miller. *Language-based generationand evaluation of NIDS signatures*. University of Wisconsin.

[30] Beyond Security Inc. ProFTPD ASCII file remote root exploit. `http://www.securiteam.com/exploits/`.

[31] A. A. Grusho, E. E. Timonina. *Hostile multiagent systems*. Materials of conference "Mathematics and security of informational techonologies", Moscow, 2004, p. 249–256.

[32] Vladimir Gorodetsky, Igor Kotenko, Oleg Karsaev. *Multi-agent technologies for computer network security: Attack simulation, intrusion detection and intrusion detection learning*. 2003.

[33] I. D. Medvedovsky, P. V. Semyanov, V. V. Platonov. *Attack through the Internet*.

[34] Roelof Temmingh. *Breaking into computer networks from the Internet*. 2001.

[35] Ariel Futoransky, Luciano Notarfrancesco, Gerardo Richarte, Carlos Sarraute. *Building Computer Network Attacks* CoreLabs, Core Security Technologies, 2003.

[36] Sviatoslav Bryanov, Murtuza Jadiwala. *Representation and analysis of coordinated attacks*.

[37] Jelena Mirkovic, Peter Reiher. *A taxonomy of DDoS Attack and DDoS defense mechanisms*, 2002.

[38] Arpit Aggarwal, Ranveer Kunal. *A Comparison of Various Port Scanning Techniques*. Indian Institute of Information Technology — Allahabad, India.

[39] *Examining port scan methods — Analysing Audible Techniques*, Synnergy Networks, 2001.

Application of Network Simulation in Informational Security Field

I. S. Batov

1. Introduction

In recent years there was an increasing amount of attention to application of simulation for informational security tasks [1], [2], [3]. It is concerned with the necessity of investigation of different complex processes, proceeding in huge computer systems. Direct experiments in such systems are often impossible because of system failure risks and costs of experiments. In contrast simulation provides handy and useful tool for security evaluation and research of new technologies. Section 2 of this paper is devoted to review of the problems that seems to have effective solution by means of simulation tools such as ns2 [4], OPNET Modeler [6] and so on. One of the main concerns regarding simulation is the evaluation of adequacy of simulation results compared to ones that can be obtained by real experiment. This problem is also called validation. More thorough discussion of problems connected with simulation validation can be found in section 3.

Modeling of computer networks has been studied for many years and now we can use for simulation a lot of commercial and free distributed modeling software among which are ns2 [4], OmNet++ [5], OPNET Modeler [6], SSFNet [7], and some others. Because of necessity to model variety of standard network components it seems appropriate to use existent software as a basis of simulation tool intended to solve informational security tasks. Section 3 is devoted to requirements such modeling software should satisfy.

Obviously, to build scalable and useful imitation system one should state general tasks such system must solve and define methods of finding solutions. Description of one of the possible levels of abstraction that can be used for modeling network attacks and computer defense systems can be found in section 4. Section 5 illustrate described approach by the example of worm attack modeling.

2. Field of application

One should note that the attacks appropriate for simulation consists for the most part from compound (or multi-stage) distributed attacks. It is worthwhile to simulate IDS (Intrusion Detection System) reaction on compound attack to get detailed characterization of system behavior in complex conditions. For example it is interesting to simulate worms' behavior in a big networks as creation of adequate analytic model for such behavior is a very complex task. Taking into account this consideration we can now state tasks that seems to be appropriate for solution by means of simulation.

1. *Compound network attacks research.* Such tasks arise when one wants to find fast, scalable and effective methods of attack detection and mitigation. For example one can investigate worms' behavior to choose statistics that IDS must take under permanent control to quick detect wide range of worms [8].

2. *IDS testing.* Characteristic example of such task is the investigation of IDS reaction in case when network is loaded with extra traffic from distributed attack in addition to typical working load.

3. *Staff training.* Simulation of network attacks can be used to assess security system reaction that is working on real network prototype and to inform network stuff about possible behavior of IDS in case of either, one or another situation. Modeling results can also give a huge amount of data for neural network training.

4. *Network infrastructure design.* With simulation tool we can analyze different possible variants of topology, service disposition, security policy and so on to choose optimal one [10, 9]. This class of tasks also includes impact assessment for determining how security measures affect system and application performance.

5. *Networks debugging.* Comparing results of network model simulation with results of real prototype testing of this model one can reveal errors in infrastructure realization of the last one. Mentioning such adaptation can be found in [11].

6. *Investigation of research protocols, algorithms, services and technologies.* Simulation results can be used to assess advisability of implementation of new technologies in terms of security and performance.

3. Simulation tool selection

Let us state requirements for simulation software that is intended to become a basis for useful modeling tool. Such system must be able to

- provide possibility of extension with new models;
- carry out different cyberattack scenarios with necessary level of model accuracy;
- simulate huge computer systems consisting from thousands of nodes;
- provide possibility of analytical, graphical and statistical analysis of simulation results.

In addition to stated above simulation software must provide possibility to conduct the following types of verification and validation [12].

- *Verification of model implementation.* This requirement implies correctness of model realization in a programming language.
- *Validation of models of protocols and services.* Models for each of the protocols and services must be correct in the sense that they correctly implement the functionality described in the specification or deduced from the real implementation (as many protocols and services do not have a complete and unambiguous specification) [11].
- *Validation of physical system models.* Simulation tool models not only technologies but also different devices and physical processes, such as signal propagation, error presence and so on.
- *Simulation scenario validation.* This task states the problem of finding appropriate scenario to provide suitable for interpretation results. It deals with choice of representational cases and includes investigation of dependence between initial scenario data (network infrastructure, model network loading and so on) and output results.
- *Methodology validation* Most simulation models use random number generators. To ensure that results from an experiment are statistically valid, it is necessary that the experiment be designed with appropriate attention paid to simulation and analysis methodology to remove statistical bias. [16, 17].

For example well-known simulator ns2 [4] can be used as a basis for the simulation tool intended to solve tasks described above (section 2) as it satisfy to all the stated guidelines [18], [19], [20], [22], [21].

4. Choice of the attack model

Choice of cyberattacks to model becomes one of the most concerns when we plan to use simulation for IDS testing. For exhaustive research one should use a taxonomy of possible attacks. Another problem is accuracy with which one should imitate attack to get valuable results. In this section we will describe possible approach for getting structure and precision of attack models.

- In the first phase it is advisable to find all network objects that can become final goals of compound attack. These objects can be systems with critical databases, monitors of important objects, servers with shared file system and so on right up to network segments as a hole. To reveal the most important attacks objects should by classified according to their significance.
- In the second phase attacks on particular object are classified by its purpose — attacks on integrity, confidentiality and availability.
- All attacks with certain purpose then classified by *implementation method*. By *implementation method* we mean the general idea of attack realization which is often defined by qualification and initial data of attacker. For example to attack integrity of network segment one can use the following methods.
  - to attack main routers of the network;
  - to insert worm in the network;
  - to exploit vulnerability of routing protocol;
  - to attack DNS system and so on;
- Attacks with certain final goal, purpose and method then subdivided to classes according to the *stages of attack*. These stages are picked out in compliance with characteristics that is planned for evaluation. For example the model of worm (section 5) consists from the following stages.
  (a) infected host A is trying to send data to another infected worm to find new victim (host B) or to exploit vulnerability of host B;
  (b) exploit is running on host B;
  (c) worm is sending on victim;
  (d) installing and execution of worm.

  There are several possible sequence orders of this stages:
  - a, b, c, d;
  - a, c, b, d;
  - a, c, d, b.

  The sequence order of the stages is important for testing IDS as attacks alerts will come in sequence in compliance with the sequence of stages. So in this example worm consists from three classes according to sequence order of stages.
- Making the model more concrete one will define *method of attack stage realization*. For example the model of the worm that is described below can have the following methods of victim searching [14].
  - *Passive scanning.* Watching on files and actions of host worm makes assumptions on existing addresses of possible victims.
  - *Hitlist scanning.* Worm uses list of vulnerable hosts that created in advance.
  - *Hidden scanning.* Slow scanning on existence and vulnerability of hosts in network.
  - *Ordinary scanning.* Fast scanning on existence and vulnerability of hosts in network.

  Similarly vulnerability exploitation stage method depends on the concrete vulnerability of the host.

One should notice that to reveal as much multi-stage attacks as possible one should use classification of objects that can become final goal of the particular attack stage. Every technology that is used in a network infrastructure can become such a “subgoal”. It may be TCP, DNS, OSPF, SMTP protocols, Web, VPN, IDS technologies and so on. Finding the set of potential “subgoals” one should find possible attacks that can achieve this “subgoal”. For every such attack its prerequisites and consequences must be formalized so that automatic multi-stage attack generation becomes possible. The example of such formalization can be found in [13].

Reaction of IDSs that is disposed on separate hosts can be modeled in abstract way — by means of probabilities to react on a particular event. Another way is to use more detailed model of host and IDS, such as SIMS [15]. Nevertheless second approach simulate systems with more precision, the first one (high abstraction) can give more adequate results because it deals with probabilistic evaluation of attack consequences rather that its concrete implementation covering in that way a wide set of possible attack realizations. This approach is especially important when one wants to test IDS on attacks that use unknown vulnerabilities. Beside that it is useful as allow us to predict IDS behavior with reduced detection probabilities.

Concrete values of mentioned probabilities can be obtained from analytical models on which IDS is based and from its testing on real systems.

5. Example of distributed attack model

In this section we will describe the model of the worm, that was implemented with the ns2 [4]. This model was created for IDS testing on a wide range of worms and for optimal tuning of security system settings. That is why it contains many variables that must be set to get a particular kind of worm.

5.1. Worm behavior assumptions

We assume that during worm propagation IDS that is installed on particular network host can generate the following alerts. These alerts can be gathered by IDS that is responsible for correlation of particular alerts from different hosts to make decisions about worm intrusion.

1. Infected host A is trying to send data to another infected worm to find new victim (host B)or to exploit vulnerability of host B.

(a) Alert from host A — suspicious outgoing packets.

(b) Alert from host A — suspicious incoming packets. This alert can be triggered by suspicious acknowledgment packets from B.

(c) Alert from host B — suspicious incoming packets.

(d) Alert from host A — suspicious outgoing packets.

2. Exploit is running on host B. Alert from IDS of B — vulnerability is exploited.

3. Worm is sending on victim.

(a) Alert from host A — suspicious outgoing packets.

(b) Alert from host A — suspicious incoming packets.

(c) Alert from host B — suspicious incoming packets.

(d) Alert from host A — suspicious outgoing packets.

4. Installation and execution of worm. Alert from IDS of B.

With this types of alerts in mind we can now formulate actions (or attacks stages) that should be included in worm model to get appropriate for our task abstraction.

5.2. Worm behavior

The worm is starting to propagate with one of the following methods [14]:

- from one network host;
- from several hosts at the same time;
- from several hosts at the different times;

1. One of the victim searching method is chosen — a, ab, ac, ae, b, bc, bd, be, c, ce, d, abc, abd, abe, ace, bce, abce, where a, b, c, d, e — are the following methods.

(a) *Passive scanning.* Watching on files and actions of host worm makes assumptions about addresses of possible victims. Probability of detection of this method by host IDS is the variable of model.

