Strong Cryptography Armoured Computer Viruses Forbidding Code Analysis: the bradley virus

Eric Filiol

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Strong Cryptography Armoured Computer Viruses Forbidding Code Analysis: the BRADLEY *virus*

Eric Filiol

N° 5250

Juin 2004

THÈME 2

ISSN 0249-6399 ISRN INRIA/RR--5250--FR+ENG

INRIA
ROCQUENCOURT

Strong Cryptography Armoured Computer Viruses Forbidding Code Analysis: the BRADLEY virus

Eric Filiol*

Thème 2 — Génie logiciel
et calcul symbolique
Projet Codes

Rapport de recherche n° 5250 — Juin 2004 — 10 pages

Abstract: Imagining what the nature of future viral attacks might look like is the key to successfully protecting against them. This paper discusses how cryptography and key management techniques may definitively checkmate antiviral analysis and mechanisms. We present a generic virus, denoted BRADLEY which protects its code with a very secure, ultra-fast symmetric encryption. Since the main drawback of using encryption in that case lies on the existence of the secret key or information about it within the viral code, we show how to bypass this limitation by using suitable key management techniques. Finally, we show that the complexity of the BRADLEY code analysis is at least as high as that of the cryptanalysis of its underlying encryption algorithm.

Key-words: encryption, viral encryption, antiviral techniques, code disassembly, key management, computer security

also Virology and Cryptology Lab, Army Signals School, ESAT/DEASR/CSSI, 35998 Rennes, France, *efiliol@esat.terre.defense.gouv.fr*

* Projet Codes - Eric.Filiol@inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Un cas d'étude de blindage de code viral par utilisation de cryptographie forte en vue d'interdire l'analyse de code : le virus BRADLEY

Résumé : Ce rapport présente un cas détude illustrant comment les techniques cryptographiques peuvent mettre en échec total l'analyse et les techniques antivirales. Une preuve de concept est présentée au moyen d'un virus générique, appelé BRADLEY dont le code est protégé par chiffrement à l'aide d'un algorithme ultra-rapide et possédant un très haut niveau de sécurité cryptologique. Dans la mesure où le principal inconvénient dans l'utilisation du chiffrement réside dans le fait que la clef secrète ou une information suffisante pour la retrouver, la concernant, est toujours contenue dans le code, il est montré comment résoudre ce problème par l'utilisation de techniques de gestion environnementale de clefs de chiffrement. Il est alors démontré que la complexité du problème de l'analyse du code du virus BRADLEY est au moins aussi élevée que celle du problème de la cryptanalyse du système de chiffrement utilisé.

Mots-clés : chiffrement, cryptographie virale, techniques antivirales, désassemblage de code, gestion de clefs, sécurité informatiques.

1 Introduction

Antiviral detection is directly based on the capability to have malware codes at one's disposal and to study them by disassembly means. Thus, viral databases can be updated and antiviral engines can be upgraded.

A few malware writers try to make this task more difficult by implementing various techniques which aim at delaying the knowledge and the understanding of their codes: obfusctating, rewriting, encryption.... These codes are denoted *armoured* codes. The first and most famous one is probably the *Whale* virus appeared in the early nineties. More recently, the *MyDoom* virus very naively tries to complicate antiviral experts' work by implementing basic encryption techniques. Up to now, none of the known malware succeeded in preventing code analysis.

The main explanation for this failure lies on two facts:

- antiviral experts always manage to obtain a malware copy (infected file). As they are widely dissiminated, malware code samples (viruses, worms...) are always very easily available. This comes from the fact that limited virulence[1] is not a feature inherent to malicious codes.
- When present, techniques aiming at making code analysis more difficult are bound to fail. The main reason is that the related problems (that is to say, problems to be solved in order to bypass code protection) belong to polynomial complexity class. As an example, encryption techniques are always relatively easy to break since the key space is too limited and allows an exhaustive search approach. Moreover, encryption algorithms that have been found in known malware codes are either very naive or do not offer high level of security.

