RESEARCH OF THE INFLUENCE OF METHODS OF COMPRESSION ON THE CO-EFFICIENT OF INFORMATION SECURITY OF OBJECTS UNDER ATTACKS

Dimitrina Polimirova, Eugene Nickolov, Cecko Nikolov

***Abstract**: In this paper a possibility for quantitative measuring of information security of objects, exposed to information attacks and processed with methods of compression, is represented. A co-efficient of information security, which reflects the influence of the level of compression obtained after applying methods of compression to objects and the time, required by the attack to get access to the corresponding object, is proposed. Methods' groups with the highest and respectively the lowest values of the co-efficient of information security for all methods' groups in relation to all attacks' groups are determined. Assessments and conclusions for future investigations are proposed.*

***Keywords**: Information Flows, Information Security, Information Attacks, Methods of Compression, Compressed Objects, Coefficient of Information Security, Level of Compression.*

***ACM Classification Keywords**: D.4.6 Security and Protection: information flow controls*

The Situation

Since the 70-th of XX century the problem for security and protection of information flows has drawn developers' and constructors' attention in the area if information technology [1]. With the first malattack in the 60th of last century [2], a progress in the area of object protection is observed and requirements for information security of objects are increased. Later the problem for creation of maximum protection for information flows arisen.

This cause the necessity to solve various problems associated with heightening the information flow security in the processes of transfer, processing and storage of different types of information flows. The researches on various methods for heightening the information security of different types of file objects became actual.

The Problem

The information flows, exposed to different attacks, are characterized with their large volume. Different methods for compression were developed and their use became necessity to reduce the volume of information flows. Compression can be used as a means of heightening the information security of information flows, especially when a password is applied during the compression.

We shell understand as a compression the transformation of the input data flow into codes. The decision for the correspondence "input data — output codes" is based on a preliminarily chosen model [3]. In case of effective compression, the obtained flow of codes is smaller in volume than the input data, but even though the compression is not effective, the file object will have a better protection against different attacks, because the output data will be presented by codes.

The purpose of research is to determine those compression methods whose application will give the object the highest security and protection against different attacks.

The Experiment

The experiments which were carried out were related to the necessity for a quantitative measure of information security. The following limitations have to be taken into consideration: 1) only the sets of potential attacks' groups (A_{pot}), methods' groups (M_{pot}) and objects' groups (O_{pot}), described respectively in Tables 1a, 1b, 1c will be analyzed (the groups' formation is described in [4]); 2) the experiments will be carried out according to standard user, not corporate or governmental, requirements.

Table 1a Attacks' groups

ATTACKS' GROUPS
I. Advertisements
V. Chat
VI. Criminal Investigations
VII. Cracking
VIII. Spying
X. Exploits
XIII. Scanners
XIV. Keyboard Modifiers
XVII. Computer Trojan Horses
XVIII. Computer Backdoors
XIX. Computer Worms
XX. Computer Viruses
XXI. Accessible information
XXIV. Content
XXV. Data Encapsulation
XXVII. Spoofing
XXXI. Social Engineering
XXXIII. Zombie Computers

Table 1b Methods' groups

METHODS' GROUPS
I. Statistical lossless methods
II. Dictionary lossless methods
III. Image lossless methods
IV. Audio lossless methods
V. Other lossless methods
VI. Dictionary lossy methods
VII. Image lossy methods
X. Audio lossy methods

Table 1c Objects' groups

OBJECTS' GROUPS
II. Scientific file formats
III. Data file formats
IV. Graphics
V. Sound
VII. Internet related
VIII. Binaries
X. Miscellaneous

A research will be conducted according to the influence of the level of compression on the time required by the attack, representative of an attacks' group from the potential set of attacks' groups, to get access ρ to the object, representative of an objects' group from the potential set of objects' groups, where ρ will present the different type of an access (read, write, execute and delete – to simplify the investigations will not make difference between its functionality). The following tasks were posed:

1) To determine the co-efficient of information security of the object.