(b) *Hitlist scanning.* Worm uses list of vulnerable hosts that was created in advance. Probability of detection of this method by host IDS is the variable of the model.

(c) *Hidden scanning.* Slow scanning on existence and vulnerability of hosts in network. Probability of detection of this method by host IDS is the variable of the model.

(d) *Ordinary scanning.* Fast scanning on existence and vulnerability of hosts in network. Probability of detection of this method by host IDS is the variable of the model.

(e) *Profile matching.* Watching on files and actions of host worm makes assumptions about standard behavior (profile) of host users. Victim searching begins when worm consider himself well-educated. Probability distribution function of time necessary for worm education and probability of detection of this method by host IDS is the variable of the model.

2. Exploit is running on victim. We consider that a particular worm can use several types of exploits with different detection probability. After successful vulnerability exploitation victim receives worm, which then installed. Every type of exploit may have corresponding type of worm delivery with corresponding detection probability. All these probabilities are the variable of the worm model.

3. After successful worm installation victim becomes an infected host and starts searching for new victim as described above.

6. Simulation results

In this section we illustrate application of simulation to one of the tasks formulated in section 2, namely task 1. We simulate propagation of particular worm that is obtained by the general model described above when the initial variables are set as follows.

- Infected host search victims with the ordinary scanning method (subsection 5.2), at that the probe packet is send on particular port (the number of port is fixed). In case there is exists vulnerable application on this port, victim answers with acknowledgment, otherwise it do nothing. Size of probe packets[1] is 1000 byte, delay between packets sending equals 0.08 seconds.
- Receiving answer infected host sends to victim packets with exploit (total size of packets 5000 byte) and packets with worm (total size of packets 100000 byte).
- Receiving packets with worm victim turns into infected host and begins searching new victim on the expiry of 0.1 seconds.

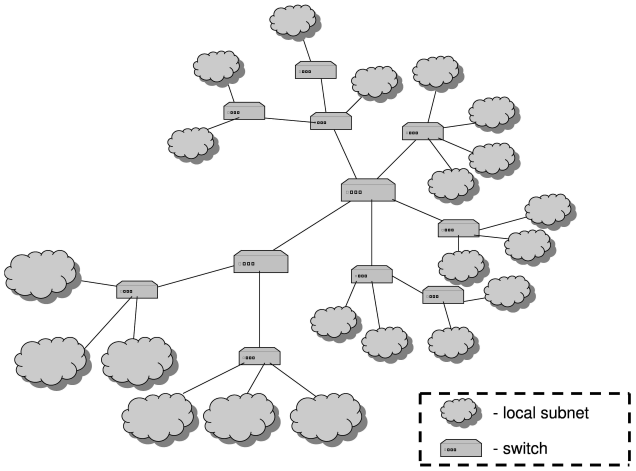

Figure 1. Network model illustration

Attack described is rather representational among distributed cyber-attacks. Furthermore, investigation of such attack requires experiments

[1]Hereafter packet denotes message of network layer protocol (IP). Maximum size of packet equals 1024 byte.

on big networks, creation and tuning of which constrained with a big inputs. Utilization of existing networks for such experiments can lead to system failures and expensive recovery process. Unlike this simulation provides useful and effective tool for carrying out such experiments on networks of arbitrary size and complexity. In our simulation we used models of two local subnets connected with 100 Mbit channel. Topology of this network is depicted at figure 1. Here we can see first subnet (137 nodes) in right up corner of figure and second subnet (123 nodes) in left down corner. Topologies of the local subnets were created with topology generator tiers2.1 [23]. Nodes in local subnets are connected with 100 Mbit channels. Propagation delays in subnet channels are at most 0.001 ms. Propagation delay in channel connecting subnets is 5 ms. All switches has DropTail as queue serving algorithm.

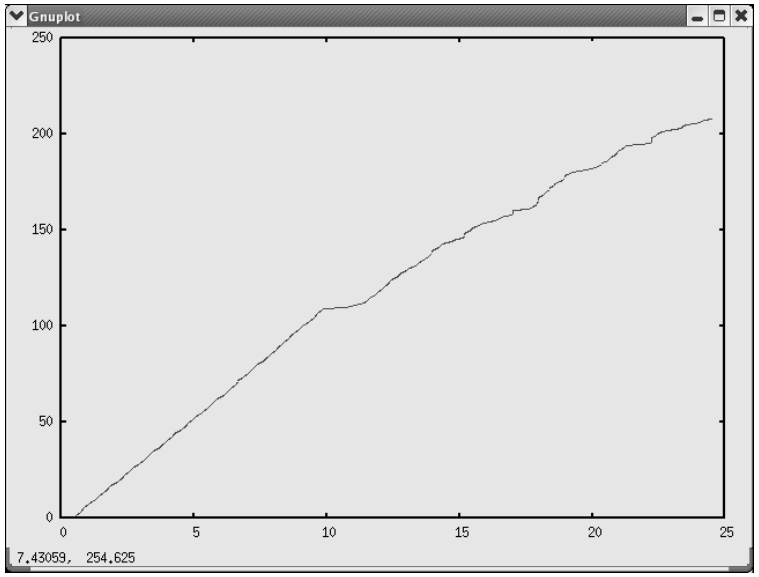

Figure 2. Increase in number of infected hosts (number of nodes/seconds) for the first scenario

Simulation of the described model allows us to evaluate almost all charateristics that is available for evaluation in the experiment on real network. However complete and detailed discription of all significant characteristics that can be obtained by such experiment is out of the scope of this article. That is why we only illustrate possible outcome

of simulation describing one of the characteristics, namely increase in number of infected hosts. This characteristic allows us to evaluate the danger of this particular kind of worm. We can also evaluate an advantage attacker gets when he uses several nodes for worm start. To estimate this we simulate to worm propagation scenarios — in the first worm began propagating from two network nodes (each subnet has its own node) and in the second there is only one node for the hole network. Figures 2 and 3 show graphics for number of infected hosts' growth for the first scenario (Fig. 2) and for the second (Fig. 3).

Presented figures show that there is a little difference in infection rate for two described scenarios.

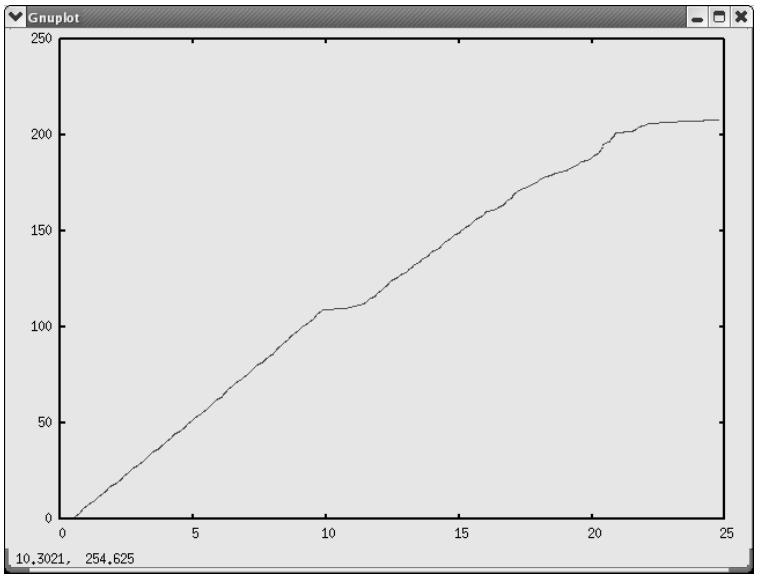

Figure 3. Increase in number of infected hosts (number of nodes/seconds) for the second scenario

In conclusion we can say that experiment described above shows that

- it is possible to model complex distributed attacks on simulation tool that is now developing in the the context of this work;
- simulation can give results that is effective for researcher;
- with the developing simulation tool one can carry out experiments that is impossible or very hard to do on a real equipment.

7. Conclusion

The paper addressed problems of application of network simulation to the field of informational security problems and proposed general approach to such adaptation. Review of possible tasks that can be solved by means of simulation presented.

Approach for selection of attacks to model and level of their abstraction is recommended. This approach was illustrated by example of worm simulation.

Results of this work are description of tasks for simulation and methods for attack modeling for building the simulation tool intended to work in informational security field. The future work will be focused on development of simulation tool and its extension with the the new models of cyberattacks, defense systems and user behavior patterns.

References

[1] Vasenin, V. A. Problems of mathematical, algorhithmic and software means for enforcement of information security in the Internet. Materials of MaBIT-03, MCCME, 2004, p. 111–143 (in Russian).

[2] Nicol, D. Modeling and Simulation in Security Evaluation, *IEEE Security and Privacy*, vol. 03, no. 5, p. 71–74, September/October, 2005.

[3] Second Workshop on Ultra Large Networks: New Research Directions in Modeling and Simulation-based Security, 2003.

[4] SAMAN, CONSER, ACIRI, Network Simulator 2. `http://www.isi.edu/nsnam/ns`.

[5] Varga, A. OmNet++. `http://www.omnetpp.org`.

[6] OPNET Modeler. `http://www.opnet.com/products/modeler/home.html`.

[7] Renesys Corporation, SSFNet. `http://www.ssfnet.org`.

[8] Aranya, A. In Search of a More Insidious Worm. Computer Science Department, Stony Brook University / Aranya, A.; Callanan, S.

[9] Carrier, B. Impact of network design on worm propagation / B. Carrier, S. Jeyaraman, S. Sellke; Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette, IN 47907-2086.

[10] Surdu, J. R. Military Academy Attack/Defense Network Simulation / J. R. Surdu, J. M. D. Hill, R. Dodge, S. Lathrop, and C. A. Carver, Jr. Department of electrical engineering and computer scitnce. United States Military Academy.

[11] Heidemann, J. Expanding Confidence in Network Simulation / J. Heidemann, K. Mills, S. Kumar USC/Information Sciences Institute Research Report 00-522, April 2000, submitted for publication to IEEE Computer.

[12] Bagrodia, R. Position Paper on Validation of Network simulation models. Bagrodia Rajive, Takai Mineo Computer Science Department UCLA. http://pcl.cs.ucla.edu/papers/.

[13] Ning, P. Building Attack Scenarios through Integration of Complementary Alert Correlation Methods, Cyber Defense Laboratory Department of Computer Science North Carolina State University.

[14] Nazario, J. Defense and Detection Strategies against Internet Worms. 2004 ARTECH HOUSE, INC. 685 Canton Street Norwood, MA 02062.

[15] Garg, A. SIMS: A Modeling and Simulation Platform for Intrusion Monitoring/Detection Systems / Garg, A.; Upadhyaya, S.; Chinchani, R., Kwiat, K.; 2003 Summer Computer Simulation Conference, SCSC 2003, July 20–24, 2003, Montreal, Canada.

[16] Pawlikowski, K. Do Not Trust All Simulation Studies Of Telecommunication Networks. Department of Computer Science, University of Canterbury Christchurch, New Zealand.

[17] Jeruchim, M. Simulation of Communication Systems : Modeling, Methodology and Techniques (Information Technology: Transmission, Processing and Storage), second edition / Michel C. Jeruchim, Philip Balaban, K. Sam Shanmugan; Plenum US; 2nd edition, 2000.

[18] Georgia Institute of Technology PDNS — Parallel/Distributed NS. http://www.cc.gatech.edu/computing/compass/pdns/.

