Concepts and Future Trends in Computer Virology

Eric Filiol
efiliol@esat.terre.defense.gouv.fr

ESAT
Laboratoire de virologie et de cryptologie
Rennes

XXth CISE 2007 Plenary Talk

Plan

Introduction

- The computer viral hazard is somehow recent: less than 30 years.
- Existence of a malign will: cybercriminals.
  - High adaptative and organisational capabilities.
  - They are well-off and very well equiped.
- Defence progress far slower than the attacking side.
- Failure of the software industry: vulnerabilities, antivirus highly limited efficiency.
- General issue of users ' "computer hygiene".

Introduction (2)

- The attackers' vision is never neither taken into account nor even proactively considered.
  - Legal Issues (France $\Longrightarrow$ LCEN 2004).
  - Publishing reproducible scientific results is a critical issue.
- The attacker's view is essential to whom has to defend.
- Antiviral protection must consider a permanent technological watch along with a proactive research.

Introduction (3)

Postulate

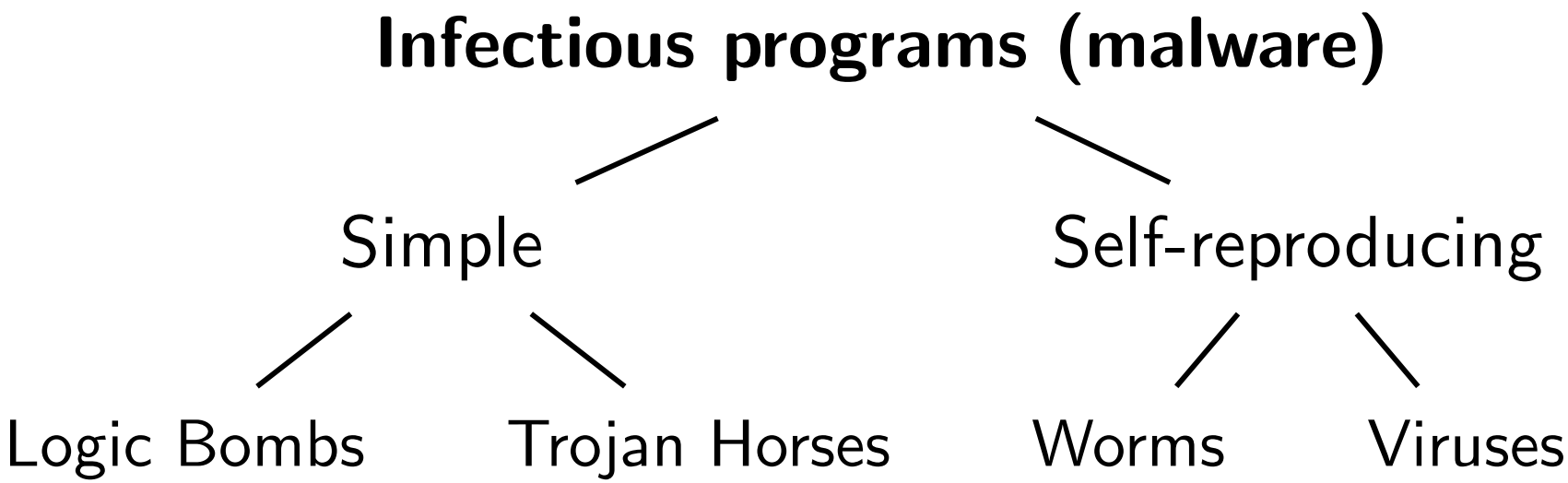
*Infectious programs (*malware*) exist for every execution-capable environment!*

- Every operating systems.
- Mobile environments (cell or smart phones, games consoles, GPS, onboard computers...).
- Almost every file formats.

Summary of the talk

Plan

Adleman's classification



- A malware is only a program!
- There is no malware normalisation yet.
- Present trend: "modern" malware cumulate all functionalities (e.g. *Botnets*).

