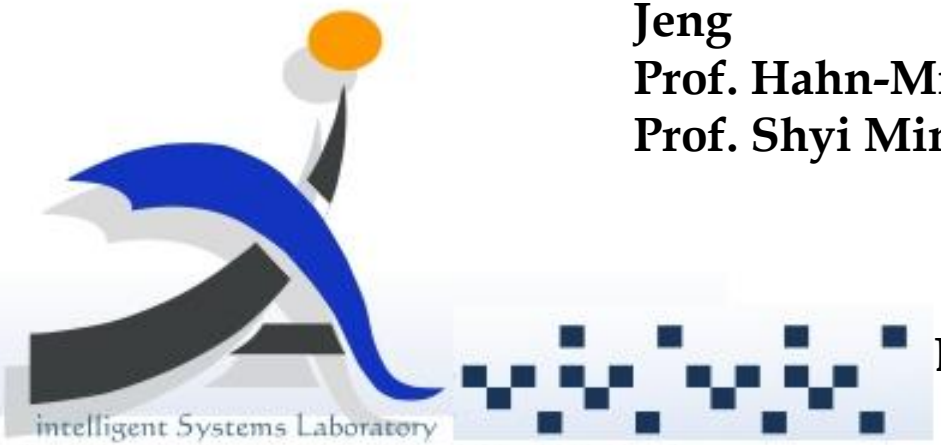
Cloud-Based anti-Malware Solutions

Speaker：Ismail AL-Taharwa
Advisors: Prof. Albert Jeng
Prof. Hahn-Ming Lee
Prof. Shyi Ming Chen



台灣科大智慧型系統實驗室

Date: Mar. 24th 2011

Outline

- Introduction
  - Malware definition
  - Cloud computing
  - Motivations
  - Deep Insight
  - Objectives
- Industry views
- Academic views
- Recommendation
- Proposed solution
- Conclusion

Malware

- Malware, short for malicious software which is written to figure-out and exploit vulnerabilities of computer systems
- Malware is not restricted to viruses only. Instead it includes Worms, Trojans, Spywares, Adware, Bots, Rootkits, Phishing, Spam, and other exploits
- Attackers now are more attracted in exploiting high protected systems rather than weak secured systems

Cloud Computing (1/2)

- Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources
- Those resources can be rapidly provisioned and released with minimal management effort or service provider interaction.
- Cloud model promotes availability and is composed of five essential **characteristics,** three **service models**, and four **deployment models**.

Cloud Computing (2/2)

- **Essential characteristics:** 1. On-demand self-service, 2. Broad network access, 3. Resource pooling, 4. Rapid elasticity, and 5. Measured Service
- **Service models:** 1. Cloud Software as a Service (SaaS), 2. Cloud Platform as a Service (PaaS), and 3. Cloud Infrastructure as a Service (IaaS).
- **Deployment models:** 1. Private cloud, 2. Community cloud, 3. Public cloud, and 4. Hybrid cloud