[19] Huang, P. Enabling Large-scale Simulations: Selective Abstraction Approach to the Study of Multicast Protocols / Polly Huang, Deborah Estrin, John Heidemann; USC/Information Science Institute University of Southern California.

[20] Dutta, D. Faster Network Design with Scenario Pre-filtering / Debojyoti Dutta, Ashish Goel, and John Heidemann; in proceedings of the International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, pp. 237–246. Fort Worth, Texas, USA, USC/Information Sciences Institute, IEEE. October, 2002. http://www.isi.edu/~johnh/PAPERS/Dutta02d.html.

[21] Estrim, D. Network Visualization with the Nam, VINT Network Animator / Deborah Estrin, Mark Handley, John Heidemann, Steven McCanne, Ya Xu, Haobo Yu; *IEEE Computer*, 33 (11), p. 63–68, November, 2000. http://www.isi.edu/~johnh/PAPERS/Estrin00b.html.

[22] The *ns* Manual. http://www.isi.edu/nsnam/ns/index.html.

[23] Calvert K. Modeling Internet Topology / K. Calvert, M. B. Doar, E. W. Zegura; IEEE Communications Magazine, June 1997. http://www.geocities.com/ResearchTriangle/3867/sourcecode.html.

Advanced Research Workshop-Round Table Discussion "Unconventional Information Warfare and the War on Terror: Critical Issues and International Cooperation"

The Idea of the Advanced Research Workshop-Round Table Discussion "Unconventional Information Warfare and the War on Terror: Critical Issues and International Cooperation"

Why this workshop?

Safeguarding information security is accepted as a critical component of national security. In the past decade, leading industrial countries have made large investments designed to protect the integrity of critical infrastructures, and to raise the overall level of awareness for the need to protect national informational assets. In the military realm, Computer Networks Operations (defense and attack) are an active area of development, and Information Operations (IO) have become the standard feature of military doctrine for most advanced militaries. And yet, while investment in Information Security has increased dramatically, the question remains open whether these measures are adequate or indeed appropriate to safeguard national security, and whether they are focused on the right set of "threats".

For example: Are present day Information Security strategies effective or appropriate against the threat posed by asymmetric actors (such as "terrorist" groups)? Asymmetric actors have the characteristic of not being bound to prevailing rules or conventions of war (and thus the difficulty in deciding on an appropriate response or target). In many case their operations and survival depend on exploiting unexpected and unintended opportunities, and thus adopting "unconventional" approaches in the pursuit of their aims (which may tend towards maximizing psychological, rather than material effects). While much attention has been focused on the so-called "electronic Pearl Harbor" scenario, a massive coordinated takeover of critical information systems, the energy grid or air traffic control system — this scenario remains in the realm of theory. At the same time, asymmetric actors have used techniques of information warfare and other more "unconventional" strategies in the pursuit

of their causes. ICTs are used to raise funds, provide command and control, wage propaganda campaigns and leverage extended diasporic and other social networks of sympathizers and supporters. As such, conventional approaches to Information Security may at best be ineffective in anticipating the "unconventional" uses to which ICTs are used by these actors. At worse, current approaches may be missing the point and building the electronic equivalent of a "Maginot Line" in cyberspace.

At the same time unconventional security threats are not exclusively bound up with the challenge of "terrorist" or criminal groups, or inter-state competition. Information Security doctrines must increasingly contend with a global ICT environment that is no longer controlled or controllable at the national level. The design and production of IT systems and software has steadily become trans-national and driven by commercial rather than patriotic considerations. Software used for critical government applications is often not designed or coded within national boundaries, or, because of intellectual property right (IPR) issues — easy to audit or review. The expertise needed to ensure compliance often dose not exist in the public sector, and the option of developing proprietary software or hardware solutions is often too heavy to bear for many countries. Moreover, given the strategic significance of these technologies to technical intelligence-gathering, these issues represents a potential liability to national information security, and may define an "unofficial" limit on the scope for international cooperation.

These issue are complex, and yet given their centrality to security, require consideration, and thoughtful reflection. While it is easy to adapt existing security paradigm to a new realm, it is much more difficult to step outside the bounds of convention and embrace the possibility that new paradigms may be needed. In an era where technological dependence and interdependence are an increasingly important aspect of national security, thinking beyond conventional boundaries is an important and critical exercise in "due diligence". At the same time, because of the inherent vulnerabilities inherent to slow-to-adapt actors (such as states) and "zero cost" options available to asymmetric actors, international workshops examining the "unconventional" side of information warfare have remained rare, as have multidisciplinary discussions that bring together scholars and practitioners.

Objectives of the workshop

The objectives of this workshop are to provide an opportunity for practitioners and scholars from NATO and CIS countries to focus on three core themes:

- Defining the threat of "cyber-terrorism" in the context of the present day "war on terror";
- Assessing the significance of other forms of "unconventional" threat to national Information Security;
- Comparing experiences, limitations and "Lessons Learned" of cooperation in combating threats to national Information Security (from an international perspective).

The workshop will be an expert, by invitation only event, bringing together practitioners and scholars from NATO and CIS countries.

Terrorism and Democracy

J. Ryder

I've been asked to say a few words about the relation of terrorism and democracy — a rather large topic, so I will limit my remarks to one aspect of the problem, which are the potential threats to democracy of efforts to defend against and combat terrorism. Couple points on background first:

1. These remarks will consider the question, consider the problem from the American perspective. I will make some judgments about the relation of terrorism or struggle against terrorism in relation to democracy that bear on American context. It would be interesting to have others of you to comment how this might bear on Russian context.

2. The details of this kind of analysis, certainly if we will develop this kind of analysis more fully, invariably will depend on what we mean by democracy. Do we mean something simply formal or procedural, do we mean something more substantial. In some of my work I distinguished between "thin" democracy and "thick" democracy. But basically one can understand democracy as primarily a formal structure whereby there are elections are more or less free and open and that is it. And there are many people including many in the leadership in the United States who regard democracy this way. But that is rather a thing that many of us could have and many of us do have a much a deeper, thick notion of what democratic society, democratic situation consists of and depending on which of those and others within that range or range between them which use of the democracy it takes when we draw different conclusions.

But while those background points are being made let me divide these remarks into two parts. One has to do with democracy and terrorism in internal/national circumstance and then external that is with prospect to foreign affairs/foreign policy.

With respect to internal situation (and remember the context is here American) Bush' Administration policies after September, 11 have run rough-shod over a number of democratic rights and liberties. It generally had thrown into question for a lot of Americans and others around the world what just democracy means we name democracy we undermine a good deal of it. And just to give a quick example for those of you who

are familiar with this kind of ethnic profile on development policies to try to identify the potential terrorist before they are able to organize, before they are able to strike, we will regenerate the opinion based on physical appearance of means or of national origin for example. The suspension of rights of the accused which is a controversial matter in my own country, question of access of people who have been arrested for a suspicion in terrorist activities or support of terrorist activities, the access to defense attorneys, by extension defense attorney's access to information that prosecution has, that government has in various cases. Some of these issues as you probably know have gone into our own federal courts and even the Supreme Court. Bush administration has been reinvented in some cases and compelled by the courts to readjust its behavior because much of what it has tried to do has undermined too much of what we regard to as democratic rights. There were issues of surveillance. One of the characteristics of the first version of the Patriot Act that was passed right after 9/11 as many people has judged to was that the government assumed to itself the right to have access to people's reading habits, for example, in libraries and so on. And this made a lot of people nervous as you might imagine since we have generally operated on the assumption people were entitled to read what they please and was none of the government's business terrorism or no terrorism. There were many other matters, like the list kept of many organizations as terrorist organizations and what damage that does to the question of the preemptive association. The tension here accords is between liberty and security. I don't pretend that it is a tension easily resolved. If it were there wouldn't be much of a tension. This is a serious problem. And it is complicated by the fact that in a less or more democratic society this is quite possible for the majority for people in the name of defending their own liberties and democratic rights to sacrifice these liberties in the name of security. The problem here is a version of what Tokwill and John Stuart Mill and many others in 19th–20th centuries have called integrity majority. It is not he majority of citizens who are suspected of terrorist activity, it is a very small minority but one can make the claim and many have that the democratic character of the society is marked not so much by the rights and liberties that he majority gives itself because that is rather easy. The democratic strength of the society is characterized by the democratic rights and liberties of the majority accords to everybody and in a case where people begin to run scared when they are nervous about security; it is quite easy for the majority to sacrifice their rights and liberties of a minority. And that is a very serious threat to a democratic character of a society. So the relation

between these two that is terrorism and democracy, in internal context is a rather serious problem to which I have no answer.

If we turn to external question, again we do see that the American Administration Foreign Policy in the name of fighting terrorism turned out to the fundamentally credit in number of important aspects. The general question here is whether democracy can be opposed/imposed. And again we need to considerate the issue of what we mean by democracy. Let me now say a word before going in it about this term we use the thick conception of democracy. I think it is worth taking seriously and it does have a variant on what we end up thinking on the relation between democracy and terrorism.

A thick democracy is a democracy that is far more than simply system of political parties, competing in open and periodic elections, that is more than just a formal procedure of political democracy. What I mean by a thick democracy is:

(1) a society which individuals, groups or communities consciously communicate in pursue common interests with others;
(2) a society which is driven not by ideology but a willingness to experiment with new ideas and solutions to social problems. It is a society not driven by ideologically but experimentally in its approach to problems;
(3) a society that communicates with, collaborates with in pursuit of common interests with those beyond its borders. For foreign policy situation, for external situation that may be the most crucial consideration. A democratic society is the one that makes a sustained an effort to communicate, collaborate and find points of commonality with those beyond its boarders.

In pursuing its policies in Afghanistan and Iraq and its interactions with Iran, Korea, China, Russia and many other nations the US currently behaves more like empire power rather like democracy. And again I am not the first to make this observation, there were a couple of very interesting books written on this topic only last year. A couple of books were arguing the virtue of American democratic imperialism. But I will argue that imperialism and democracy, certainly democracy in the thicker sense, are not compatible. So to the pursuit of one undermines the other. I would also argue that it is possible in fact to behave internationally in such a way as to advance democracy even in its thicker sense and it would be a better part of wisdom of our international community to begin to think in terms of maintaining democratic values as perhaps in a long run the best way we have to undermine the foundations of terrorism and to defend ourselves against it.

Devoted to the Creation of the Regional Informational-Psychological Zones Doctrine

V. I. Tairyan, E. I. Tairyan

The modern stage of the society development is characterized by increased role of the informational sphere, which is represented by the summation of the following components: information, informational infrastructures, subjects which perform information collection, forming, use, and distribution and regulation systems. The informational sphere is the backbone factor of social life. It actively affects the following constituent elements of national security: political, economic, defensive and others. National security essentially depends on the information protection, and that dependence will increase in the course of technological progress.

State information security is the protection of national interests in the informational sphere, which are determined by the summation of balanced individual, society and state interests.

Human development shows that the security of a separate state depends on the regional stability. So we should consider state security in the network of regional security.

The last resolutions of NATO devoted to the problem of international terrorism call to the world community to be in earnest about that problem, otherwise, the consequences may be catastrophic.

We have to detect the reasons and centers of an international terrorism and find out the ways of its information-physiological neutralization.