Simple Malware

Definition

Logic bomb.- *Resident malware, which installs itself into the system and waits for some* trigger *incident or event (data present or absent in the system, a specific system date...) before performing an offensive function (trigger mechanism).*

Definition

Trojan horse.- *Program made of two parts namely the* server *module and the* client *module. The server module, once installed in the victim's computer secretly enables the attacker to access to victim's hardware and software resources. The attacker can use them via networks (via the client module).*

Self-reproducing Malware

Definition

Virus.- *A virus can be described by a sequence of symbols which is able, when interpreted in a suitable environment (a machine), to modify other sequences of symbols in that environment by including a, possibly evolved, copy of itself.*

Definition

Worms.- *Network-oriented virus. The essential difference lies on the fact that some worms are no longer attached to an infected file (malicious process only; e.g* Slammer *or* CodeRed*).*

Computer Worms

Three main classes.

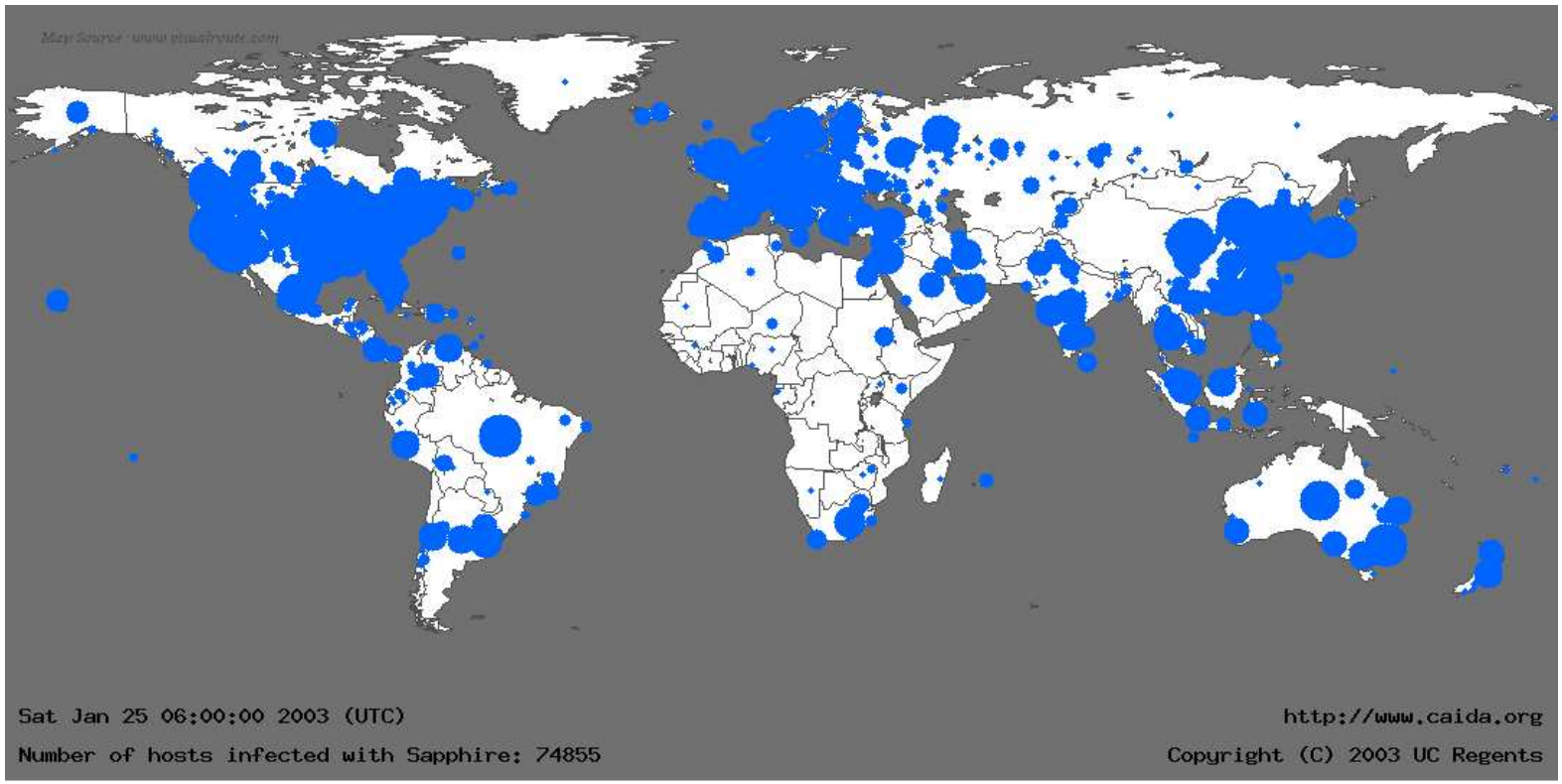
- *I-Worms* (or simple worms). Operate by using software security vulnerabilities (*Slammer, Sasser...*).
- *Macro vorms*. Use of social-engineering and of a malicious email attachment (document; e.g. (*Melissa*).
- *Email-worms* (or mass-mailing worms). Use of social-engineering and of a malicious email attachment (executable file; *Bagle, NetSky*).