The co-efficient of information security as a quantitative value, which depends on several fundamental parameters, can be determined. The co-efficient of information security can be presented as a value, dependent on several main parameters. These parameters can be represented as ratios of separate values before and after certain impact. The parameter *TIME* can be represented as a ratio of *time-prim* for the attack of the object before impact of the method of compression to the *time-second* for the attack after impact of the method. The parameter *SIZE* can be represented as a ratio of *size-prim* of the object before impact of the methods of compression to the *size-second* of the object after the compression. In this way the rest parameters as password, encryption etc. can be dynamically described in the necessary depth and concreteness.

1.1) The first stage of this task is to determine the *LEVEL OF COMPRESSION* of the objects processed with different methods of compression.

Three typical representatives for each objects' group, from the set of potential objects' groups (totally 21 representatives) and three typical representative for each methods' group, from the set of potential methods' groups (totally 24 representatives) are chosen. Each object is processed with the different methods and results are obtained for the corresponding level of compression (*L*) (totally 504 levels of compression, 9 for each relation object's group—method's group). The level of compression (*L*) can be defined as [5]:

$$L = \left(1 - \frac{\text{uncompressed size}}{\text{compressed size}}\right).100 = [\%]$$

To improve the visualization in Table 2 are shown the average values of the level of compression for each relation method's group—object's group. Maximum and minimum values are marked.

Table 2 Average values of the level of compression (*L*).

		METHODS' GROUPS									
		I	***II***	***III***	***IV***	***V***	***VI***	***VII***	***VIII***	***IX***	***X***
OBJECTS' GROUPS	***I***	-	-	-	-	-	-	-	-	-	-
	II	78%	32%	-	-	62%	-	-	-	-	-
	III	64%	92%	-	-	78%	-	-	-	-	-
	IV	-	-	81%	-	-	**96%**	94%	-	-	-
	V	-	-	-	73%	-	-	-	-	-	86%
	VI	-	-	-	-	-	-	-	-	-	-
	VII	56%	53%	-	-	48%	-	-	-	-	-
	VIII	**5%**	9%	-	-	7%	-	-	-	-	-
	IX	-	-	-	-	-	-	-	-	-	-
	X	6%	8%	-	-	-	-	-	-	-	-

1.2) The second stage is connected to the experiments which have to be carried out measuring the *TIME* required by an attack $a_i \in A_{pot}$ to get access ρ to the object $o_f \in O_{pot}$ in the case when the object is processed with method of compression $m_j \in M_{pot}$, and in the case when the object is not processed with method of compression.

The equipment used for the experiments includes: two server configurations with two workstations each. The first server configuration (in conjunction with two workstations) is used for research of "attacking behavior", another server configuration (in conjunction with another two workstations) is used for research of "protective behavior".

The workstations in the respective server configurations assume the role of "managing station" and "model station". The research techniques are based on the utilization of program systems for scientific research of the firm "The MathWorks", including the *family Matlab 7.2* and the *family Simulink*.

Three typical representatives are chosen for: 1) each of 18-th attacks' groups; 2) each of 8-th methods' groups; 3) each of 7-th objects' groups from the potential sets of groups, shown in Tables 1a, 1b, 1c.

An estimation is done with respect to the time required by the attack (totally 54 attacks) to get access to each object (totally 21) in these cases: 1) objects are not processed with methods of compression; 2) objects are processed with methods of compression (as the time for applying of each one from 24th methods to the corresponding object is investigated).

Results are obtained for totally 27216 triple relations attack—methods—object and the obtained values are compared with the values of relations attack—object. Thus the time growth for realizing an attack to an object, after its processing with the corresponding method, can be determined (in percentage). To improve the visualization in Table 3 are shown only the average values for obtained time growth in percentages, required by the attacks to be realized to the objects in case when the objects are processed with methods of compression. Maximum and minimum values are marked.

Table 3 Average values for obtained time growth in percentages, required by the attacks to get access to the objects processed with methods of compression.

