Vade Secure
Predictive Email Defense

White paper

Heuristic and behavioral methods for countering 0-day malware and ransomware attacks

1 Malware in e-mails: insignificant volume, significant consequences

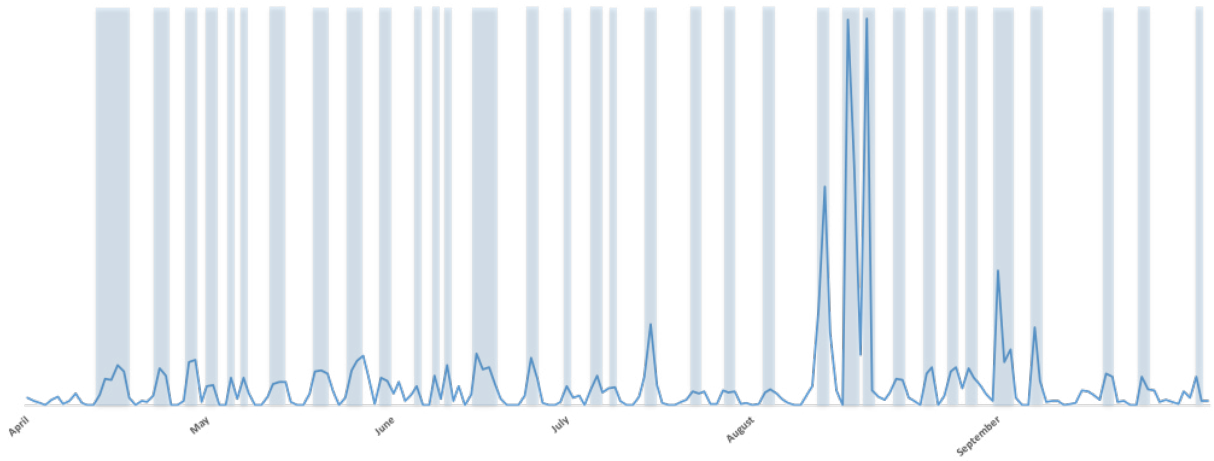
E-mail is the leading vector for malware of all types. It indeed allows reaching a significant number of users easily. Malware may have different objectives:

- Mass mailing of spam,
- Causing denial of service attacks,
- Gathering confidential data (documents such as patents or contracts, credit card details, identifiers, social security numbers, etc),
- Encrypting the contents of a machine with the aim of demanding a ransom for its decryption.

Malware represents a very small percentage of global e-mail traffic:

- As a matter of fact, only **0.02%** of e-mails sent worldwide contain malware.
- In a corporate environment, the presence of malware is even stronger. It represents **0.7%** of e-mail traffic.

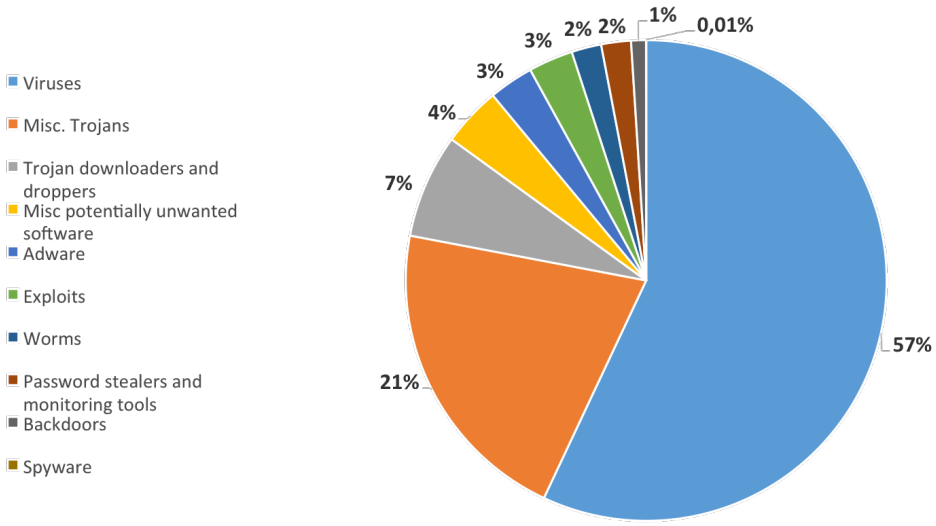
Malware is sent in seasonal waves. It may depend on current events (e.g. terrorist attack, earthquakes, tsunamis, etc) to take advantage of the user's heightened attention, or may stem from a need to maintain or extend the environment of infected machines. The graph below represents the trend and frequency with which malware have been transmitted from April to September 2015.



Data measured from 300 million mailboxes in 76 countries.

The main anti-malware techniques used by the major specialized players are not effective enough to thwart 100% of attacks. Since anti-malware detection techniques are not confidential, they are now sufficiently known to cybercriminals for them to craft countermeasure techniques that can be controlled remotely or fully automated.

Malware takes on very different forms, each of which have a particular objective. The graph below represents the distribution of currently active types of malware:



The corporations listed below are some of the institutions with the strictest security policies and use solutions from the biggest names in information security. Despite intense focus on the protection policy, attacks still manage to weave through security barriers:

- State of South Carolina - 1.9 million Social Security numbers stolen (October 2012)
- LinkedIn - 6 million passwords stolen (June 2012)
- Global Payments - 1.5 million credit card numbers stolen (March 2012)
- State of Utah - 780 000 medical records stolen (March 2012)
- Sony – Playstation Network data theft (April 2011)
- RSA – Intellectual property theft on SecurID (March 2011)

These events can be explained by the fact that malware developers focus above all on going through the most widely used anti-malware products. Indeed, cybercriminals place priority on developing malware that are transmitted through the most frequently used technologies and products in order to reach the largest number of victims. The cybercriminal will therefore maximize the return on his investment.