Computer Worms (2)

Worms	Propagation
CodeRed (2001)	14 H
Slammer (2003)	30'
P-o-C (2005 - 2007)	1"

- Very high potential propagation speed.
- The current trend (since 2004) consist in reducing the propagation speed to the benefit of stealth.

Slammer Worm Attack (2003)



Anti-antiviral techniques

Definition

Stealth.- *Techniques aiming at convincing the user, the operating system and antiviral programs that there is no malicious code.*

Definition

Code mutation.- *Capability to self-modify (mutate) his own code (rewriting, encryption) in order to bypass any sequence-based detection.*

Definition

Code armouring.- *Techniques whose goal is to delay, complicate or forbid code analysis during either the execution or through the disassembly.*

Malware Life Cycle

There are five phases.

- Design and testing phase.
- Transmission and infection phase.
- Incubation incubation.
- Offensive phase.
- Detection and eradication (removal) phase (if any).
  - **The last phase does not systematically occur!**

Operational Aspects

Ways of disseminating malware:

- Data exchange.
- Mobile and onboard environments.
- Social engineering.
- Software vulnerabilities.
- Security policy deficiencies.

Plan

State-of-the-Art

- There are very few theoretical results. In the last 20 years:
  - Less than 15 theoretical papers.
  - Less than 10 PhD thesis.
- The lack of true and independant research in the field is beneficial to the attacking side.
- It is the AV community's direct responsability.

Fred Cohen's Results

Seminal research of Fred Cohen (1984 - 1988)

- Formalisation work on self-reproducing programs.
- "*Virus detection is an undecidable problem.*"
- Theoretical concept of virus mutation.
- Propagation studies.
- Study of some security models: the only efficient model consists in totally isolating systems.

The Other Works

Mainly studies on the complexity with respect to some classes of the detection problem.

- Adleman (1989).
- Spinellis (2003).
- Zuo & Zhou (2004, 2005).
- Bonfante, Marion & Kaczmarek (2005).
- Filiol (2006 - 2007).

Most of the viral class are at least NP-complete. Consequently, viral detection becomes untractable in practice, very soon.

Consequences

Corollary

Claiming to "detect any virus, including unknown ones" is a lie.

