情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

Using a Database in the Cloud for the Static Analysis of Malware

October 17, 2012

Hiroki Hada
Institute of Information Security

Malware analysis techniques

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ Dynamic Analysis and Static Analysis

- Static Analysis is harder and takes more time than Dynamic Analysis
- Static Analysis is more exact than Dynamic Analysis
- Performing static analysis when dynamic analysis is not sufficient

	Dynamic Analysis	Static Analysis
Method	Execute and trace actions	Read assembly codes
Accuracy	Not accurate	(Theoretically) perfect
Time	several minutes (or several hours)	More than 1 week
Performed by	Automatically	Manually

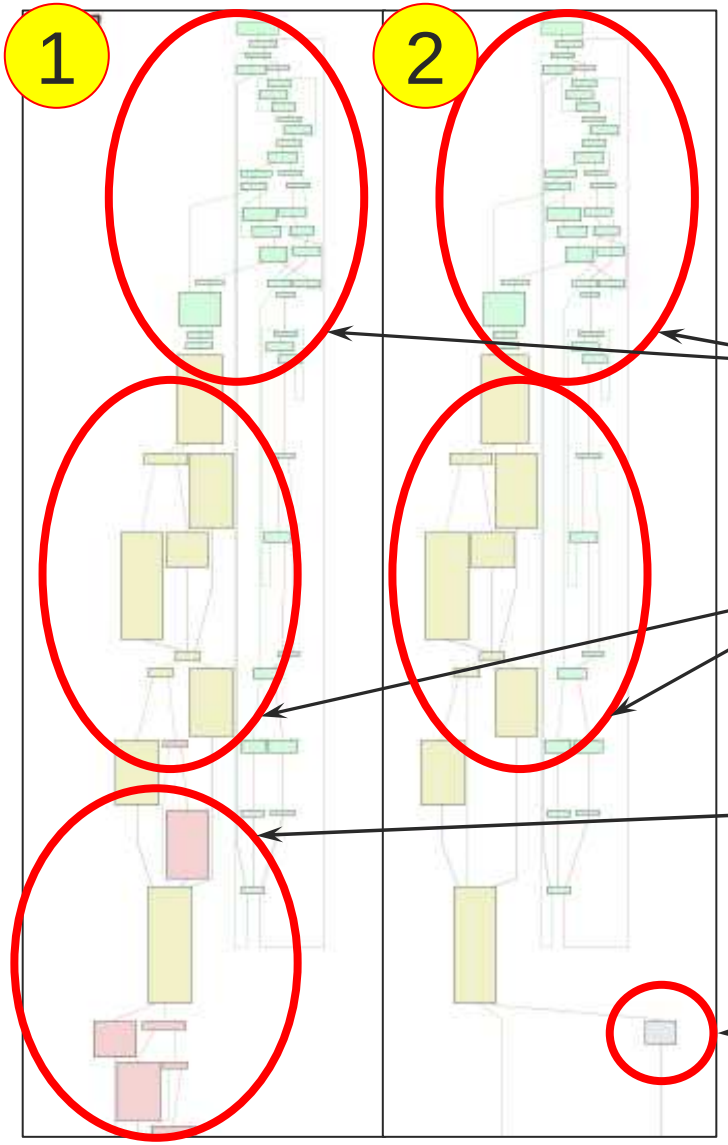
- If necessary, a researcher must do static analysis with too much time.

Our proposal

- ♣ Extracting the difference of two malware programs
  - To reduce the cost of Static Analysis
  - We use the software named "BinDiff" which is developed by H. Flake
- ♣ BinDiff
  - A program is divided into some functions (This is called Call Graph)
  - A function is divided into some basic blocks (This is called Control Flow Graph)
  - Compare Call Graph and Control Flow Graph between two malware programs

Demonstration of BinDiff

♣ Difference between two control flow graphs



Perfectly the same
(Green)

Partly the same
(Yellow)