To achieve that, certainly, a network of the regional information-physiological security zones should be created, by the collective security provided international structures. The general goals are cooperation in the regional stability and security; detection, localization and neutralization of an international terrorism's reasons and centers.

Below we discuss some political aspects of the necessity to create regional information-physiological security zone's on the example of South Caucasus.

> *"Balkans & Small Asia have the most important strategic positions in the world. They are the core and the Centre of the Old World, they separate and at the same time connect three continents: Europe, Asia & Africa... They are situated in such a place where from they can threaten and make assaults against the three continents."*
>
> J. Bucker.

It took decades for the obvious truth, which was said by an ordinary English politologist in the 20th century, to become equally obvious and vitally important today.

So, the South Caucasus and Central Asia are at last declared NATO's priority zones. And the countries of the South Caucasus are involved in the EuroUnion's "Widened Europe: new neighbours" program. So, a quite obvious question rises: to what extent are these steps adequate to the European security provision in the nearest years and whether NATO lingers, wanting to put a stable basis not on words but on concrete actions.

It's quite evident that the South Caucasus region is nothing but the opposite side of the Balkans and it is not equally "untidy": series of unsolved international conflicts, destroyed economy, and as a consequence social tension inside the countries, various confessional and ethnonational contradictions, etc.

All these contradictions will finally be settled by the states themselves, however it will take years and here the question arises — does the EuroUnion have time? According to the unclear consideration of geopolitical processes the answer is negative, and that's why:

Russia. Naturally, Russia fully realizes the region's strategic role, it realizes it not only as its base, but also as "critical" strong point for the whole Euroasian continent. The fact is just that Russia went away from the region, if not forever, but at least, for a long time. It's not difficult to understand it: if Russia was to be "the host" in this region, it would never allow ANYONE to take away different strategic recourses from the region, not talking about gold, oil; even if we don't mention the creation of the transport corridors (TRACECA) etc.

In fact the presence of Russia, in the region only creates the illusion of its presence, from one hand, for concealing tiny, unimportant interests and for "luring" NATO, which doesn't want to take responsibility on itself to regulate the region conflicts and to create in the region such a security system, which will not only allow to use the region as a platform against Europe in future, but also to make it the most important bastion of Europe.

It is worth mentioning here, that surely NATO has mass of its "own" problems — the Balkans, however one can't cure the teeth, if the gums are also sick, and no Colgate will ever help it, as Russia just can't solve the "Caucasian" problems any more.

The sooner NATO gets convicted of it, the quicker they will give up the most unimportant programs in range of "Partnership in sake of Peace", which are lead by the individual plans.

Such "cooperation" is not only useless, but also harmful, as it creates some illusions in this region's countries, allows them to speculate in their own interests, prolonging the acceptance of necessary concessions for reaching the compromise decisions, which must objectively be good for everyone and become a good basis for hospitable relationships, in future, become the guarantee of general interdependable security and stability.

And, finally NATO should once again analise for them what is the meaning of the recent reforms of Collective Security Treaty Organisation (ODKB): is it the creation of real mechanisms of collective security provision or in reality it is invitation for the goodbye party!

So others will take the same place!

Russia's leave from the region is not unexpected event for those who are not only waiting for a suitable time to "capture" such a good slice, but also made much effort to lead Russia to the decision of such a question.

The formation of two geopolitical and geostrategic macroregions of Big Europe and Big Nearest East has two scripts which are not so much desirable for Europe.

1. "Goble's Plan" or Turan etc.

This plan could first be realized in range of globalization carried out by the USA: on one hand, getting stuck in Iraq the USA won't be able any more to finish up its "Megaprojekt" in Iran and create optimal conditions for Turkey's entrance from the South Caucasus to Central Asia for getting from the USA the "gamestarting stick". On the other hand, the events in Iraq made USA understand at last the simple truth in the relationship with Turkey: Turkey's dual macro regional location doesn't at all mean the two models of its political behavior. Turkey is a "beast" too strong and aggressive for obeying its trainer; if it feels that the trainer is more delicious than the offered rabbit, it will never linger with the choice.

Analogical "failures" of the USA in Israel and in Pakistan allow us to come to a conclusion that even if the USA doesn't reject the plan of the region's Turk-making, in this case the overestimation of its real possibilities and its connection with the reality, the search of new partners

etc. will require from the USA too much time, which it doesn't have any more.

2. The second (the most realistic) script — China's "coming" to the region of the South Caucasus.

Although the USA managed to solve some of its strategic tasks of taking control over Central Asia and perspectivelly to stop the process of China's getting to the East like a very good advertised show, anyway this was the end of everything.

What is it — an intermission or finish?

Alas, if it is an intermission it will last for years, and China will not give anyone the gradual strength. Besides, the perspective of "Great China" recreation is not a ghost any more, and it is still unknown where the Great China Wall will again be built.

The perishing commune-cosmopolitism in Russia gave China a great impulse to the organization. Having gotten through the mason ideology of cosmopolitism with the loss of 50 million human lives in the cultural revolution, China borrowed one of the best systems: mass operations leading system with a profound impact on the consciousness of the masses and the experience of creating phychocomplexes of behaviour with the impact from subconsiousness.

And as a result China is already a quite wealthy state and possesses more than sufficient recources for reaching almost all the aims.

It must be also mentioned here that one shouldn't expect from China the so-called brilliant shows analogical themes, which the USA demonstrates. There won't be any China "boom", as it has already begun, having borrowed the method of "crawling agression" from the Turks, the China people under the mask of migrants, traders and refugees, and gradually spreading all over our planet, but they mainly try to concentrate, to remain in those places where it is functionally profitable for China's patriots. One of the most vivid examples of this fact can be brought the recent events, when on the financial attack against the Soros fund against yuan the Chinese answered not only with a very good defence of national currency, but also using its diaspora in the USA, lead the answering attack and practically ruined the Soros fund.

That's the reason why China doesn't hurry: time is its best ally. And what is now, being unexperienced, we assume, that we managed to persuade NATO to take more radical steps and to make influence on the South Caucasus we shall make an attempt to focus on two more aspects: the first one — what shall we begin with?

Information formats the most important component which comprises in itself the types of human thinking, its psyche and consciousness, the soul state and the morality formula, attitude towards the reality and to the abstraction, materialistic and celestial, Earthly and divine.

Consequently, by defining the primary component and the chain of stocks of security zone creation on the basis of South-Caucasian region, it is necessary to take into consideration. That spiritualism influences on consciousness, the consciousness formats the thought, the thought is expressed by words and deeds, which finally is realized by integral actions, which define the person's, state's or region's potential.

Consequently, this primary component (the fundamental one, in fact) can and must be the creation of information security zone, general and unique for the whole TransCaucasus, but in perspective its connection with the security system of Iran and that's the reason:

Iran. Having formed the historical thinking in its nation's minds and influencing its consciousness by a stable modeling, the Iran elite constantly try to give rebirth in its nation the forgotten legend heroes which is just the final element of the strengthening the connection of time in the mind of this nation. And, as a result, at the proper time, millions of Iranians by the challenge of mullah will be turned to "shakhids". Here, where it is the most important argument of Iran, when in the conversation with the USA it signs out that Iran is not Iraq. And it's quite well that "wise Europe" prefers to widen business, trade and other relationships just with Iran and instead of all this the USA invites it to search nuclear arsenals.

Europe's such far-sighted position can be supported by NATO with the creation of a system, comprising informational security zones of South Caucasus and Iran concerted with each other and which correspond to EuroUnion's interests.

In conclusion we express our high respect to the head of Russian — Armenian (Slavonic) state university professor A. R. Darbinyan for cooperation in the conducted in RAU scientific investigations in the solving information security problems.

References

[1] V. P. Cherstyuk. An information security problems in the modern world. 18.04.2003.

[2] A. A. Strelcov. Providing an information security of the Russia. Theoretical and methodological basics.

[3] A. A. Salnikov, V. V. Yachenko. A methodological problems of an cybert-errorism's opposition. 26.03.2004.

[4] E. N. Mochelkov, V. A. Nosov. A deal of Consortium work group "An informational technology influence on the national security" in 2002–2003. 25.06.2003.

[5] A. V. Manojlo. An information-psychology war as a mean for reaching political goals. 2.01.2004.

[6] A. V. Krutskih, I. L. Safronova. International collaboration in the information security sphere. 18.04.2003.

[7] A. V. Manojlo, A. I. Petrenko. An information-psychology security of the social-politic relations in the modern society. 5.03.2004

[8] A. V. Krutskih. War or peace: an international aspects of the information security. 26.03.2004.

[9] Tairyan Vasiliy. Humanitarian problems of the information security. In Proc. of the NATO ASI — Armenia, Nork, Armenia, 2005 (to be published).

Educational Aspects of Ensuring Information Security during the Growth of Terrorism

A. N. Kurbatsky

The task of building an informational society, the growing role of information and information resources in the development of the society and the state, force us to consider information security matters as the major problems in the forefront. The transition of the information into the category of a major society resource instigates an active struggle for the possession of this resource. As a consequence, we witness dramatic growth in the importance of information wars and information weapons. They have experienced rather unexpected development in the beginning of the 21st century, after drastic growth in the irrationality of terrorism. A certain kind of destructive democratization in information technology became obvious.

Terrorism forces the society to balance between freedom and safety. In the XVIII century Benjamin Franklin wrote: "He who is willing to sacrifice freedom for security deserves neither freedom nor security". Rapidly growing danger from international terrorism undermines the authority of the state.

For instance, citizens observe inefficiency of government in defense from international terrorism, and cease to regard the state as the main and irreplaceable form of public organization. It is especially dangerous when this view merges with the current growing tendencies of ethnic self-identification in various groups.

These processes, as a rule, are explained in "historical researches of the past, language, borders". Current world events and tendencies support Winston Churchill's words: "If we open a quarrel between the past and the present, we shall find that we have lost the future". The most common form of ethnic self-identification is terrorism (national terrorism).

The examples here have already become classical: the Irish Republican Army (IRA) in Great Britain; Front of national liberation of Corsica in France (FLNC); Basques' organization "Euscadi ta asakata-

suna" (ETA) in Spain; the Albanian terrorism on the Balkans, etc. It has become practically impossible to clearly distinguish terrorists from freedom fighters.

Who knows, maybe the Spanish philosopher Hose Ortega-y-Hasset was not far from the truth when reflecting about national states: "So, it looks like the only thing left to us is to disregard the old, habitual, misleading concept of the national state with the traditional three whales (language, blood, motherland territory) as the main supports of the nation, that allegedly create the foundation of it, and start looking at the them as the primary obstacles on the way of the state's formation ... We have to be daring enough to see the answer to the riddle of the national state in the very character of the state, something that is inherent to it as a state, in its policy, instead of the extraneous foundations of biological or geographical properties".

The growing tendency of ethnic self-identification, that became especially intensive in Europe after the USSR's collapse, overlapped the process of cultural and religious self-identification of violently growing Islamic Diaspora, as long as cultural assimilation of Muslims by the Western civilization, even on its own territory, has not occurred.

European civilization (and Western as a whole) has proved to be practically helpless against the scheme according to which the initiative, preparation and direct performance of terrorist acts are carried out actually from within. Terrorist acts like those in London during the summer of 2005 are an obvious visualization of this scheme, which has demonstrated that there exists a real, internal infrastructure and a demand on terrorist acts.