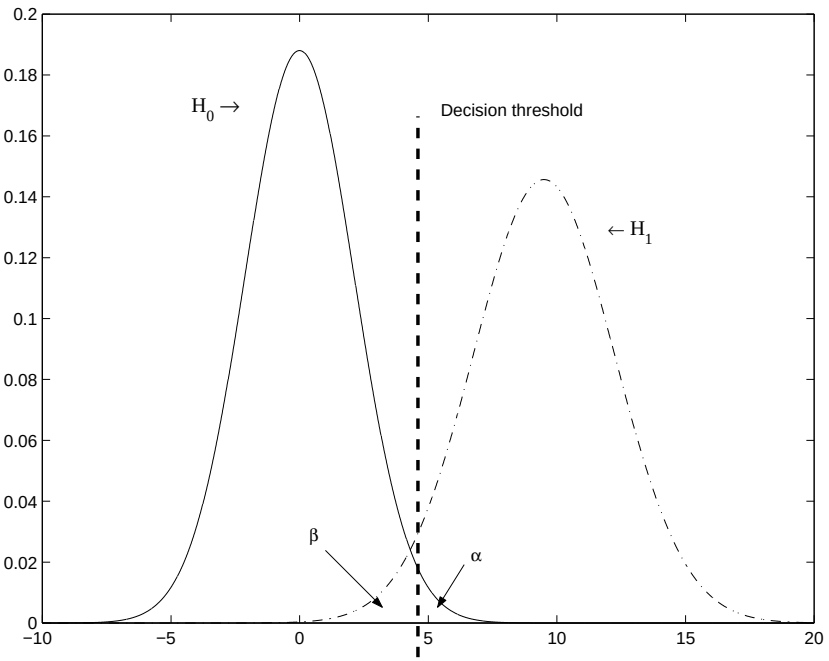
There is an equivalence between the problem of detecting many classes of virus with some other well-known problems:

- Cryptanalysis of public-key cryptosystems.

It remains still very easy to bypass any existing antivirus software.