OBJECTS' GROUPS	METHODS' GROUPS I	II	III	IV	V	VI	VII	VIII	IX	X
I	-	-	-	-	-	-	-	-	-	-
II	27,90%	**24,53%**	-	-	60,33%	-	-	-	-	-
III	41,23%	38,29%	-	-	38,13%	-	-	-	-	-
IV	-	-	51,86%	-	-	43,64%	44,28%	-	-	-
V	-	-	-	38,29%	-	-	-	-	-	36,05%
VI	-	-	-	-	-	-	-	-	-	-
VII	46,76%	44,36%	-	-	**85,88%**	-	-	-	-	-
VIII	34,17%	31,10%	-	-	30,93%	-	-	-	-	-
IX	-	-	-	-	-	-	-	-	-	-
X	43,29%	39,65%	-	-	-	-	-	-	-	-

1.3) The third stage is to determine the *CO-EFFICIENT OF INFORMATION SE*CURITY (K_{INF}).

The coefficient of information security can be determined as a combination of the coefficients of time (K_T) and size (K_S). K_T is the ratio of the time, required by the attack to get access to the object before impact of the method, to the time, required by the attack to get access to the object after impact of the method.

$$TIME = \frac{\text{time required by the attack before impact of the method } (t')}{\text{time required by the attack after impact of the method } (t'')}$$

K_S is the ratio of the uncompressed object's size to the object's size processed with method.

$$K_S = \frac{\text{uncompressed object's size } (s')}{\text{compressed object's size } (s'')}$$

Then K_{INF} can be denoted as:

$$K_{INF} = K_T + K_S = \frac{t'}{t''} + \frac{s'}{s''}$$

After the investigations, described in 1.1) and 1.2), which were carried out, the co-efficient of information security can be determined for the representatives of corresponding objects. To improve the visualization in Table 4 are shown only the average values for obtained coefficients of information security from applying methods to objects, exposed to attacks. Maximum and minimum values are marked.

Table 4 Average values for obtained coefficients of information security from applying methods to objects, exposed to attacks.

OBJECTS' GROUPS	METHODS' GROUPS I	II	III	IV	V	VI	VII	VIII	IX	X
I	-	-	-	-	-	-	-	-	-	-
II	5,27	2,23	-	-	3,03	-	-	-	-	-
III	3,37	13,12	-	-	5,16	-	-	-	-	-
IV	-	-	5,74	-	-	**25,56**	17,22	-	-	-
V	-	-	-	4,32	-	-	-	-	-	7,78
VI	-	-	-	-	-	-	-	-	-	-
VII	2,81	2,68	-	-	2,06	-	-	-	-	-
VIII	1,71	1,79	-	-	1,77	-	-	-	-	-
IX	-	-	-	-	-	-	-	-	-	-
X	**1,63**	1,69	-	-	-	-	-	-	-	-

2) To determine the methods with the highest coefficients of information security

A database can be build after experiments which were carried out during the first posed task. The base includes matrices with the values of obtained coefficients of information security when methods of compression were applied to objects. The number of matrices is 18 (one for each group of attacks). Each matrix includes the average values of the obtained K_{INF} from the applying 8 methods' groups to 7 objects' groups. Thus for each matrix the methods' group with the highest co-efficient of information security for every objects' group for the respective attacks' group can be determined. In Table 5 are shown the average values for the best methods' group, after applying the highest values of the co-efficient of information security for all objects' groups, exposed to the corresponding attacks' group, are obtained.

Table 5 Methods' groups after applying the highest value of the co-efficient of information security for all objects' groups, exposed to the corresponding attacks' group, are obtained

Attacks' groups	Group of methods with the highest values of K_{INF}
I. Advertisements	Dictionary lossless methods
V. Chat	Dictionary lossy methods
VI. Criminal Investigations	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
VII. Cracking	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
VIII. Spying	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
X. Exploits	Statistical, dictionary lossless, methods
XIII. Scanners	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XIV. Keyboard Modifiers	Image lossless
XVII. Computer Trojan Horses	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XVIII. Computer Backdoors	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XIX. Computer Worms	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XX. Computer Viruses	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XXI. Accessible information	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XXIV. Content	Statistical, dictionary lossless, dictionary lossy methods
XXV. Data Encapsulation	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XXVII. Spoofing	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XXXI. Social Engineering	Statistical, dictionary lossless, dictionary lossy, audio lossy methods
XXXIII. Zombie Computers	Statistical, dictionary lossless, dictionary lossy, audio lossy methods

From the experiments which were carried out and after the obtained results methods with the highest co-efficients of information security in relation to all attacks for every object can be determined. Results are shown in Table 6.