In this paper, we present a new concept about malicious codes combining efficient key management with high-level security encryption algorithm. Different analysis and experiments have confirmed the impossibility to study the code, under the assumption that we managed to get a copy of it. By limiting the code presence and virulence in the computer, we show also how to make this assumption very unlikely. We illustrate these concepts (proof-of-concept) by presenting from an algorithmic point of view the most simple example of a new virus family called the BRADLEY viruses. As a main result, we show that the general problem of BRADLEY code analysis is equivalent to the cryptanalysis of a secure encryption algorithm in the sense that it is exponentially complex.

This paper is organized as follows. In Section 2, we first define precisely the background and the known cases where attempts to use cryptography in viral codes have been made. We show why all this attempts were bound to fail. Section 3 recalls key management techniques presented in [11]. In Section 4, we present the generic viral concept[2] using strong encryption combined with optimal key generation and key management. At last, Section 5 prooves that BRADLEY viral code analysis is equivalent to encryption systems cryptanalysis and that it is, in fact, of exponential complexity. Conclusion will address the problem of fighting against such armoured malware.

The purpose of this paper is purely academic and draws our attention on the evolution of viral risks. It shows how the malware risk may evolve very quickly (if not already the case) and cause great concern among the antiviral community. This is the reason why we will not give any detailed code. The activity of our laboratory is dedicated to defensive aspects. Our mission consists in identifying new risks, in testing them in practice and assessing the level of potential threat precisely.

2 Definition and Background

2.1 Computer virology

The reader is supposed to be familiar with basic definitions about malware (virus, worms, trojan horses...) and antiviral techniques. We just recall the following starting definition of *armoured* codes.

Definition 1 *(Armoured codes) An armoured code is a program which contains instructions whose goal is to delay, complicate or forbid its own analysis during either its execution or through its disassembly.*

[1] *Virulence* is an index measuring the level of risk for self-reproducing codes. This index, hence the risk itself, is related to the number of copies of the malware code. For details, see [7, chapter 4, pp. 89ff].

[2] Without loss of generality, we will use the general term " *virus* " but everything presented in this paper may apply to any other malware type: trojans, worms, logical bombs...

The best known example is probably the *Whale* virus which appeared in september 1990. The virus did actually represent a very limited risk but it intended obviously to make its analysis very difficult. Its code contains roughly a dozen of program traps and tricks hampering trace, disassembling and code analysis: dynamic decryption/encryption, code obfuscation, code nesting...Once activated, the viral code tries to detect the potential use of a debugger and consequently freezes the keyboard. Using polymorphism techniques, about 30 different random variants were possible for an infected file.

What the *Whale* virus easily managed to cause is not a terrific epidemic but a waste of anti-virus experts' time and a nearly three-day delay to eradicate it. Nowadays, the main part of the viral action is completed during the first thirty minutes after the beginning of the infection (a good example could be the *Slammer* worm which appeared in january 2003); therefore, such a delay in code analysis cannot be acceptable. That is the reason why *armouring* code techniques must be seriously taken into account.

These techniques can be divided into different classes, as follows.

- code obfuscation: the aim is to transform a program into another which both is functionally identical to the original and more difficult to decompile and reverse engineer. In other words, the program is written in order to reduce readability and understandability. Three types of transformations are generally used: lexical transformations (variable name exchange), control flow transformations (making the program flow more complex by using code nesting or placebo code) and data flow transformations (action on data structures by changing storage, encoding, aggregation and order of data). More details can be found in [3, 5, 10, 14]. In a viral context, these techniques are of limited interest.

  Obfuscated codes are generally too slow and of too large a size to be efficiently used by undetectable malware. Moreover, the results presented in [2] show there is no transformation able to prevent the code of every program from revealing any other information except the program's input-output behaviour.

- polymorphism: the aim is to make the code change as often as possible by using rewriting techniques (equivalent functionnality but different code). The code analyst has thus to face up not a single version of the malware code but several versions of it; hence the difficulty to efficiently fight against it. Fortunately, code analysis becomes always possible in the end and, consequently, all polymorphic techniques are overcome.