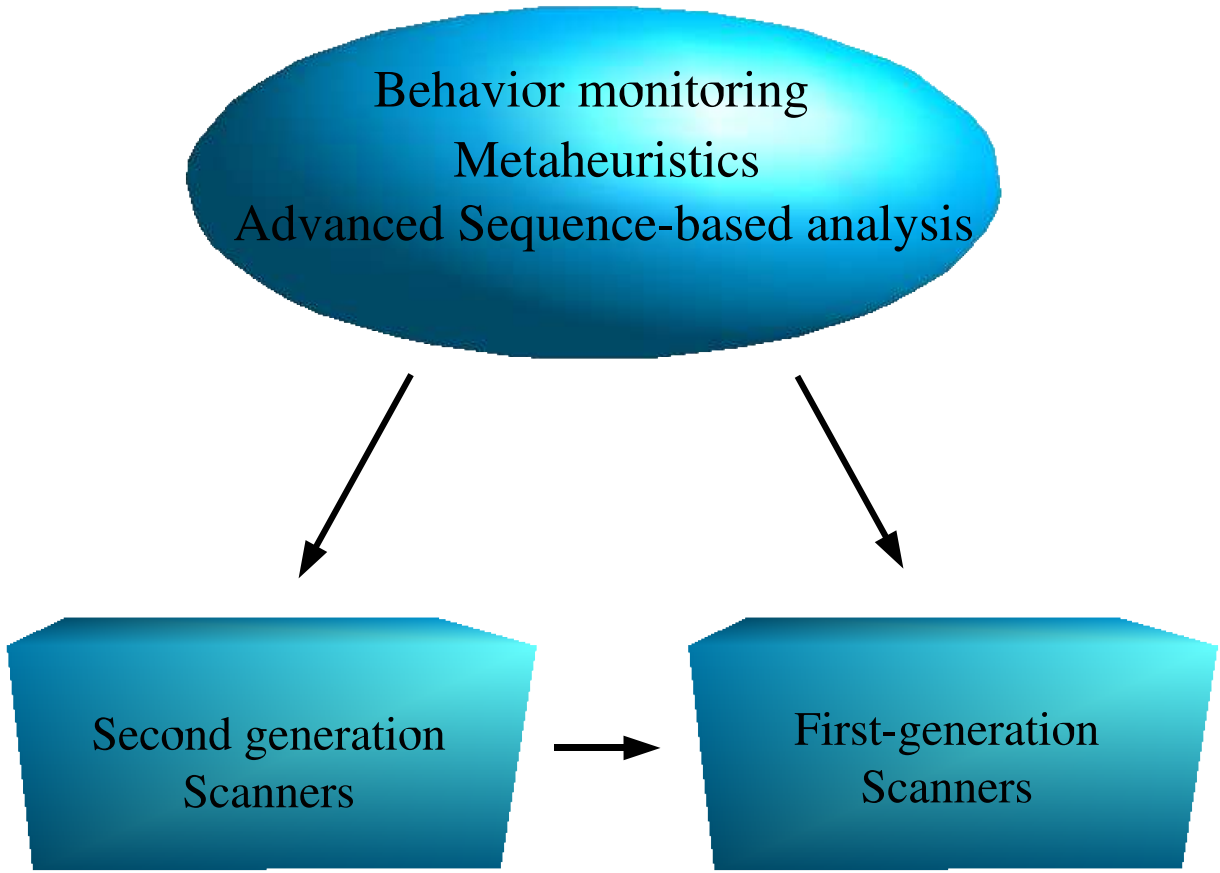
Plan

General Principles



- Any set of detection techniques can be modeled as a statistical testing (Filiol & Josse - 2007).
  - False positive and non detection probabilities.
  - These two different errors are opposite one of this another. Any AV designer has to make a strategic choice between them.
- The probability law which describes the infectious process ($\mathcal{H}_1$) is generally unknown.

General Structure of Antivirus



Sequence-based Detection

The code is analysed in a non-execution context.

Fact

(Filiol 2006; Filiol - Jacob - Le Liard 2006) Every existing antivirus still relies quite exclusively on sequence-based detection.

- The 14 main antivirus have been analysed:
  - All the detection functions and patterns are all weak and trivial.
  - There exists a large similarity from one antivirus to another one.

$\Rightarrow$ Existing AV are can be bypassed far too easily!

W32/Bagle.P Detection Scheme

Product	Pattern size (in bytes)	Signature (indices)
Avast	8	12,916 $\longrightarrow$ 12,919 12,937 $\longrightarrow$ 12,940
AVG	14,575	533 $\longrightarrow$ 536 - 538 - ...
Bit Defender	8,330	0 - 1 - 60 - 128 - 129 - 134 - ...
DrWeb	6,169	0 - 1 - 60 - 128 - 129 - 134 - ...
eTrust/Vet	1,284	0 - 1 - 60 - 128 - 129 - 134 - ...
eTrust/InoculateIT	1,284	0 - 1 - 60 - 128 - 129 - 134 - ...
F-Secure 2005	59	0 - 1 - 60 - 128 - 129 - 546 - ...
G-Data	54	0 - 1 - 60 - 128 - 129 - 546 - ...
KAV Pro	59	Identique à *F-Secure*
McAfee 2006	12,1278	0 - 1 - 60 - 128 - 129 - 134 - ...
NOD 32	21,849	0 - 1 - 60 - 128 - 129 - 132 - 133 - ...
Norton 2005	6	0 - 1 - 60 - 128 - 129 - 134
Panda Tit. 2006	7,579	0 - 1 - 60 - 134 - 148 - 182 - 209...
Sophos	8,436	0 - 1 - 60 - 128 - 129 - 134 - 148...
Trend Office Scan	88	0 - 1 - 60 - 128 - 129 - ...

Testing of www.virus.gr - August 2006

Produits	%
KAV	99,62
F-Secure	96,86
Bit Defender	96,63
NOD32	95,14
McAfee	93
Norton	83,18
Sophos	69,48
eTrust	50,36

- Exhaustive scanning of 147,184 known malware.
- Optimal configuration for the detection.
  - Optimised setup.
  - Heuristics all activated.

Behaviour-based Detection

The code is analysed in an execution context. The potentially dangerous actions are searched for.

- These techniques are in fact not frequently used directly (Filiol - Jacob - Le Liard 2006).
- Sequence-based detection is used for validation purposes.
- When implemented, behaviour-based detection can be easily bypassed (τ-obfuscation, polymorphic behaviours...).