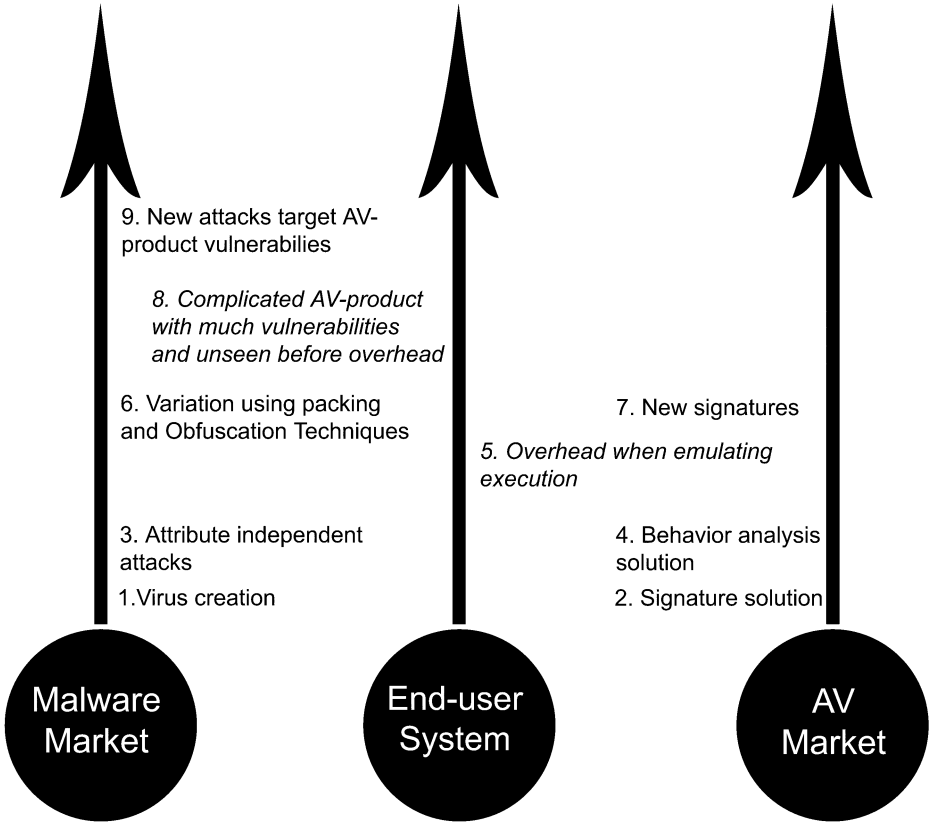
Motivations (1/3)

- In 2008, Symantec created over 1.6 million new signatures in addition to the existing six hundred thousand signatures in 2007
- Host-based AV began to cause problems and resource burdens to the end-user host
- Malware industry has evolved from curious hackers to profit motivated attackers
- Panda Security has detected more samples in year 2008 than the previous 17 years all together

Motivation (2/3)

- A cyclic negative phenomenon arise when host-based solution attempt to countermeasure new attacks by adopting more complex solution
- The more complexity the AV has, the more vulnerable it becomes

Motivation (3/3)



Deep insight

Host-based AV

- End-user's heavy resources consumption
- Big overhead on the network performance
- Increasing complexity
- Reducing end-user's productivity
- Missing many kinds of new attacks, and making FP

Malware industry

- Online malware generators
- Toolkits that facilitate generation of new variants given a malware instance
- Attacker's exploiting tools created to protect programs and developer's copyrights
- Availability of many organizations that train novice attackers

Objective

- Take the advantages of cloud computing environments in detecting malwares
  - detection of zero-day attacks
  - Minimization of vulnerability window
  - Preservation of end-users privacy
  - Protection of any potential attack that may infect end-user machines
- Release end-users systems from increasing burdens of host-based AV solutions
  - Complexity of host-based AV solutions
  - Overload on the end-user systems

Industry views

- Two representative AV-Product providers

Corporation name	Products	Underlying Technologies
Trend Micro	1. InterScan Messaging Security Virtual Appliance 2. Trend Micro Smart Protection Network 3. Trend Micro OfficeScan 10 with File Reputation	1. Web and file Reputation service, 2. Machine learning techniques 3. Autonomous Malware Signature Discovery System (AMSDS)
Panda Security	1. Malware Radar	1. Expert systems 2. Rule-based policies 3. Machine learning techniques

Trend Micro

- Developed Smart Network solution to collect more information about different clients' behavior and reaction to the suspicious files
- Their solution is primarily based on the File Reputation (FR) strategy
- Provided many products where each one is designed to handle different leakage point
- Come out with a novel malicious file exchange mechanism (AMSDS)