- encryption: the purpose is both to provide polymorphism (encrypted code change with every different key) and to prevent code analysis. So far, known techniques suffer from a lack of efficient key management. The key must be securely available to the malware code only. Hardware solutions which are generally envisaged to prevent code analysis are not suitable for mobile codes like virus or worms. Mostly, the key elements are somehow or other contained in the code itself.

As these first two techniques are concerned, code analysts fortunately will always get to the end of it since these techniques always produce deterministic results (even if some algorithms may be partly probabilistic). The only thing is that the analysts need time to study the code behaviour instruction by instruction. Therefore, this study is likely to be time-greedy and requires many human resources and much effort dedication.

Encryption is maybe less obviously easy to handle contrary to what experience tells us so far. Fortunately, only naive or very unsecure encryption methods have been used in known malware [4]: constant masking (like in earlier macro-viruses for example), rot13 encryption or other weak encryption systems (as an example, *DarkParanoid* virus used very simple arithmetic functions – ADD, SUB, XOR, NEG, NOT, ROL, ROR – as encryption functions).... Besides, weak key management always allows to recover the key and then to decipher the malware code when dealing with strong cryptosystems.

2.2 Cryptology

The reader is supposed to be familiar with basic concepts of cryptology as well. A detailed monography about cryptography will be found in [8]. We will just recall previous uses of cryptology inside malware and a few useful concepts we will use throughout this paper.

Cryptology has previously been envisaged to provide computer virology with very efficient tools. On the one hand, cryptographic techniques have recently be considered as a means for optimal worm propagation [1]. The use of cryptographic hash functions, for instance, is suitable for speeding up *Curious Yellow worm* propagation [15].

On the other hand, the combination of cryptographic techniques with viral technologies led in 1996 to the concept of " *Cryptovirology* " as presented in [6, 16, 17]. Cryptovirology consists in applying cryptography tools to malicious codes in order to strengthen, improve or develop such codes.

Particularly, cryptography appears to be very efficient in designing payloads. Several convincing examples are presented in [16]. The main goal is to make a victim host dependent upon the virus – *i.e.* a virus can survive in the host if it makes the host depend in a critical way on the very presence of the virus itself. These results are mainly obtained with public-key cryptographic techniques[3] combined with limited symmetric cryptography techniques.

Though very efficient, these approaches aim only at protecting the action of the virus (its *payload*) but not the virus itself. In other words, if a copy of a cryptovirus is somehow or other obtained and analyzed by reverse engineering, none of the cryptographic tools it contains will totally protect them against its code analysis. Thus, the exact knowledge of the code is likely to allow antiviral software update and limit/forbid the malware's action. The main limitation comes from the fact that cryptovirus as defined by Young and Yung, is not able to manage secret key part in a suitable and efficient way for that particular purpose.

The basic technique we discuss in this paper can effectively forbid such code analysis and thus, properly complement all the approaches developped in [6, 16, 17]. Other more sophisticated techniques are being tested in our laboratory.

3 Environmental Key Generation

Malware are mobile agents by nature. If they pass through an " insecure network " or environment (from the malware's point of view), they may be analyzed (disassembled) so that their code will be completely accessible to the attacker (the analyst). As previously explained , traditional encryption systems are dealing with static keys. Actually, the key is present somehow or other in the agent (hardware or software).

In 1998, B. Schneier and J. Riordan [11] introduced the notion of *environmental key generation* to address this problem. In other words, keying material is contructed from certain classes of environmental data. Environmental key generation can thus be useful when the sender wishes to communicate with the receiver such that the receiver could only receive the message if some environmental conditions are true. Environmental key generation can even be used in circumstances where the receiver is not aware of the specific environmental conditions that the sender wants his communication to depend on. This latter case corresponds exactly to our malware code analysis problem. The receiver here is the malware code present in a computer (the environment) and the sender is malware code author or the target system itself.