Table 6 Methods' groups with the highest coefficients of information security in relation to all attacks' groups for every objects' group

Objects' group	Groups of methods with the highest values of K_{INF} in relation to all attacks' groups
II. Scientific file formats	Statistical lossless methods
III. Data file formats	Dictionary lossless methods
IV. Graphic file formats	Dictionary lossy methods
V. Sound formats	Audio lossy methods
VII. Internet related	Statistical lossless methods
VIII. Binaries	Dictionary lossless methods
X. Other	Dictionary lossless methods

The following assessments could be made from the experiments which were carried out:

1) With respect to the LEVEL OF COMPRESSION:

a) With respect to the selected objects which will be processed by methods of compression, the assessment is positive, and the assumptions don't impact the obtained results. With respect to the chosen methods of compression, the assessment is also positive, and the conducted experiments can be applied successfully enough for the other methods as well.

b) The best results are shown with objects from the graphics group (*IV*), processed with a method of compression belonging to the group of dictionary lossy methods (*VI*). The worst results are shown with objects from the group of other file formats (*X*), processed with a method of compression belonging to the group of statistical methods (*I*).

2) With respect to the TIME, required by an attack to get access to an object:

a) The selected server configurations, defined as "high-level professional server", and the workstations, defined as "middle-level professional station", are sufficient for making the necessary conclusions and recommendations. The conducted experiments on these configurations are also valid for other standard user not corporate (government) configurations as well.

b) The best results are shown with objects from the group of “Scientific file formats”, processed with a method of compression belonging to the group of Statistical methods with respect to the attack belonging to the group of Spying. The worst results are shown with objects from the groups of Data file formats and graphics file formats, processed with a method of compression belonging to the group of “Other” lossless methods with respect to the attack belonging to the groups of Criminal Investigations and Computer Viruses.

c) Based on the conducted experiments, we can make the conclusion that increasing the level of compression leads to increasing the time required by an attack to get an access to an object.

3) With respect to the CO-EFFICIENT OF INFORMATION SECURITY:

a) The best results are shown with objects from the group of graphics file formats (*IV*), processed with a method of compression belonging to the group of dictionary lossy methods (*VI*) with respect to the attack belonging to the group of Zombie Computers (*XXXIII*). The worst results are shown with objects from the group of other file formats (*X*), processed with a method of compression belonging to the group of dictionary lossless methods (*I*) with respect to the attack belonging to the group of Criminal Investigations (*VI*).

b) It is necessary to conduct research by using a password in determining the coefficient of information security.

Conclusion

The influence of compression methods on different objects is substantial and can be utilized for making decisions with regards to the information security of these objects in relation to different attacks, and for evaluation of methods with the lowest risk with respect to the coefficient of information security.

Bibliography

[1] Dorothy E. Denning, A lattice model of secure information flow, Communications of the ACM, v. 19 n. 5, p. 236-243, May 1976.

[2] http://www.sptimes.com/Hackers/history.hacking.html

[3] Тимоти Стенли, Компресиране на данни, изд. “Интерфейс България”, София, 1998

[4] Dimitrina Polimirova, Eugene Nickolov, Cecko Nikolov, Investigating The Relations Of Attacks, Methods And Objects In Regard To Information Security In Network TCP/IP Environment, Proceedings of International Conference Information Theories and Applications: Cyber Security, 28-29 June 2006, Varna, to be published.

[5] David Salomon, Data Compression: The Complete Reference, Springer Verlag New York, Inc., 2004.

Authors' Information

Dimitrina Polimirova – PhD Student, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, e-mail: polimira@nlcv.bas.bg

Eugene Nickolov – Professor, DSc, PhD, Eng, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, e-mail: eugene@nlcv.bas.bg

Cecko Nikolov – PhD Student, Research Associate, National Laboratory of Computer Virology, Bulgarian Academy of Sciences, Phone: +359-2-9733398, e-mail: nikolov@nlcv.bas.bg