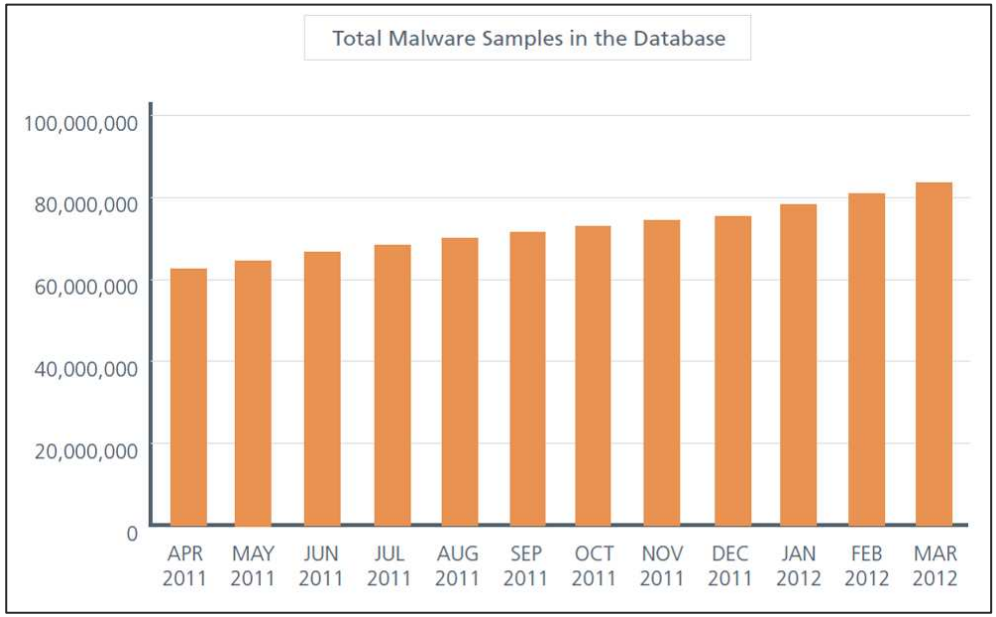
Exsists in only No.1
(Red)

Exsists in only No.2
(Blue)

latest threats of malware

♣ Massive malware programs have been discovered

- McAfee has detected over 80 million malware so far.
  (i.e. one per 1.5 sec)



McAfee Threats Report: First Quarter 2012

- The techniques of generating malware variants.
  (e.g. metamorphic, polymorphic, frequency maintenance)

Recent tendency of malware damage

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

- ♣ Enhanced technique of infection
  - Our social infrastructure systems are under targeted-attack
- ♣ The incident of Mitsubishi Heavy Industries (MHI)
  - In September 2011, 83 servers and PCs have been infected
  - They were infected by famous malware such as Gumblar and SpyEye
  - Forensic experts have researched and concluded that there was no leaks of important data

November 18, 2011

Mitsubishi Heavy Industries, Ltd.

Bulletin Board Notice re Current Status of Investigation on Virus Infections (4)

MHI has investigated possible data leakage from the computers and servers that were suspected of being infected by a new type of virus. MHI has completed a thorough investigation into the matter involving nuclear power-related data, and has concluded that the incident led to no leaks of nuclear power-related data requiring protection.

The company will continue to investigate into the incident relating to its ... ...gations under way by the police authorities.

Considering the importance of its products and technologies, the company ... ...ensure a high level of information security. Taking special notice of the latest incident, the company will pursue ... ...information security measures.

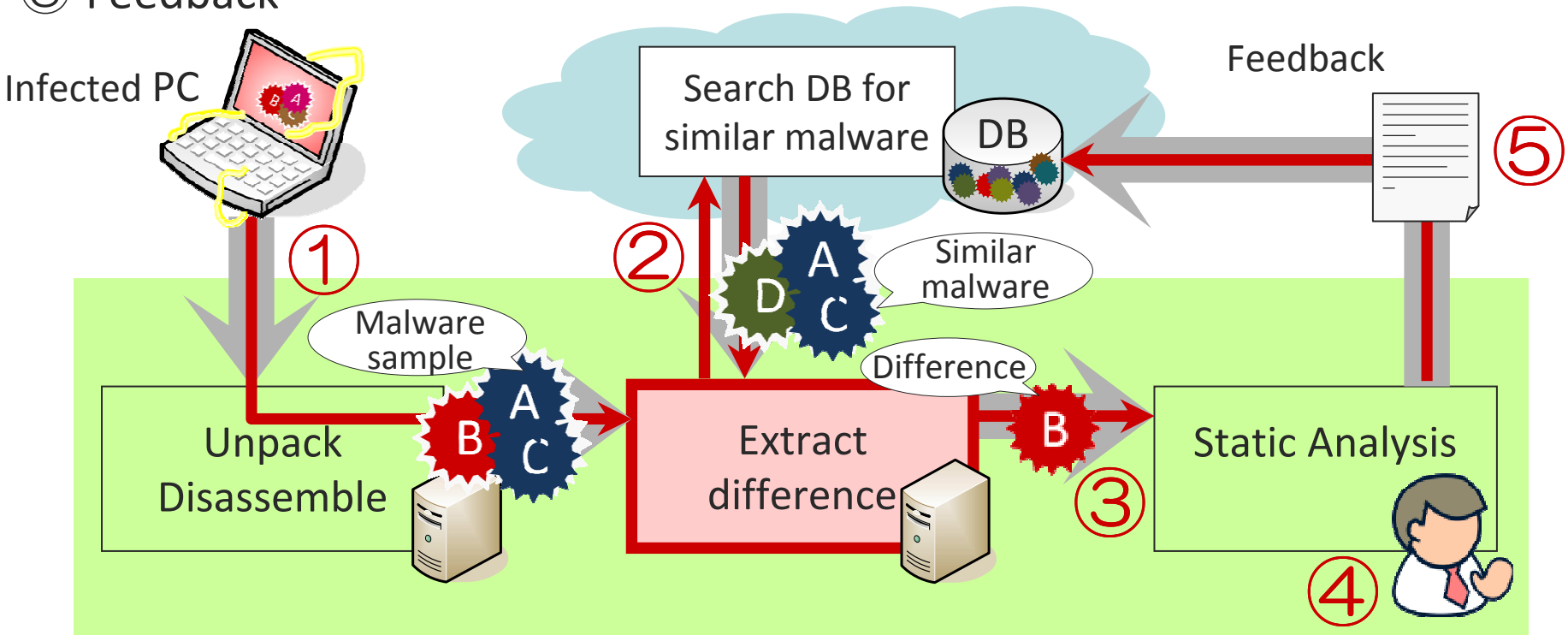
Forensic analysis after the incident is important