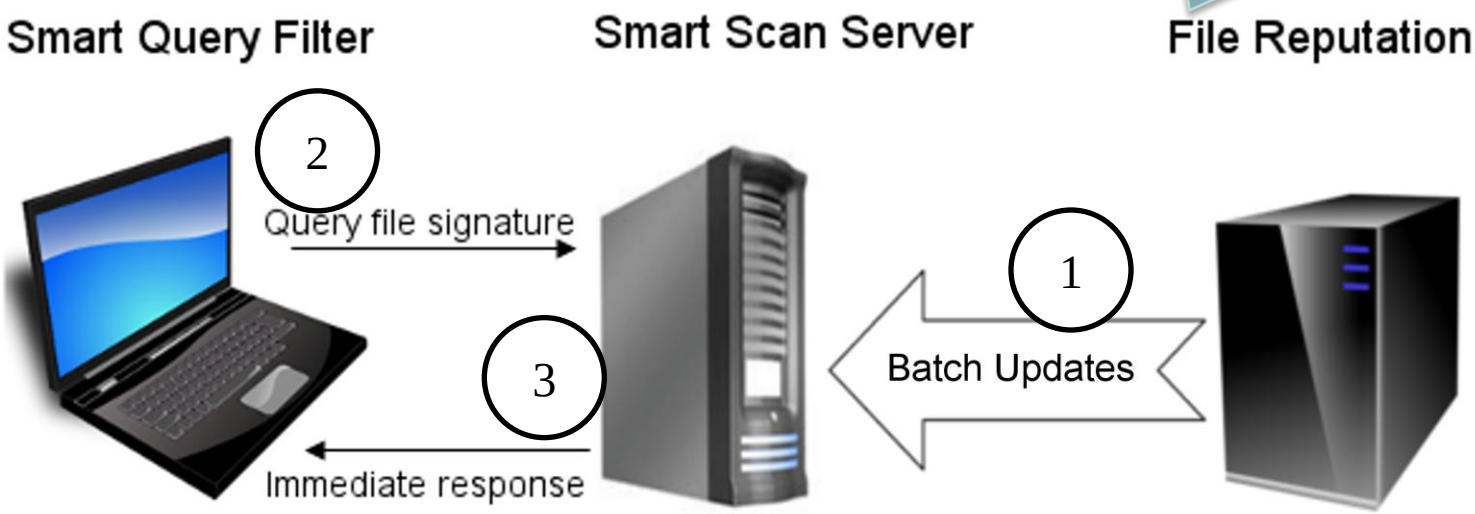
Trend Micro Architecture

This architecture consists of two main parts

1. **File reputation technique**
2. **Communication patterns (AMSDS)**

Leverages the anonymous software usage patterns of millions of Trend Micro users to automatically identify new threats

❖ *Does not query the cloud service for every single file*



❖ *No need to transfer suspicious file over the cloud*

Panda Security

- They invest their earlier solution, that allow to transmit suspicious files to be emulated in the Panda labs
- Integrated their Collective Intelligence (CI) technique to invest the results of their labs in building trusted Database and use cloud services to deploy these information
- Their solution is based on a set of AI techniques and tools to understand the complete behavior of suspicious files, by integrating it with result of interactions generated by other client machines

Effective comparison

✕✕✕ means perfect performance, ✕✕ means good performance, and ✕ means OK performance

Issue	Trend Micro	Panda Security
Storage and memory optimization in the end-user machines'	✕✕✕	✕✕
End-user network bandwidth consumption	✕✕✕	✕✕
Bandwidth consumption on the corporation front	✕✕	✕✕
Processing time	✕	✕✕✕
Data Privacy preservation	✕✕✕	✕
Flexible to end-users security preferences	✕✕✕	✕✕
Forensic tracing support	✕✕✕	✕✕✕
Retrospective detection support	✕✕	✕✕✕

Important characteristics (1/2)

- Two main issues
  - Preservation of data privacy
  - Complexity of processing time
- SplitScreen is an interesting technique proposed by researchers in CMU
  - To minimize signature file size, we will make a trade-off between accuracy and processing time
  - However, this solution provides better processing resource optimization than before
  - Also, it preserve data privacy