American scholars from Washington Research Institute (Nixon Center) have found out that the largest Islamic terrorist groups in Europe and Northern America originated not in the Middle East or developing countries, but in the West. As a basis for the research, they used a database where they registered about 400 terrorists, who within the period from 1993 to 2004, have been taken to court, sentenced to prison, or killed in Europe and Northern America. From the facts in this database they drew the following conclusions: less than half of registered terrorists were born in the Middle East; 41% have EU or US citizenship; and 36% are from the Maghrib countries; only 17% of terrorists came from the Middle East, and 3% came from Asia.

We face a situation, where the actual and potential terrorists are not from the desolate, poorest Muslim layers of society, but are young people with some level of educational background. The most dangerous thing is that in the very near future there is a possibility that the mass of recruits

for the international terrorism might come not only from Muslim youth, but from traditional western (in the broad sense of the word) youth. Let me explain why, in our opinion, such a tendency is possible.

It is common knowledge that terrorism never exists in vacuum. Even the medieval organizations, which in a certain sense can be considered the ancestors of modern terrorists, such as Isma'ilite, tried to attract the attention of contemporary society to their actions. For example, they would declare out loud in a market place about the latest murder they had committed. As early as the last century, terrorists understood the ultimate importance of newspapers, because terrorism always "cooperates" with structures of the society and the state.

International terrorists have managed to involve *the information field* and they are using it with great efficiency.

- Today terrorist acts in the "traditional" sense are used only as *elements of information technologies*, and actually terrorist acts are no more the purpose and goal of terrorism. Thus, we can observe the growth of the "quality" of terrorist acts as *information products* due to the fact, that terrorist acts are losing their primary role of *local action tools*, and pass into the category of *global management tools*. Therefore basic changes have occurred in the list of national security threats, namely drastic decrease of the society's tolerance threshold to information attack — both external, and internal.

We cannot live without newspapers, TV, and the Internet, but after watching the endlessly repeated "TV news live pictures" after terror acts (let us recollect the coverage of events of 9/11 in 2001 in New York, explosions in the underground and "Nord-Ost" theatrical performance in Moscow, tragedy in a Beslan school, explosions in Madrid, etc.), a number of reasonable questions arise:

- Do or do not the information society, and mass-media as its main tool, "provoke" the terrorist element that can exist only when it is being "covered"?
- Does mass-media sometimes assist terrorists in creation of exaggerated fear?
- Does mass-media always oppose terrorism, even in the case when terror acts might serve as a fascinating informational prerequisite for creating a sensational show and thus satisfy some ambitious reporters? A terrorist act is a show, and a show, taken out of context, is the mass-media priority element.

Terrorist organizations try to impose on society a new model of economic and political regulation, in which the informational and physical

violence over an individual becomes the basic *method of management* of the society and its institutes.

All this is accompanied by "TV news live pictures". Even Zbignev Bzhezinsky in an October 11, 2005 "Los Angeles Times" article states: 'One is not born a terrorist, one becomes a terrorist — under the influence of concrete events, personal experience, concepts, phobias, national myths, historical memory, religious fanaticism and intentional "brainwashing"'. We can add: under the influence of a "TV news live picture". And now let us take a look at the main occupations of school-age children. One of the questions now actively discussed in the Internet is: "What will come of our children who are growing up in the conditions of the 21st century technologies?"

Today youngsters spend more and more time behind a computer, preferring the virtual world to the real one. The virtual world created by aggressive computer games is often a much worse kind of world. But now "live pictures" on TV screens and computer monitor screens are practically the same. Murder and violence are becoming a life norm.

The existence of computer game addiction, similar to drug addiction, is still doubted by many scientists and experts, but what is doubtless is that the quantity of young people, taking great interest in the search for virtual reality is growing at an alarming rate. The person, spending long hours in such an environment, transfers its laws to the real world and starts to feel more vulnerable. Many psychologists confirm that games with a lot of violence, form aggressive thoughts in the mind of a person. However, such aggression does not necessarily show itself immediately. It accumulates gradually, sometime taking a long time. An adult today is capable of drawing a line between the virtual and real worlds, and can do it more or less precisely. Children who have not yet received sufficient conceptions about the real world around them are less and less able to form such ability. There starts a substitution of morals. Actions in the virtual world are transferred to reality. As a result — mass executions of passers-by by maniacs, inexplicable suicides, etc. In fact even stages of forming computer (virtual) addiction are becoming similar to the stages of "binding" to a drug.

Modern terrorism is the distributed system of a *network type*. This fact, to an essential degree, complicates the struggle against terrorism: network systems cannot be destroyed by "pinpoint strikes", they are capable of self-restoration. The analogy of a network principle with many popular aggressive computer games arises here.

Consequences of today's inefficiencies in raising and educating children can show up much earlier than we expect.

Much to our regret, information technology is frequently introduced to education not with the goal of fulfilling thoroughly planned educational tasks, but only in order to keep pace with the times. In many cases, those who are engaged in the computerization of school training spend too much effort and means on hardware and internet connection, and pay too little attention to the vocational training and support of the teachers. As a result — boring classes at schools are replaced by out-of-school games, mainly of a very poor quality.

In fact, we have no mass formation of youth culture based upon information and communication technology application, which could promote introduction in real life of the knowledge and skills received by young people in the virtual environment.

What is security? What is a threat? What is terrorism?

V. I. Muntiyan

Information is the most valuable global resource for the development of mankind and the state. Our life, thought, consciousness, and intellect are all — information. The universe, life, and reason are built on information as the basis.

Information is both the main type of strategic weapon of a state and its main resource.

There are three global threats that can bring the whole of mankind to cataclysm:

- thermonuclear catastrophe;
- ecological, technogenic catastrophe;
- informational collapse of civilization.

Without exaggeration, terrorism belongs on the list of global threats. There is no country, state or nation that can provide absolute security from terrorism.

Today this threat has emerged on a planetary level. The focused efforts of the whole global community are necessary for its neutralization. Summarizing all threats, starting with ecological, and finishing with threats of a technogenic character, it is possible to draw a conclusion, that mankind is sitting on the "powder keg". And terrorism can become the match that might cause this "powder keg" to explode. This is simple in form, but horrible in essence, as the threat to the very existence of the human race emanates from terrorism. It would be wrong to blame only underdeveloped countries for providing the sources of terrorism. The negative contribution to the growth of this threat was brought by the egocentric policy of developed Western countries in relation to other countries of the world. It has generated an inequality between the rich and the poor, lack of balance between scientific/technical progress and spiritual development of mankind.

Infringement upon 8 buffer zones which provide safety to the planet has provoked irreversible processes of biospheric instability. And instead of solving problems of peaceful existence, ecological, energy, and demographic safety, poverty and the spread of diseases, instead of all this — mankind spends $1 trillion for military purposes every year. There is an amazing proportion: to destroy 1 square kilometer with conventional arms, it is necessary to spend $2,000 US, with nuclear weapon — $800, chemical — $600, biological — $1. And through cyber-terrorism, it is necessary to spend only 5 to 10 cents for this purpose.

Therefore, today it is necessary to direct information resources not on information wars, but on the development and introduction of peaceful existence models for men among themselves, and in harmony with nature.

Information becomes more complicated in quality. The quantity of its sources grows, and different types influence the person with various purposes. Let us look upon two sides of information revolution as the two sides of a coin. The first comprises the positive factors of information influence on the development of the world community. The second comprises the negative factors — when the advanced information technologies and resources come into the terrorists' hands. Then it presents huge danger which brings tragedy and human suffering. In our opinion, solving this, one of the mankind's most complicated problems, lays in the field of knowledge of information sphere.

In the information sphere it is possible to establish the cause-and-effect relationships of the threat of terrorism. It also includes methods and mechanisms of liquidation and neutralization of these threats at the earliest stages of origin. The problem of security from terrorism threats is many-sided. But it is information that is simultaneously the original cause, the weapon, and the protection. Therefore, in the first place information should become an object of scientific research, an example of which is this conference, in which we all participate.

Our civilization, which is guided by its immediate consumer priorities and develops spontaneously, as a whole is inevitably approaching the border of bifurcation where the mankind should decide: whether to prompt ecological catastrophe (destruction), or to create new priorities and principles of harmonizing with nature.

The role of intellect (scientific thought) in the development of civilization increases. In the points of bifurcation we need to have a certain amount of "critical weight" of intellect in order to prevent the informational and ecological collapse.

Information factors obtain increasing value in the sphere of political interests. The economic potential is also, to a great degree, defined by the volume of information resources and the level of the information infrastructure development. Along with this, the vulnerability of

economic structures caused by unreliability, delay and illegal use of commercial information, industrial espionage, and hacking grows. Another reason for the increase in the role of information-psychological security is the intensification of the threat of using an information weapon in the international information exchange. It predetermines the necessity to solve the problems connected with the opportunities of information war, negative information influence on individual and public consciousness, mentality of people, and on computer networks and other information systems.

In the past, information struggle took place in practically all military actions, being shown in such basic forms, as conducting reconnaissance and counteraction to it, misinformation, rumors, and the struggle against them.

Information struggle can be defined as struggle of the parties for gaining advantage in quantity, quality, speed of receiving information, and its timely analysis and use.

By information warfare we mean the actions directed at achieving informational advantage in national military strategy, by imposing influence on the opponent's information and information systems and at the same time providing security and protection of one's own information and information systems. These are the features of information war:

- interception of all kinds of information and information systems with cutting off information from its sphere of use;
- objects can be both the weapon, and the object of protection;
- widening the territory and space of conducting wars both at declaration of war, and in crisis situations in various spheres of vital activity;
- conducting war with both special military units and civilian structures.

The concept of informational confrontation, according to experts, provides for:

- suppression (during military actions) of infrastructure elements of the state and of military management (destruction of command and control centers);
- electromagnetic influence on elements of information and telecommunication systems [radio-electronic struggle (RES)];
- reception of intelligence information by interception and decoding of the information streams transferred by liaison channels, as well as collateral radiations by means of electronic devices specially installed in the premises, and means for interception of the information (radio-electronic reconnaissance);

- unauthorized access to information resources (by use of hardware-software means of breaking the opponent's information and telecommunication protection systems) with the subsequent distortion, destruction, plunder or infringement of normal functioning of these systems (hacker war);
- formation and mass distribution on the opponent's information channels or global information networks — misinformation or biased information in order to influence evaluations, intentions and orientation of the population and the persons making the decision (psychological war);
- reception of the necessary information by means of interception and processing open information transferred by unprotected communicational channels, circulated in information systems, and published in mass media.

The information weapon, as well as informational confrontation, is subject to changes in the process of development of the society and information technologies. In modern practice, we understand the term "information weapon (IW)" as means of destruction, distortion or theft of information bits, extracting from them the necessary information after breaking systems of protection, restriction or prohibition of access to them, disorganization of technical means of work, incapacitating telecommunication networks, computer systems, all highly technological means of life support of the society, and interference with the functioning of state.