The difficulty with building an environmental key generation protocol is that the threat model assumes that any attacker (the malware code analyst) has total control over the environment. All information available to the malware program can be found by the attacker as well. All inputs to the program are supplied by the attacker and the program states themselves is completely determined by the attacker during the code analysis. As such, the constructions must resist direct analysis and dictionnary attacks in the form of Cartesian deception, that is to say, in which the attacker tells lies about the environment.

The authors of [11] discuss several construction for environmental key generation. To illustrate their approach, let us consider the following basic construction. Let N be an integer corresponding to an environmental observation, $\mathcal{H}$ a one-way function (typically a hash function), M the hash of the observation N, $\oplus$ the bitwise exclusive-or operator, $||$ the concatenation operator, R a nonce and K a key. The value M is carried by the agent (the malware code in our case). Hash function can be used to conduct tests and construct the keys so that examination of the agent does not reveal the required environmental information. Then possible constructions, among many others, are:

- if $\mathcal{H}(N) = M$ then let $K = N$.
- if $\mathcal{H}(\mathcal{H}(N)) = M$ then let $K = \mathcal{H}(N)$.
- if $\mathcal{H}(N_i) = M_i$ then let $K = \mathcal{H}(N_1, N_2, \ldots, N_i)$.
- if $\mathcal{H}(N) = M$ then let $K = \mathcal{H}(R_1, N) \oplus R_2$.

Let us note that the first construction is used in most of static encrypted password authentication schemes. The most important feature of each of these constructions is that knowledge of M does not leak any information on K.

Riordan and Schneier proposed several efficient constructions which provide efficient environmental key generation protocols using various techniques: thresholding (protocol using the ideas of cryptographic secret

[3] A *cryptovirus* is defined as a computer virus that contains and uses a *public key*.

sharing), nesting (action of the mobile agent is ruled by several environmental keys used in the sequential way), time indexation (part of the environmental date required to generate the key are based on time)...

Environmental key generation has only been proposed from a theoretical point of view by the authors. Some aspects still need to be thoroughly tested. Particularly, for most of the constructions they proposed, the attacker is likely to find the key by observing both the agent and the environment. The search space for the activation data may always be small enough to allow an exhaustive search approach. Moreover, by observing mobile agent actions, the attacker may easily determine where and which kind of data the agent is interested in. That implies that a patient analyst will obtain information about the agent at the same time this latter is activated by the suitable environmental data.

We now present a practical and efficient use of environmental key generation in the case of viral code armouring.

4 A Generic Armoured Virus: the BRADLEY virus

Let us discuss now the generic family virus named BRADLEY. Without loss of generality, we choose to describe only a basic but powerful example. Some More complex protocols have been developped or are currently under study (see Section 6). Two different codes have been developped and tested:

- a directed but generic virus which aims at specifically infecting a given group a machine/people (variant A),
- a directed virus dedicated to specifically strike only any given user (variant B).

Minor variants have been tested as well and will be listed later on. The codes have been designed both for Windows and Unix systems. They successfully managed to bypass antiviral software which all remained silent. Since BRADLEY viruses are only proof-of-concept viruses, we will focus only on the armoring protocol part. Complete source code is not available. The general structure of the codes is given in Figure (1) and summarized

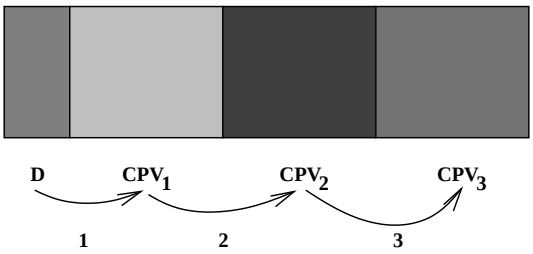

Figure 1: Overall structure of BRADLEY codes

as follows:

- a decipherment procedure D which purpose is to collect activation data, test and evaluate them and finally decipher the different encrypted parts of the code;
- a first encrypted part EVP_1 with encryption key K_1. Once deciphered, this part installs all anti-antiviral functions (passive and active).
- a second encrypted part EVP_2 with encryption key K_2. This part contains the infection functions and the polymorphism procedures. When replicating, the virus will always and completely change its form (including the decipherment procedure).
- a third part EVP_3 (optional) with encryption key K_3. It contains the payload functions (in our case, a simple opening window issuing a infection warning in order to keep control over the virus).