Important characteristics (2/2)

- Panda Security technique ensures faster processing time, while they didn't declare any technique to support data privacy
- They use set of rule-based policies to satisfy different preferences of their clients, but they didn't specify if any of them to preserve end-users privacy

Academic views

- We studied newest works in the area of cloud-based anti-malware solutions, those are:
- *CloudAV: N-Version Antivirus in the Network Cloud*
- *A framework for behavior-based malware analysis in the cloud*
- *Retrospective Detection of Malware Attacks by Cloud Computing*

Cloud AV

- This solution exploit cloud capabilities to run multiple detection engines in the cloud, that are able to scan suspicious files in parallel
- Detection coverage increase as the number of the detection engines increased
- Many new industrial approaches support the same mechanism
  - Hitman Pro from SurfRight
  - Virus Total

Behaviour-based malware analysis

- Leverage fast-provisioning cloud capability to allow end-users delegate security labs in the cloud the execution of their suspicious programs
- Program executed in the cloud in the behave of its original end-user
- Cloud combines the results of behaviour analysis from different end-users to get comprehensive understanding of considered file behaviour in real environments instead of limited single synthetic environment

Retrospective detection

- A model to clean-up infected machines from already detected malwares
- This model monitor Portable Format (PE) files written/created operations logs to figure-out malware
- Two types of indexing
  - File indexing: specify which operation related to which machine
  - File relation indexing: novel technique to detect attacks based on persistent programs attributes instead of volatile file attributes
- It can overcome the problem of PE packing and obfuscation

Comparison in terms of cons and pros

Technique	Pros	Cons
CloudAV	1. Multiple heterogeneous detection engine 2. Tuneable parallel processing 3. Enhanced forensic capabilities 4. Support retrospective detection	1. Works for private networks 2. Exposure of enterprise data 3. lack adaptive integration among the different detection engines
Behaviour-based analysis	1. Resembles the exact execution behaviour of the desired system 2. Leverage tracing of all possible execution paths 3. protect user machines' from potential attacks	1. Compromise users sensitive data 2. No guarantee that end-users will execute such suspicious data 3. Memory and CPU overhead are proportional to the number of system calls
Retrospective detection	1. Reliable detection performance 2. Overcome some kinds of evasion techniques (e.g., obfuscation, and packing)	1. Need long time to complete file relation indexing 2. Compromises users' privacy

Recommendation

- any cloud-based solution should meet the following criteria to qualify itself as an applicable solution:
  - **Preserve end-users privacy**
  - **Minimize the overhead on the end-users systems**
  - **Be applicable for both enterprises and stand-alone end-users systems**
  - **Support different levels of protection according to the end-users preferences**
  - **Provide more than a single detection engine in the cloud**
  - **Provide organizations different levels of user privacy to monitor their clients according to their application type**