It is possible to relate to the category of information weapon, seven corresponding sets or groups of means, which can be applied for destructive (misleading, misinforming, disorienting, destabilizing, destroying, suppressing, etc.) information influences on substantial components of strategic control systems:

- mass media (radio, press, TV) and "propaganda" means (videocassettes, electronic textbooks and encyclopedias, etc.);
- phychotronic means (special generators, special video-graphic and television information, video);
- means of nano-technologies in the field of information (of "the Virtual reality" type, etc.);
- electronic means (optical and radio-electronic means) — special transmitters and radiators of electromagnetic waves and impulses;
- electronic computer means — "computer viruses", which destroy program bookmarks, etc.;

- linguistic means (linguistic units, "special" terminology, idiomatic phrases that have semantic ambiguity when translated into other languages, etc.);
- psychotropic means (specially-structured pharmaceuticals, psychopharmacological means, tranquilizers, hallucinogens, drugs, alcohol, etc.).

Unlike usual destruction means, the information weapon is characterized by the following features:

- secrecy — an opportunity to reach the purpose without preparation and declaration of war;
- scale — an opportunity to cause irreparable harm without violating national borders and sovereignties, without habitual restriction of space in all spheres of human activity;
- universality — an opportunity for multiple types of uses by both military, and civilian structures of the attacking country against both military and civilian objectives of the target country.

The sphere of IW application includes not only military sphere but also economic, banking, social, and other branches of potential use with the following purposes:

- disorganization of administrative structures, transport streams, and means of communication;
- blocking of separate enterprise and bank activity, as well as of strategic industries by disrupting multi-sectional technological ties, mutual payment system, and financial transactions;
- initiation of serious technogenic catastrophes in the opponent's territory as a result of infringement on management of technological processes and objects which are connected with manufacturing numerous dangerous agents and high concentration of energy;
- mass introduction and promulgation to people's consciousness of certain notions, habits and stereotypes of behavior;
- provocation of discontent or panic among population, and destructive actions of various social groups.

Thus the main targets of IW both in peaceful and wartime are the following:

- computer and communication systems, used by the state organizations while performing their administrative functions;
- military information infrastructure that deals with managing of troops and combat means, collecting and processing information in the interest of armed forces;
- informational and administrative structures of banks, transportation, and industrial enterprises;

- mass media, primarily electronic (radio, TV, etc.);
- telecommunication units, centers of satellite communication, and channels of international information exchange;

IW is extremely dangerous today for information computer systems at all levels of government, management of troops and weapons, chemical and biological objects, and also for atomic engineering, finance and banking, the national economy. It is no less dangerous for people when they are under information-psychological influence with the purpose of changing and controlling their individual and collective behavior.

At the same time, IW productivity can be compared to that of weapons of mass destruction, making it extremely dangerous when IW is in the hands of terrorists. We should do everything in our power to prevent it from happening; otherwise the processes might become uncontrollable.

IW, that can be applied effectively both in the times of war and peace, includes means of destroying informational computer systems and means of influencing people (their psyche).

Means of influence on people and their mentality are categorized in accordance with the purpose of their application in psychological war. They are the following:

- distortion of information received by opponent's political leadership, armed forces command and staff, and forcing on them false or meaningless information which deprives them of the opportunity to correctly analyze events or current situation and to make reasonable decisions;
- psychological processing of troops and population;
- ideological sabotage and disinformation;
- support of favorable public opinion;
- the organization of mass demonstrations under false slogans;
- propaganda and spreading of false rumors;
- change and control of individual and collective behavior.

Along with the use of traditional means (printed and electronic mass media), special means of influence on the person both through mass-media, and through computer networks are actively developed and tested: means of information-psychological (psychophysical) influence.

Application of these and other kinds of IW in conditions of openness and growth of the international information exchange defines protection methods of information systems from its influence.

Mental degradation of society in the near future will become a reality in case the state leaders do not analyze world tendencies in ecology of consciousness and do not draw corresponding constructive conclusions.

Instead of observing passively the growth of destructive forces and capacities, the state should take necessary measures to protect mankind against any opportunity of violence by means of Psychological Weapon (PSW), to protect each citizen's intellect (the most valuable gene fund of the nation and the state), and to make this protection open and easily accessible for wide international participation in work and control. Only then can we prevent the possibility of psychotronic war.

More and more often, we hear that in the third millennium the world leadership will be mostly defined not by the economic potential of the state, but by its ability to control the information processes. In the book *War and Anti-War*, Tofler claims that information technologies transform societies of the second wave (Industrial Forms) into societies of the third wave (Knowledge Forms). At the present time the world industry of information and communication computer technologies, by estimations of the World Bank, makes more than $1 billion US. In other words, there is a transition from economic to informational era in the development of the civilization.

Reorganization of societies based upon the new information bases has caused the demand for new approaches to providing national security in the information epoch.

Information as a complex of the factual data and the correlation between them, has become a highly marketable product — cost of the information and its timely delivery to its intended place increases like an avalanche.

Information penetrating all activities of the state, obtains concrete political, material and cost expression. The problem of information security, from the point of view of state interests, has now become especially important and is looked upon as one of priority state issues, as an important element of national security.

New information technologies integrate the world through global networks of instrumentalism. Communication performed through computers has given birth to a host of virtual communities.

It is no secret, that now one of the basic ways for a state or any organization to maintain their interests on the international scene is to gain information space by developing information technologies and creating information systems (IS) that can give access to various achievements in the fields of science, engineering, economy, etc.

At the same time systems of higher level, as a rule, have an opportunity to operate and control systems with lower levels of informational capabilities, directing and constantly supervising the activity of the latter according to their interests.

Security is the major concern in introducing new information technologies in all societal spheres of life, an internal system of influence recognition and formation of the subject's reaction to environmental actions. The basic purpose of immunity is the protection of the subject against infringement on his integrity. Hence, information immunity is a system of protection for a person from any informational aggression of environment.

The struggle against terrorism will become most effective when not only information security will be provided at the state level, but also the economic security, which should deprive terrorist organizations of resources. Similar levels of information and economic security should also be provided at the international level. They will serve as additional protection in case some country cannot cope with the task of ensuring reliable security from international terrorism.

IS is such a state of protection for vital personal, societal and state interests at which minimum losses are induced by such factors as incomplete, false or untimely information, harmful information influence, negative consequences of information technologies functioning, and unauthorized distribution of the information.

Regarding the question of national security as a system, it is possible to claim that information security holds a special place within it for the following reasons:

- first, information ties and processes penetrate into all relations that exist in the society;
- second, in modern conditions, when various information technologies are widely used, questions of information security receive meaning of their own;
- third, the system of external and internal threats to information security has complex character, universal for all spheres of human activity: individual and social.

Terror is a special form of political violence characterized by cruelty, ambition and illusive efficiency. Terrorism was a widespread tool in the struggle of revolution and counterrevolution during periods of social shock in the society. In modern conditions, escalation of terrorist activity in extremist organizations is observed. Its character becomes more complicated and anti-humanity of terror acts increases (taking hostages, hijacking planes, explosions, acts of genocide in religious conflicts, direct threats during political struggle, kidnapping or assassination of politicians, and other dangerous actions directed at mankind).

Terrorism is organization and realization of explosions or other actions, that jeopardize human lives, lead to significant material losses and

other socially dangerous consequences, if these actions have been carried out with the purpose of infringement of public security, intimidation of the population or imposing pressure on decision-making authorities, and also a threat of carrying out similar actions with the same purposes.

Terrorism is a crime against public security.

Attributes of terrorism:

1. Terrorism is a form of organized violence.
2. Terrorism influences broader layers of society, is not limited to direct victims of violence.
3. The formation of objectives in most cases is not connected with concrete displays of violence, i.e. there is no direct connection between the victims and the objective of terrorist actions.
4. The tactical goal of terrorism consists in drawing attention to a problem, strategic — in reaching certain social changes (freedom, independence, disposal from forced labor establishments for certain contingent of persons, revolution, etc.).
5. Terrorist acts in their essence are traditional forms of general criminal actions.
6. Terrorism paralyzes counteraction from the public.
7. The instrument of influence is the psychological shock which is generated by the comprehension that anybody can fall victim, regardless of what layer of a society the person belongs.
8. Terrorists accept no rules or laws. Victims of terrorist acts can be adults, as well as women and children.
9. They rely on the effect of suddenness, unexpectedness.
10. Publicity is the basic attribute of terrorism.
11. Making a spectacular show out of terror acts because of the desire to impress the broad masses.
12. Terrorism presupposes "political demand" therefore it is not connected with spontaneous rebellions and riots of population.
13. Terrorism requires immediate satisfaction of the demands brought forth, otherwise the realization of threats and the escalation of violence starts.
14. Can be used by the organizations of any political "color".
15. Practically always takes responsibility for committed acts of violence because they are the means to achieve the goal, but not the goal itself.
16. Represents an antithesis of political murder. Unlike the selectivity typical of a political assassination, it shows indifference in relation to victims.
17. The gap between the direct victim of violence, and the group which represents the object of influence and is the actual goal of violence.

Threats in the social sphere, which can be the roots of terrorism, are: low standards of living and social security of the population; a significant amount of citizens of working age who are not involved in any socially useful activity; and/or political opposition of separate social layers of the population and regions of the countries.

Besides, nuclear and biological terrorism are among the most dangerous types.

Actually, there are no reasons why our planet could not provide for the life of many more people than its present population. But in fact, the distribution of fertile soils and favorable conditions for their processing mismatches the distribution of population. This situation is deteriorated by growing degradation of land resources. Almost 2 billion hectares of soil is going through degradation due to the human activity. This fact jeopardizes the means of survival for nearly 1 billion people.

The main reasons of that are soil salinization as a result of irrigation, erosion caused by excessive pasturing, and the deforestation which results in a reduction of biological variety.

Globalization, along with positive factors, brings to mankind the whole cascade of threats.

1. The increase of population by one billion people might become a serious problem as long as economic stagnation and high percent of unemployment are an obstacle to the immergence of new workers or emigrants on the labor market.

The inadequate urban infrastructure and social services in the majority of cities will create conditions that will promote instability and disorder. Migration from the South to the North will become the basic source of tension, forcing the USA and European countries to distance themselves from the developing countries.

2. Growth of the population will promote reduction of the area of arable lands, reduction of fresh-water resources and biological variety. Shortages of resources, fresh water in particular, will become the main problem both for the countries with developed market economies, and for developing countries. This problem will lead to the reduction of agricultural production and to the increase of migration from rural areas to cities.

3. Introduction and distribution of technological innovations will occur slowly owing to economic stagnation and political uncertainty. Destabilizing effects of new technology introduction will prevail, which in its turn can lead to proliferation of the weapon of mass destruction. Information technology will create extra advantages for terrorists and criminals. Only several rich countries will receive advantages from the introduction of new technologies, and the majority of countries will fall behind.

4. Economic recession in the EU countries and the USA will lead to economic stagnation. The global agreement on support of market reforms will be broken, thus undermining "the American economic model", transforming the USA into a more sensitive state and, by that, leading to a decrease in the USA's role on the international scene. Today and in the near future there will be a struggle for the influence on the Asian region, not on Europe. This tendency will become the main world making policy. Annual rates of economic growth in China since 1978 have been 8–10%, in India for the last 10 years — 8%. In the coming 15–20 years, these countries will be in second and third place in the world in economic potential, after the USA For the same period, average annual rates of economic growth in the USA was 2.1%, and in the developed EU countries — less than 1.2%. While the critical level of economic security should be no less than 2.7% — this will give an annual increase of Gross National Product.