Note that these three encrypted parts are exactly of the same size in order to give the slightest information on the underlying code.

Let us now describe the key management protocol. The activation data – in other words the data required to construct the different keys – are (variant A):

- the local DNS address (e.g. @company.com), denoted α,
- current system time (hours hh only) and date (mmdd), denoted δ,
- a particular data present in the target system(s) (in our case a particular file), denoted ι,
- a particular information under viral code author control, located outside the system (public channel) but easily accessible to the virus (in our case, a given web page containing a particular value whose presence is limited in time and related to the value δ); it is denoted π and obtained from the hash of this information[4]

For the variant B, data ι is a given public key which is present in a *pubring.gpg* for example. Thus, the virus may target a particular user or users communicating through encrypted emails/data with any given user. The viral code uses the hash function SHA-1 [9] as one-way function (here denoted $\mathcal{H}$). Then, the environmental key protocol is described as follows:

1. the decipherment procedure D collects the activation data either directly (α, δ and π) or repeatedly[5] (ι) and compute a 160-bit value V given by $\mathcal{H}(\mathcal{H}(\alpha \oplus \delta \oplus \iota \oplus \pi) \oplus \nu)$ where ν is the first 512 bits of EVP_1 (in its encrypted form).
2. if $V = M$ where M is the *activation value* contained in the viral code, then $K_1 = \mathcal{H}(\alpha \oplus \delta \oplus \iota \oplus \pi)$ otherwise the decipherment procedure stops and disinfects the present system from the whole viral code.
3. D deciphers EVP_1 producing $\mathrm{VP}_1 = D_{K_1}(\mathrm{EVP}_1)$ and launches it. Then D is computing $K_2 = \mathcal{H}(K_1 \oplus \nu_2)$ where ν_2 is the 512 last bits of VP_1.
4. D deciphers EVP_2 producing $\mathrm{VP}_2 = D_{K_2}(\mathrm{EVP}_2)$ and launches it. Then D is computing $K_3 = \mathcal{H}(K_1 \oplus K_2 \oplus \nu_3)$ where ν_3 is the 512 last bits of VP_2.
5. D deciphers EVP_3 producing $\mathrm{VP}_3 = D_{K_3}(\mathrm{EVP}_3)$ and launches it.
6. After virus action is completed, the virus disinfects itself totally.

Some remarks can be made about this protocol:

- from replication to replication, the whole code (including procedure D and value M) has completely changed every time. This implies a total control of the polymorphic procedure relatively to the key management protocol by the author of the viral code (*i.e.* the evolution of the activation data – in practice only the values δ and/or π).
- the purpose of values C_x is to make the data span the whole input space (512 bits).
- the different parts VP_i may be compressed before encryption.
- the keys K_1, K_2 and K_3 can be made independant by using additional environmental data.
- the auto-disinfection may be delayed in order to handle the time and date values in a less strictly way. In that case, the decipherment procedure D remains active in system memory.

Other variants have been tested as well, particularly to produce the most optimal code in terms of size and stealthiness. The most significative variant is the following:

- the underlying code is compressed,
- instead of embedding compression and encryption functions within the virus code, this latter will borrow local resources if present,

[4] One may object that the presence of the webpage url within the procedure D could give a useful information to the analyst. Since the webpage is under malware author's control, it is a very dunious hypothesis that the attacker could successfully access to the suitable activation data, especially if the data availability is very limited in time. Nonetheless, we have developped a variant of the basic protocol which is discussed in this section. Instead of only one data π, we use two external activation data π and π'. Each of them come from two different webpages. The second webpage's url is encrypted by means of the key contructed from data α, δ, ι and π. Once decrypted, the virus gets the second activation data π' and a secret permutation function P (at the very beginning of EVP_1). Finally, the key K_1' is build from data $P(\alpha), P(\delta), P(\iota)$ and π' in the same way as key K_1 is. In this variant, K_1 is superseded by K_1'.