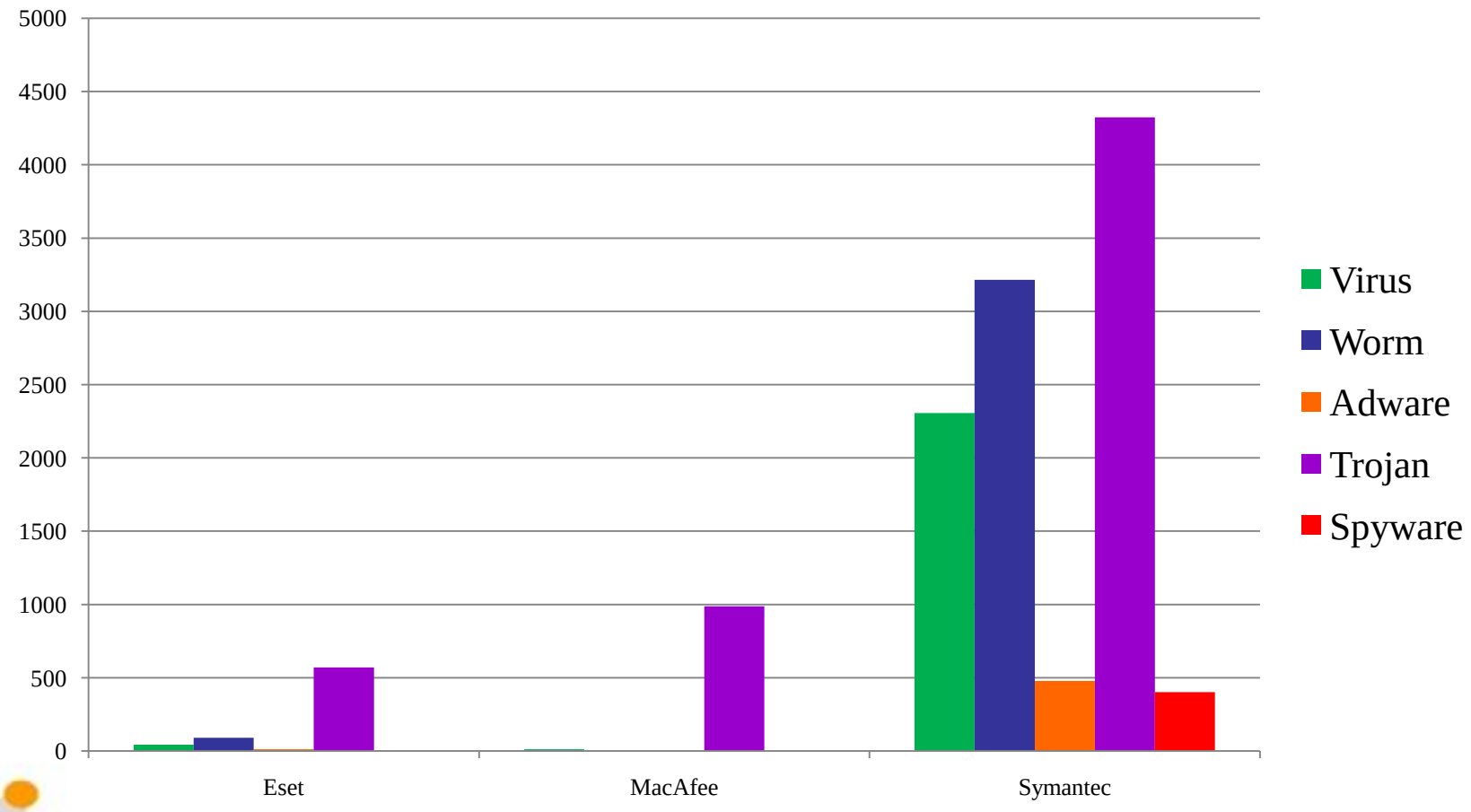
Proposed solution

- Our model aim to exploit different detection capabilities of AV products with minimum overhead
- Using cloud to coordinate and integrate two models of file exchange
  - Depending on the vulnerabilities point in end-user machine
  - Selection of the most appropriate detection engine in the cloud
- Our motivation is the difference in the detection capabilities of distinct AV products