Economic stagnation will negatively affect the state of affairs in the countries with developing market economies, and also in the majority of developing states.

5. In many heterogeneous states religious/ethnic divergences will aggravate. Social tension and violence in Africa, Central and Southern Asia and in some areas of the Middle East will increase. The political influence of Islam will increase. The probability of terror acts against the objects connected with globalization and the USA will increase. The opportunities of the government at all levels, among both developed, and developing countries will reduce. In the near future the following unsolved challenges will remain:

(a) Uncontrollable growth of the world population which leads to the further deterioration of water supplies, foodstuffs, and power resources, and accelerates degradation of the biosphere as a whole, will lead to an increase in the frequency of conflicts of different sorts — inter-regional, regional, internal.

(b) The US aspiration to build up its military domination and create the new world order, and the policy of the industrially developed countries of the world which slows down the development of the Third world countries cause a growth in hostility of developing countries toward the USA.

(c) Increase of emigrants and refugees from dangerous regions of the world in the countries of the West.

(d) Strengthening of the role of the international terrorism in solving the inter-regional conflicts, in counteraction to military dominance of the USA and their allies, as well as the process of globalization.

(e) As a means of intimidation of the population and pressure on political leaders, acts of terrorism, in all their cruelty, do not remove principal causes of conflicts — overpopulation, struggle for natural resources, and lack of ecological balance.

(f) Negative factors of globalization processes are a stimulus for criminal revolution in the world.

6. Experts claim that in the 21st century the Third world countries, especially those with weak economy and lack of access to modern military technologies, from all the variety of weapons will prefer biological as the most profitable. It will be used mostly, not in combat conditions, but in time of peace, and not by the armed forces, but by terrorists. The main targets will be not only the adversary's armed forces, but civilian population as well.

The biological weapon can become the weapon of revenge, and also a means of realization for political murders in the hands of dictatorial regimes and terrorist organizations. An even more dangerous threat, by the scale of destruction, especially for industrially developed countries, is cyber-terrorism. The governments of separate countries alone cannot cope with today's global challenges. Therefore it is necessary to form world public opinion and to unite efforts of the world community to fight the global threats. The possibility to solve the problems mentioned above lies, in our opinion, in the introduction of informationalized economy as the bases of harmonious coexistence of the world community.

Thank you for your attention.

Round-Table Discussion "Comprehensive Security in the Fuel and Energy Industrial Complex"

Certain Aspects of Providing Systematic Protection to the Objects of the Unified Gas Supply System

B. N. Antipov

Mr. Chairman,
Dear participants of our round-table discussion,

For your attention I would like to offer my presentation which will highlight certain aspects of providing protection to objects of the unified gas supply system. As of today the open joint stock gas company Gazprom (JSC) is one of the major fund suppliers to the budget of the country. The proper operation of the gas supply system ensures the functioning of all our other industries. We export gas to foreign customers and are fully responsible for the contracts signed. As of today, the gas transportation and gas extraction objects are reliably covered by a protection system, which to a certain extent provides integrated protection to these units. But the worldwide growth of terrorism, which can cause serious consequences to our country in case of sabotage and terrorist acts at the objects of gas extraction and transportation, forces us to improve current protection systems against the unauthorized access to the objects of the gas industry. Terrorism triggers the necessity of complex reorganization of the entire system, as well as upgrading of software and hardware according to new scientific and technological achievements. As of today, the JSC Gazprom security service has developed a new concept of country's gas transportation and production units' protection which besides the creation of new technical means of preventing unauthorized access to them suggests a new classification system for these objects. The system divides these objects into:

- serviced objects;
- rarely serviced objects;
- non serviced objects.

If we consider all the possible threats that might arise at these objects, we can identify two types: external and internal threats (each of

these threats is presented on the slide). One should mention the importance of both areas, which will be estimated in this comprehensive protection program, and adequate measures will be developed. They will cover and include all the possible consequences of the adversary's action. In the concept, which we hope to finalize in the near future, we will pursue the following goals:

- security threat prevention;
- personnel life and health protection;
- damage or destruction prevention of installations, technical means, property, and valuables.

These purposes will be achieved through timely threat identification and elimination, through providing a high level of engineering and technical protection systems and antiterrorism protection system exploitation, through the efficiency increase of all security divisions, and through specification and improvement of the JSC Gazprom regulatory base in the field of providing security. I would like to point out that at the present moment this regulatory base, which would allow improving the system and providing necessary high-quality technologically advanced protection system, is practically non-existent. Of no less importance is the improvement of personnel recruiting and training. Let me focus on some organizational aspects of the object protection. If we review this issue from the view point of comprehensive object protection insurance, we have to take into account the following factors of protection system functionsing:

(1) its continuous action;
(2) ability of the system to function throughout the lifetime of the object;
(3) centralized control of the protection system functioning;
(4) standard methods use;
(5) and ability of the protection system to develop and improve (I would like to mention here that the concept of the protection system assumes that the systems will be continuously developed and improved in compliance with scientific and technological progress).

I have to emphasize here that the expenses of creating such protection systems are completely justified. I would like to draw your attention to the fact that there are no approved methods of calculating possible expenses on the means of preventing the unauthorized access or destruction of gas production and transportation objects. However the preliminary estimation made by the Gazprom security service shows that the losses

that the Gazprom experienced due to unauthorized access to its objects are very significant. The major principles for finding a solution to this challenge consist in observing the federal law of the Russian Federation. During the last few years, five regulatory documents aimed at providing protection and defense of the objects were adopted and approved in the Russian Federation. An analysis of threat prediction to the comprehensive security of the objects is also required. All activities of the current systems, as well as of the systems to be installed on the new units and upgraded at the old ones, have to be interlinked put in accordance with the actions of law enforcement agencies in this region. Unfortunately, at present, this connection is very weak.

There are two more aspects that are more or less clear. First is the creation of databases for the protected objects. One of the main innovations in the above mentioned concept will comprise the establishment of 6 or 7 regional centers which will not only register the objects' protection system work, but also control its functional reliability and in case of emergency will provide manual control of the protection system. The real time data from these regional centers will be transferred to the Gazprom central control board. Therefore, the Gazprom management will be fully aware of the protection system status, each case of unauthorized access to the protected object and also about the measures taken to prevent or eliminate the consequences of such unauthorized access.

I would also like to say a few words about providing protection to the objects. Let me begin with the means that, we think, can ensure such protection.

Continuous access control to the objects is comprised of: guard TV; the building and construction perimeter signaling system; control of mobile objects (I mean the COs of Gazprom); search and detection of subjects, substances, materials, and stationary or portable equipment, which are not to be carry to the object (what we are suggesting here is step by step object protection, which includes first, protection form vehicles' then from individuals' intrusion on the object, who will be checked for explosives with the help of special detectors); creating barriers to unauthorized access (one of the ideas of our concept is building a mechanical barrier to stop the unauthorized penetration of vehicles onto the territory of the object. We also consider the possibility of blocking the vehicle at the violation point). It should be mentioned here that the international practices used by terrorists in the modern world show two major ways of conducting a terrorist act. It is either the use of a suicide bomber who carries a certain amount of explosives or the use of

a vehicle loaded with explosives. I am not considering here a possibility of a missile attack or aircraft bombing.

We believe that the methods that should provide protection to the objects are:

- armed guards;
- centralized guard system;
- integrated protection systems, which are interconnected and synchronized;
- control of protection status of the Gazprom objects on the basis of the unified informational-analytic database, which I have already mentioned above;
- offensive measures of a temporary effect for the rarely serviced objects (it means that there are certain objects where we do not have permanent personnel. Therefore these objects have multilevel protection systems. On the first level, the protection system sends a signal of an unauthorized access. On the second, it will show if the unauthorized access occurred in an enclosed space, where certain chemicals will produce a temporary impact on the intruder. The chemical agents used for this purpose have to be certified, laboratory tested, and have an expiration date);
- and approbation and construction of new systems on the specialized testing range. (It is a special item of the concept, that all systems that we want to use have to be tested and certified on the very object where they will be used. For these purposes the testing range will be modified, newly equipped, and all systems will go through the process of technical approval there).

As I have mentioned earlier speaking of logistics, one of the major ideas is the establishment of regional centers. The very number of objects (which is enormous) complicates centralized maintenance and control. But the system of regional centers, which will be connected with the central control board, will help solve this problem and will also reduce the number of false alarms.

- regulation of object's security service activity, cooperation with private companies, accredited in the field of providing security (I do not think that we will have any problems here. An overwhelming majority of private protection agencies are certified, though in order to work for Gazprom they will have to receive a special access to the basics of Gazprom security);
- use of modern control methods in equipping the engineering and technical protection systems and antiterrorism protection system objects.

Regarding personnel, we suggest that the security service employees are covered by compulsory insurance. As I have mentioned before, the gaps in the existing regulatory base will have to be eliminated. One of the gaps concerns the voluntary equipment certification, which at present is not properly developed from a legal point of view. But we believe that only the equipment, fully certified in efficiency, will be used at our objects.

One of our major challenges is personnel and raising the level of the personnel skills and qualifications. It is necessary to say here that judging by the information we have today, personnel are lagging behind the technological level of our equipment. It means we have to continuously train our employees because the technologies are developing very fast and we need to have personnel capable of controlling this equipment.

We also need to mention that there is no legal interaction today between the security service of the JSC "Gazprom" and regional protection forces of the Ministry of Internal Affairs, Federal Security Service (FSB) and the Russian Emergency Control Ministry (EMERCOM). We believe that there is a need to develop and strengthen the relationship on the basis of bilateral agreements. To give a better idea of the interaction of these three agencies and the Gazprom security service, we have prepared the data in the table that you can see. First of all, we have to ensure the coordination between the Gazprom security service and regional divisions of the Ministry of Internal Affairs, FSB, and EMERCOM not only for elimination of a possible terrorist attack or unauthorized access consequences, to prevent this unauthorized access. As of today, our law enforcement agencies widely use the information about the location of suspicious individuals in this or that region, moreover these agencies have technical and technological means to conduct visual and tactical surveillance of a suspicious object at significant distance. The interaction between local law enforcement structures and the Gazprom security service will allow us to eliminate possible consequences of the objects' performance interruption, but also to prevent a possible unauthorized access.

Thus, I think it is necessary to mention that a proposed special comprehensive protection program for 2005–2007 will allow us to start technical upgrading of the Gazprom objects protection system and will also enable us to develop more advanced methods and means of protection. At the first stage out of 30,000 Gazprom objects our program will identify the ones that are crucial for the smooth functioning of gas distribution and transportation system. These vitally important objects

will be the first to have installed or upgraded protection systems. The major objectives of this comprehensive program are:

- reducing objects' vulnerability level;
- perfecting the protection system;
- centralizing the control and monitoring of security systems;
- and installing at the objects the uniformed integrated protection means.

All these goals can be achieved, though as I have mentioned, it will not be possible to cover all objects. Within two years we have to work out uniform methods and technical means of object protection. And again I will use the word unification so it will also cover the technical documentation of the objects. We can name the following tasks:

- early threat warning;
- reduction of threat level and its neutralization;
- development of new systems and technologies;
- and providing objects with integrated comprehensive means of protection.