Plan

General Principles

The attacker will more and more exploit the fact that any antivirus is a commercial product above all else!

- Antivirus and malware do not share the same constraints.
  - A malware can operate within tens of minutes. Not an antivirus!
- Design of malware as difficult, complex or undecidable instances of the detection problem.
- New viral models.

Stealth

Fact

(Mike Danseglio - Microsoft - 2006) " When you are infected by very sophisticated *rootkits* or *spyware*, the only solution is to start again from scratch. In some particular cases, there is no other way to go back to a stable system than formatting and reinstall everything!"

- Virtualisation-based rootkit:
  - *SubVirt*-like techniques (Microsoft/Univ. Michigan 2006).
  - *BluePill*-like techniques (Vista attack - Rutkowska 2006).
- Detection must now be done from outside the system.

Advanced Code Mutation

Polymorphism and metamorphism techniques will become too complex. Protection is consequently bound to fail in the future.

- Modelisation by formal grammars and languages (Filiol - CISE 2007).
  - Classical code mutation: the mutation language can be easily decided.
    $\Rightarrow$ the word "easily" is an english one.
  - Advanced code mutation: the mutation language is difficult to be decided or even undecidable.
    $\Rightarrow$ is the word "dot" an English, French or an Indonesian one?
- Behaviour-based detection can be easily bypassed:
  - Slowing-down of the "translation" process in a metamorphic malware.
  - Behavioural or mimetic code mutation (Filiol - Jacob - Le Liard, 2006).

Code Armouring

The code analysis enables to guess what the malware really did, to understand how it works and eventually to update antivirus.

- Software-driven analysis frequently fails where human-driven analysis always succeeds (up to a time factor).
- Light armouring techniques by τ-obfuscation (Beaucamps - Filiol 2006).
- Total armouring techniques (*Bradley* codes, (Filiol, 2005)).

New Viral Models

Present viral models are not the only existing ones.
K-ary Codes (Filiol 2007) :

- The malware information is divided up among many files.
- Sets of k codes in cooperative mode:
  - Parallel mode.
  - Sequential mode.
- Every of the k parts looks like an innocuitous one.
- Detecting K-ary codes is a NP-complete problem.

Plan

Conclusion

- Gloomy future with respect to the existing context only:
  - Antivirus are essential but their efficiency will be more and more limited.
  - Detection versus eradication.
  - Antivirus just notice an already old problem.
- Facing some sophiticated malware, the only solution is to prevent them from infecting the system.
- Security policies must be prevalent over any antivirus.
- Malware are a social problem:
  - Can we keep on opening systems?
  - Can we accept network interconnexion without limits?
- Security et ergonomics are mutually exclusive.

Future Prospects

- Antiviral protection must be supported by a theoretical and applied, independant research.
- Dual problem of results reproducibility.
- Computer world actors must have their responsability redefined:
  - Decision-makers.
  - Software editors.
  - Users (including administrators).

Many thanks for your attention.

Bibliography

- P. Beaucamps et E. Filiol. *On the possibility of practically obfuscating programs - Towards a unified perspective of code protection*. WTCV'06 Special Issue, G. Bonfante & J.-Y. Marion eds, Journal in Computer Virology, 2 (4), 2006.
- E. Filiol. *Computer Viruses: from theory to applications*. IRIS International Series, Springer Verlag, 2005.
- E. Filiol. *Techniques virales avancées*. Springer Verlag France, 2007 (an English translation is due end of 2007).
- E. Filiol - G. Jacob - M. Le Liard. *Evaluation Methodology and Theoretical Model for Antiviral Behavioural Detection Strategies*. WTCV'06 Special Issue, G. Bonfante & J.-Y. Marion eds, Journal in Computer Virology, 2 (4), 2006.
- E. Filiol. *Malware Pattern Scanning Schemes Secure Against Black-box Analysis*. EICAR 2006 Special Issue, V. Broucek & Paul Turner eds, Journal in Computer Virology, 2 (1), 2006.