[5] " Repeatedly " means here that the virus scans any data contained in the system. In our case (a given file), the virus looks recursively for that data through the tree file system.

- that implies that one more activation data is required and repeatedly scanned for: existence or not of compression and encryption softwares.

For all variants we developped, encryption algorithms that have been used are RC4 [12] and RC6 [13] while gzip compression has been chosen.

5 Viral Code Analysis and Cryptanalysis

To evaluate the code analysis complexity, two cases must be considered:

- the analyst has the viral binaries at one's disposal,
- the analyst does not have them.

The second case is more likely to happen if we consider that, in any case, the virus limits its presence inside the target system by disinfecting itself from it.

But let us suppose that the analyst, even if it is very unlikely, has managed to get one copy of the virus binaries. Let us show that the environmental key generation protocol presented in section 4 efficiently forbids code analysis unless a cryptanalysis problem of exponential complexity is solved.

Proposition 1 *The analysis of a code protected by the environmental key generation protocol defined in Section 4 is a problem which has exponential complexity.*

Let us now prove this proposition.

Proof.
Firstly, let us remark that decipherment procedure D leaks only the following information:

- the activation value V,
- the fact that the virus looks for the system time and date,
- the fact that the virus scans for specific data α, ι and π.

Moreover, the analyst is able to analyze the virus if and only if he knows the secret key K_1. It can be obtained either by direct cryptanalysis or by guessing the exact values of the different activation data required to generate the good key. This guessing is equivalent to dictionary attacks.

The cryptanalysis approach aims at finding the value K_1 such that $\mathcal{H}(K_1 \oplus C_1) = M$ where M and C_1 are easy to identify in the decipherment procedure D. A hash function is highly non injective by nature. Thus it cannot be computationaly inversed in any way (*preimage resistance*). Consequently, this problem must be reformulated as a collision search problem (for more details, refer to [8, Chapter 9]). In other words, find all pairs of input x and x' such that $\mathcal{H}(x) = \mathcal{H}(x')$. This problem itself is computationally infeasible. To be more precise finding such a pair requires $2^{\frac{n}{2}}$ operations for n-bit input values ($n = 512$ for SHA-1). Since the analyst must absolutely find the exact key K_1 (secret key really used to encrypt the viral code), he must beforehand compute all the values x such that $\mathcal{H}(x) = M$. For a n-bit input, m-bit output hash function, there exists 2^{n-m} such x (2^{352} for SHA-1). Then, to summarize, recovering the key requires $2^{\frac{n}{2}} \times 2^{n-m}$ operations – that is to say $2^{\frac{n^2-m}{2}}$ operations ($\approx 2^{131,072}$ for SHA-1).

Let us now consider the dictionary attack approach. It consists in enumerating all the possible values that might have been used as activation data. Note that, in that particular case, the analyst must simultaneously consider both the encrypted viral code and the system in which the code has been found. The analyst can try all the possible data relevant to the system (that is to say α, ι and δ) over which he has control during the analysis. Unfortunately, data π remains out of his control and thus he will not be able to determine its exact value. Thus there is no any other more efficient approach than searching exhaustively for the value $\alpha \oplus \delta \oplus \iota \oplus \pi$. Since at least π will be chosen randomly by the viral code author, this exhaustive search has complexity 2^n if a n-bit input hash function has been used (2^{512} for SHA-1). All things considered, the overall complexity of the code analysis is $\min(2^n, 2^{\frac{n^2-m}{2}}) = 2^n$. ■

6 Conclusion

The proof-of-concept virus BRADLEY has been designed and discussed to illustrate the fact that efficient armouring is possible. BRADLEY and other efficient viruses of same kind pose the problem of a threat which so far, is impossible to deal with. The polymorphic nature of such codes, when optimally implemented, forbids any detection based only on the decipherment procedure D. During the experiments, detection based on behaviour monitoring and analysis has been successfully bypassed as well.