In conclusion, I would like to go back to the expected results of the program. I believe that we have much of work to do and a long road to travel. The Gazprom security service pays utmost attention to this program and it has been approved by the Board of Directors. The major task that we plan to achieve due to the implementation of this program is providing comprehensive security to the Gazprom objects. And now let me quote what makes the backbone of providing protection, "basing on the control of protection level of the critically important objects, basing on determination and state of the mobile objects, basing on detection of explosives at the objects checkpoints and preventing their use, basing on prevention of the unauthorized vehicle movement, and protection of major gas pipelines on the coastline and riverside, if the objects in the sea can somehow be protected, they should lay down to 50 meters depth, which is common in our practice, to prevent the unauthorized access from individuals with aqualungs".

I would like to make one more observation in conclusion. The implementation of this program is based on the complex approach to the protection and defense of various Gazprom objects, development of security system database, and establishment of the regional centers. Every impact on every Gazprom object might bring severe environmental consequences which is also one of the most important factors that makes the development of the efficient integrated system of physical protection of Gazprom one of the crucial tasks.

Infrastructure of Complex Security Systems for Enterprises: Urgent Problems

V. F. Pustarnakov, V. N. Kustov

Dear colleges,

Allow me to thank the organizers of this conference for the opportunity to give a report on behalf of the "GazInformService" leadership on such a remarkable forum. I hope that we will all meet quite soon at the PKI security forum in Saint Petersburg and at the exposition dedicated to information issues, which will take place on November 8–10, 2005 in LenExpo. "GazInformService" participates in these events and organizes the session of the Gazprom's 5th section.

Allow me to begin my presentation with the definition of "object of protection", which we believe consists of the following components: people (objects personnel, enterprise's clients); material and financial valuables (equipment, production objects, office buildings, infrastructure objects); and confidential information. Sometimes we forget that besides the material and financial valuables, object of protection can also include relations that emerge between the personnel working at these objects during the production activity, and the material valuables such as financial relations, production relations, and other relations of administrative nature. Thus this scheme is incomplete and these relations can be added to the list of the objects of protection. Besides the objects of protection mentioned above we can add the following ones: preventive and counter measures against physical threats; prevention of unauthorized access to the guarded objects, buildings, zones, facilities or to guarded subjects; creation of efficient alarm signal; control of employee, visitor and vehicle access to the object's territory and to the secure facility; providing employees' schedule of work; supervision of the area adjacent to the object territory and its traffic; creation of video archives; and computer analysis of the object's security. In our opinion, the principles of complex security are the following: universality; complexity; rational adequacy; efficiency; adaptability; continuity; systematic character;

purposefuness; layering; balance; echelonment[1]; compatibility; simplicity; environmental cleanliness; obscurity; friendliness; invulnerability; documentation; and relevance. These are, in our opinion, the major principles of creation of a complex security system.

What do we need integration for? First of all, an integrated technical security system assumes unification basing on modern information technologies, and program-hardware integration of subsystems, which are functionally and informationally connected to each other. Even on the lowest level of integration, the interaction between the subsystems can be carried out in such a way that occurrences in one subsystem can influence the others, and produce certain reactions. In comparison with the simple aggregate of protection means, in our opinion, the use of integrated systems provides the following advantages: faster and more precise reaction to current events; optimal analysis of current situations; substantial reduction of risks related to human factor; equipment cost reduction; facilitation of personnel's labor due to process automation; cost reduction on installation and exploitation of security systems; and the reduction of personnel and the expenses of training and salary. The main components of infrastructure of a complex security system, in our opinion, are: engineering protection system (one of the main systems); alarm signal system(a standard component); fire safety system; control and access monitoring system; if necessary — video observation system; centralized information collection and processing system; intercommunication system; and system of real-time monitoring of moving objects, which are used not only for Cos, but also for valuable cargo transportation for reason of taking adequate measures in case of terrorist threat or criminal activity; explosive alarm system with chemical means of temporarily neutralizing a terrorist or criminal until first response forces arrive; system of audiovisual warning in case of personnel emergency evacuation. Besides, as the recent events in Moscow have shown, one of the major components is a system of guaranteed power supply.

The above mentioned documents issued in the open joint stock company "Gazprom", primarily the plan and comprehensive special program, identify five directions of activity. First of all, it is an establishment of infrastructure and regulatory legal base for security of unified gas supply system objects; equipping objects with uniformed modern complexes of engineering and technical protection systems and antiterrorist protection system; training of skilled specialists in the field of protection means exploitation; establishment of quality-guarantee system for providing security to the objects of gas supply industry; and establishment of an interaction system between counter terrorism agencies. In the first we distinguish the following objectives:

(1) formation of engineering and technical protection systems infrastructure for providing security to the objects of gas supply industry;
(2) creation of automated control system of object's protection;
(3) development of systematical software for state of protection assessment;
(4) creation of computer data exchange system;
(5) preparation of standards in the open joint stock company "Gazprom" for dividing objects into groups depending on degree of possible danger they are facing, and the terrorist vulnerability of these objects to terrorism;
(6) determination of critical objects that should be the first equipped with engineering and technical protection system and antiterrorist protection system.

In the second direction of activity, the following objectives are distinguished:

(1) equipment of objects with uniformed modern complexes of engineering and technical protection systems and antiterrorist protection systems;
(2) development of standard design decisions;
(3) development of uniformed complexes and their integration into existing systems.

Modernization of security complexes of critical objects is executed through accomplishment of these three objectives.

The objective of the third direction of activity is the organization of specialists training program in the field of engineering and technical protection systems and antiterrorist protection systems (as you probably understand, in the modern training system such educational institutions do not exist).

[1]I would like to draw your attention to the fact that the multi layering and echelonment are not equivalent. The word 'multi layering' just means a number of layers (rings of security). It is quite possible, for example, to use two notification layers in the information system, these layers will execute equivalent functions but the notification itself will becomes multifactor. The echelonement considers that each layer consecutively performs its tasks and next layer is calculated basing on the presupposition that the previous layer was successfully passed by a criminal.

System Aspects of Providing Security to the Open Joint Stock Company "Gazprom" Objects with the Use of Risk Index

V. N. Pozharsky, V. S. Safonov, V. V. Lesnykh

Russian unified gas supply system is a unique administrative and technical system characterized by a large number of production facilities and their expanse. A fundamental feature of the Russian gas supply system is the continuity of the operating process that takes place in real time and needs complete compatibility in material balances and technologies of all the system sections (production, transportation, refinery, storage). Providing secure and stable functioning of such a system is a very complicated scientific/technical and organizational problem. This problem becomes more and more urgent because of the growth of internal and external threats. The most significant ones are illegal actions including terrorist acts.

It is very important to make a decision concerning the structure of the security system and optimal material and financial resource allocation based on the rating of Gazprom production and administrative facilities as the objects being exposed to illegal actions depending on level of systematic risk.

Regardless of the nature of initiating event, the emergencies at the Gazprom objects can cause social and economic consequences in a certain region or in the whole country. Post accident analysis, including level of expected damage, is conducted with the use of optimized and simulation models that describe gas flow distribution during normal operating and emergency modes.

While allocating resources for organizing security at the Gazprom objects, quantitative evaluation of systematic risks and efficiency analysis of preventive measures have to become integral stages of Gazprom's rating of objects that require protection. Expenses for specific preventive measures should be considered as an investment in security and sustainable development. Thus, expenses on security and stability of the Russian unified gas supply system should be validated and allocated depending upon the level of expected damages.

Urgent Problems in Providing Information Security to the Gas Industry

A. I. Efimov

The following is the list that comprises the urgent problems of providing information security to the gas industry:

- protection of basic and technical means of Automatic Process Control System (APCS);
- protection of APCS information resources including databases and Scada system;
- security system of APCS process equipment;
- control of used in APCS means of telecommunication, communication, computer engineering, Scada system, etc.;
- development of APCS in its guarded construction and its certification according to information security requirements;
- control of set exploitation regulations and security system adjustment with the help of Russian organizations certified for such kind of activity (in this case it means that, as a rule, objects of APCS are served by Russian companies but there cases that some objects since the Soviet times have been technologically and technically maintained by neighboring countries. The matters of information security is a certified type of activity, the technological objects are situated on the territory of the Russian Federation that is why we believe that such service should be provided by Russian organizations which are licensed for such type of activity).

Special emphasis should be placed on drawing a physical and logical line of distinction between APCS and other Enterprise Information Systems, for example, between APCS and production and commerce operation control systems. In other words APCS is a separate logical or physical net segment, while bookkeeping, inventory of material and technical resources, and warehouse supplies are other segments. There should be a clear distinction line between them. The information system of other industries can also be related to APCS. For example, the automatic system of commercial energy accounting can be related to APCS. There is a requirement that all manufacturing enterprises should

have automatic energy accounting systems (AEAS). In this case, it is necessary to understand that AEAS is a system which belongs to one department, for example, to RAO Unified Energy System of Russia, but APCS can belong to some oil or gas production enterprise. A logical division should be made between these two.

Taking into consideration that terrorism, including technological terrorism, aims primarily at technical infrastructure and its purpose is to cause maximum damage to industry and population, it is necessary to mention that above named approaches to providing security to information systems of critical industries should be legally set forth on the federal level in the technical regulations on information security, that was already mentioned in the beginning of this presentation.

In our opinion the second problem of providing security for critical industries is the necessity of providing information security to technological objects in conditions of enterprise or branch reorganization. This problem became very obvious after the accident that occurred in Moscow, the Moscow region, and neighboring territories in May 2005. Obviously, during any reorganization of a large enterprise that operates geographically distributed technological objects, the question of adaptation of APCS and its protection systems providing information security in a new organizational structure, has to become a top priority. This adaptation should involve technical improvements to APCS and its protection system; engineering, construction, and operational documentation updating; personnel retraining, development of interaction among elements of a new technological structure and related services including dispatch. Other organizational problems must be solved during the restructuring process, among them: assignment of responsibility in the field of information protection to leaders of newly formed organizations; certification of modernized APCS according to information security requirements; development and initiation of provisions for interaction with related bodies of technological control.

The third problem of providing security to technological industries, in our opinion, is a necessity of providing a minimal level of information security to technological objects which were built in the 70s–80s of the last century. We all know that preparing gas for transportation or preparation of some other technological object is very expensive. The whole country used to aim efforts at realization of such missions. Such objects serve for a long time. There are such objects in our country, and this is where the problems come from. There are some difficulties with taking prompt action on improving information security of such objects. These difficulties arise from the fact that informatization means

of technological infrastructure objects belong to the last century, to the previous generation of protection means, and it is difficult to adapt them to modern information security apparatus. The opinion is that these objects have an acceptable level of security, which is determined by their low level of automatization and the use of antique hardware support. In reality, a study of the deconstructive impact on such infrastructures never took place. In this respect, threats to these infrastructures continue to exist. This fact makes the inspection of the APCS vulnerability of old technological objects very urgent.

The last problem is related to the necessity to develop competitive Russian-made elements of automatic control system of technological processes, and to guarantee users' independence from foreign manufacturers. This will help to solve a number of problems related to information security. And, of course, we must develop our own instruments of hardware support for automatization of technological objects with continuous production cycle.

In conclusion, I would like to mention that problems have a complex nature, and their solution will require efforts not only from corporations, but also from government agencies.