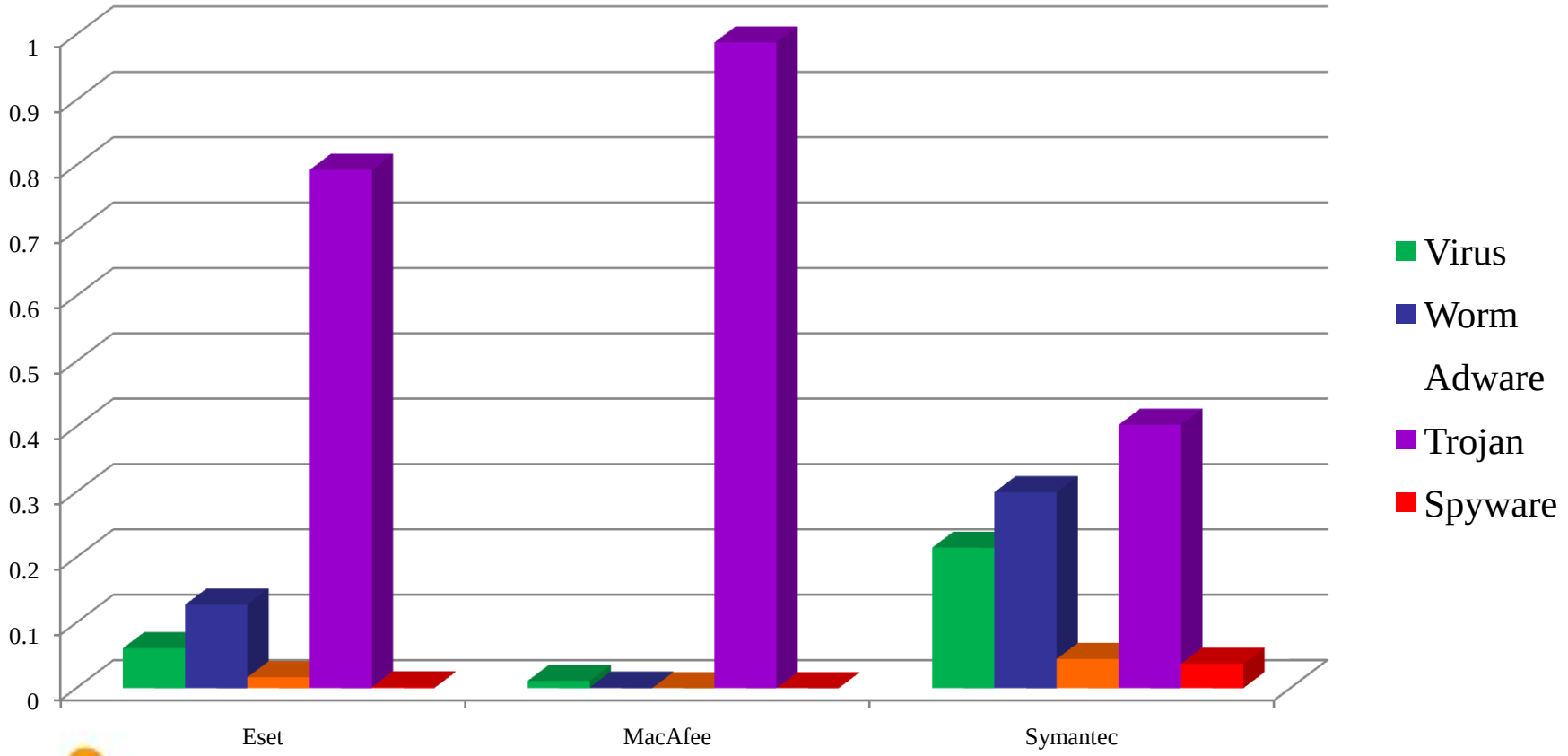
Theoretical evidence (1/3)

- Comparison among the detection rates of some AV products based on their threat databases
- **MacAfee: latest 1000 threats**
- **Symantec: A-Z threats**
- **Eset: detected threats along one year ago**

Theoretical evidence (2/3)



Theoretical evidence (3/3)



Conclusion (1/2)

- Cloud computing is proofed to be the definitive countermeasure for malware prevalence
- Both industry and academic people realized the importance of cloud computing to handle malware detection problem. However, there is a shortage in the academic research in this area comparing with industry works
- Industry people tend to hide their underlying technologies and shortcoming of their products from customers. Thus, we tried to explore those techniques and measure their performance based on the best available information

Conclusion (2/2)

- Good practices from both industry and academic perspectives have been emphasized
- A new adaptive efficient solution has been proposed
- Until now we are collecting data and setting up environment to operate our proposed solution