Permament and direct memory monitoring might be a solution to deal with such efficient armoured codes. Besides, heavy system ressources are required and this approach implies to be aware of this particular threat (efficient code armouring). Current research carried out in our laboratory aims at proving that even memory management and monitoring can be bypassed. We particularly designed a far more complex variant directly drawn from the *DarkParanoid* virus: at any time, only a single instruction can be found in an unencrypted form in memory. But it requires a far more complex environmental key generation than that simple one presented in Section 4.

Unless a solution is rapidly found to fight against these virus, this study outlines that an isolation of critical networks and a strict computer security policy is absolutely essential. Moreover, this implies that the antiviral companies must develop cryptanalysis skills in the very near future, under the assumption that it is possible to obtain a viral code sample and that breakable cryptosystems have been used.

References

[1] Balepin I. (2003), *Superworms and Cryptovirology: a Deadly Combination*, `http://wwwcsif.cs.ucdavis.edu/~balepin/new_pubs/worms-cryptovirology.pdf`

[2] Barak B., Goldreich O., Impagliazzo R., Rudich S., Sahai A., Vadhan S, Yang K. (2001), *On the (im)possibility of obfuscation programs*, Advances in Cryptology, Crypto 2001, LNCS 2139, pp. 1–18, Springer Verlag.

[3] Chow S., Eisen P. Johnson H., Zakharov V.A. (2001), *An approach to the obfuscation of control-flow of sequential computer programs*, Information Security, ISC 2001, LNCS 2200, pp. 144–155, Springer Verlag.

[4] Ciubotariu M. (2003), *Virus Cryptoanalysis*, Virus Bulletin, november. `http://www.virusbtn.com/magazine/archives/200311/cryptoanalysis.xml`

[5] Collberg C.S., Thomborson C. (2002), *Watermarking, tamper-proofing and obfuscation - Tools for software protection*, IEEE Transactions on Software Engineering, Vol. 28, No. 8, August 2002, pp. 735-746.

[6] `http://www.cryptovirology.com/`

[7] Filiol E. (2003), *Les virus informatiques : théorie, pratique et applications*, Collection IRIS, Springer.

[8] Menezes A., van Oorschot P., Vanstone S.A. (1997), *Handbook of Applied Cryptography*, CRC Press.

[9] National Institue of Standards and Technology, NIST FIPS PUB 180, *Secure Hash Standard*, U.S. Department of Commerce, May 1993.

[10] Project funded by the Fund for Scientific Research - Flanders (2003), *Coordinated Research of Program Obfuscation*, `http://www.elis.regent.be/~banckaer/obfuscation/proposal.html`

[11] Riordan J., Schneier B. (1998) *Environmental key generation towards clueless agents.* In *Mobile Agents and Security Conference'98*, G. Vigna ed., Lecture Notes in Computer Science, pp 15–24, Springer-Verlag, 1998.

[12] Rivest R.L. (1992), *The RC4 Encryption Algorithm*, RSA Data Security Inc.

[13] Rivest R.L., Robshaw M.J.B. Robshaw, Sidney R. and Yin Y.L. (1998), *The RC6 block cipher*, Proc. of the 1st AES Candidate Conference, August 1998, Ventura.

[14] Shah P. (2002), *Code Obfuscation For Prevention of Malicious Reverse Engineering Attacks*, `http://islab.oregonstate.edu/koc/ece478/02Reports/S2.pdf`

[15] Wiley B. (2002), *Curious Yellow: The First Coordinated Worm Design*, `http://blanu.net/curious_yellow.html`.

[16] Young A, Yung M. (1996), *Cryptovirology: Extorsion-Based Security Threats and Countermeasures*, IEEE Symposium on Security and Privacy, Oakland, CA, 1996.

[17] Young A, Yung M. (2004), *Malicious Cryptography: Exposing Cryptovirology*, Wiley & Sons.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399