2 The main techniques used

All the main techniques used involve analyzing the content of the infected file and its possible impact on the system.

The main technologies used are set out below.

File signatures:
A digest of the file is calculated, in general in the form of a hash with a varying degree of agility (MD5, SHA1, SSDEEP). A centralized database stores signatures and spreads them directly on anti-malware filters.

Heuristic rules on file contents:
A signature allows detecting one or several malware. However, it is often strict and can easily be bypassed by polymorphic malware, meaning that they can produce an infinite number of hashes for the same operational objective. The aim of heuristic rules is to rely on the malware's operational behavior in order to be less sensitive to variations in its contents. Heuristic rules therefore allow identifying malware that has yet to be known to laboratories by identifying similar behavior in malware with different contents.

Sandboxing:
Sandboxing consists of executing a suspicious file in an isolated virtual environment. In general, sandboxing techniques provide disk space, RAM space and controlled network access in order to intercept packets there.

Although security vendors innovate and improve their systems, cybercriminals remain agile and adapt to updates. Now, malware are able to identify security layers and adapt their operation accordingly.

3 Countermeasures used by malware

a) Mutant malware: In order to evade signature systems and heuristic systems on file, many malware become mutant, meaning that they are created from an existing malware but with modified content. As such, anti-malware vendors would have to detect this evolution then create a new signature. The length of time between the publication of the mutant malware and the creation of the new signature allows the creator of the malware to infect new victims.

b) Polymorphic malware: These malware are able to dynamically change forms each time they are replicated, preventing anti-malware software from identifying them by their signature. However, the way the malware operates remains unchanged - the algorithms are the same - only their translation to machine code is modified.

c) Detection of a virtual environment: In order to evade sandboxing, an increasing number of malware are equipped with techniques for detecting virtual environments, as they only wish to be operational on physical machines. In order to do this, they detect specific drivers on virtual machines or even files that are only found on the disk when a hypervisor such as Vmware, KVM, Xen, etc is used.

d) Waiting for user interaction: Also for the purpose of evading sandboxing, malware are equipped with chunks of code that will only be executed if the user interacts with the environment, for example by moving the mouse, the presence of a browsing history, keyboard usage, etc.

e) Partially clean code: In yet another tactic to evade sandboxing, malware embed clean code, which poses no danger to the machine on which they are installed. This clean code is launched at the beginning for a certain period. The aim of this is to appear harmless during the analysis by a sandboxing system. After this period, the malware will then activate its malicious code.

4 Extended heuristic approach

(transmission vector + content)

In order not to take into account all the countermeasures set up by malware creators, heuristic methods perform a fully behavioral analysis based not only on the file alone, but on the transportation mode and all the information available about the file, including the file itself.

Heuristics based on the transmission vector: In an e-mail filtering context, the extended heuristic analysis concentrates its analysis on the characteristics of a message. These algorithms can therefore be completely independent of the content and operational function of the attachment. In an e-mail context, a heuristic analysis allows focusing on the sender of the message, headers, the structure of its content, the characteristics of attachments (name, type, etc), among other properties - in other words, a set of significant data that can allow blocking a message that contains a malicious attachment.

Heuristics based on macros: Macros contained in Office documents have once again become a widely used vector for spreading malware and contamination. In general, macros simply download and execute truly malicious code. The creators of such scripts therefore ensure that they will not be identified as malware so long as the actual malware file has not been downloaded. A heuristic analysis on macros allows distinguishing malicious macros from harmless macros and thereby blocks dangerous messages before they can reach the user's mailbox.

Heuristics based on the analysis of PDF files: PDF files are also widely used for spreading malware. Especially in the corporate world, a PDF document may appear perfectly legitimate. Therefore a PDF-oriented heuristic analysis is a series of criteria detected within PDF documents combined with the analysis of the e-mail to identify whether it is part of a fraudulent communication.

Extended heuristics: By definition, heuristic analysis may cover all the elements of the e-mail as well as the file. As such, heuristic rules aim to detect points in common between various attacks and take on malware with an advanced analysis that the cybercriminal would not be able to notice. They allow defusing polymorphic techniques on malware that attempts to bypass signature and sandboxing systems.

5 Heuristics avoid malware countermeasures

As an expert in e-mail for more than 10 years and having protected more than 235 million mailboxes worldwide to date with a unique imprint on the French market, Vade Retro is in a position to combine a targeted heuristic analysis on commonly used file types (office documents containing macros, PDF documents, etc), with another heuristic analysis of the e-mail containing the file.

This double analysis makes it possible to be fully independent of the content of the malware file with very high execution performance due to the fact that the file does not need to be fully loaded for the analysis.

In addition to its heuristic analysis, Vade Retro generally provides a more conventional signature-based anti-malware technology which should be enabled for optimum effectiveness. The signature-based anti-malware will be used after the heuristic analysis in order to optimize the use of system resources.

About de Vade Secure

Vade Secure is the global leader on anti phishing, spear phishing, malware and ransomware filtering. Language independent, the filter analyzes globally all incoming emails (links, attached files, content…) to detect all threats in zero-day, even the most targeted attacks. After elimination all threats, we eliminate the nuisance of low priority emails with the Graymail Management. Ads, social networks notification and newsletters are automatically sent to the graymail folder while the Safe Unsubscribe button eliminate them forever.

Protecting more than 400 million of mailboxes in 76 countries, our solutions are used by major ISPs, OEM and Enterprises worldwide. Vade Secure is implanted in 5 countries (USA, Canada, France, Hong Kong and Japan) to offer a 24/7 support.

Vade Secure
180 Sansome Street, Floor 2, San Francisco, 94104, CA