MHI News Release, Nov 18, 2011

Proposed architecture

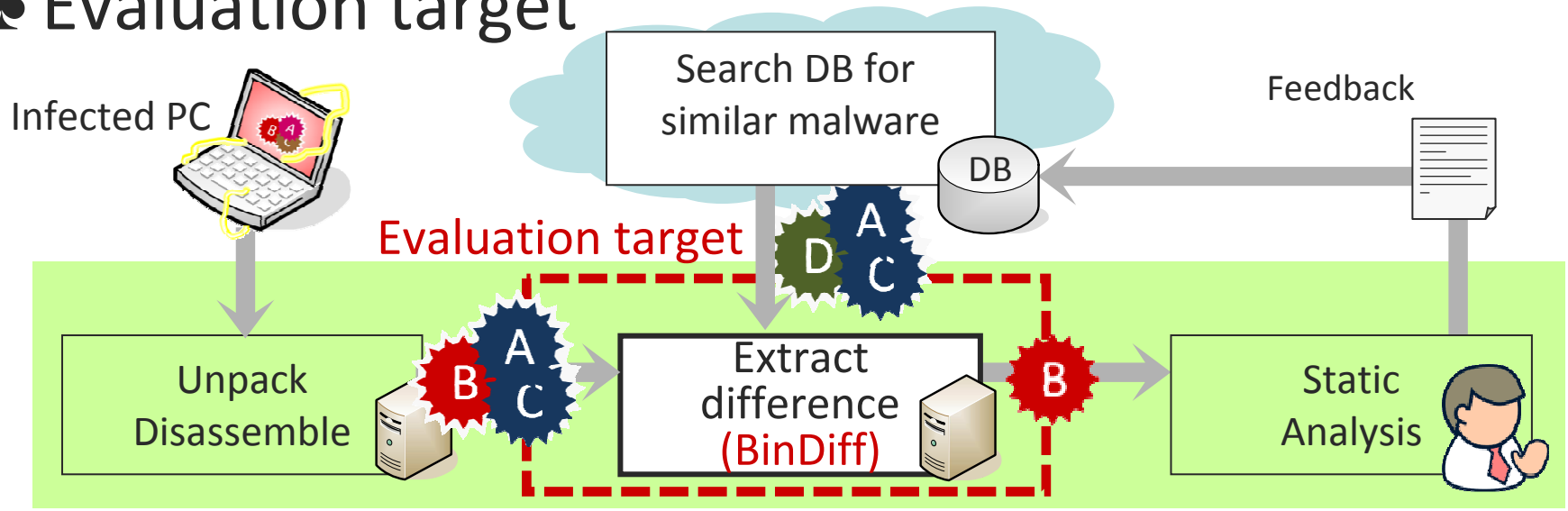
情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

① Obtaining malware sample and extracting assembly codes
② Searching malware which is similar to malware sample
③ Extracting difference between two malware programs. This difference is the part to analyze
④ Static Analysis (in manually)
⑤ Feedback



Evaluation

♣ Evaluation target



♣ Evaluation measure

- How many functions can we remove ?
- The aim of this architecture is to help with the manual analysis.

Result 1

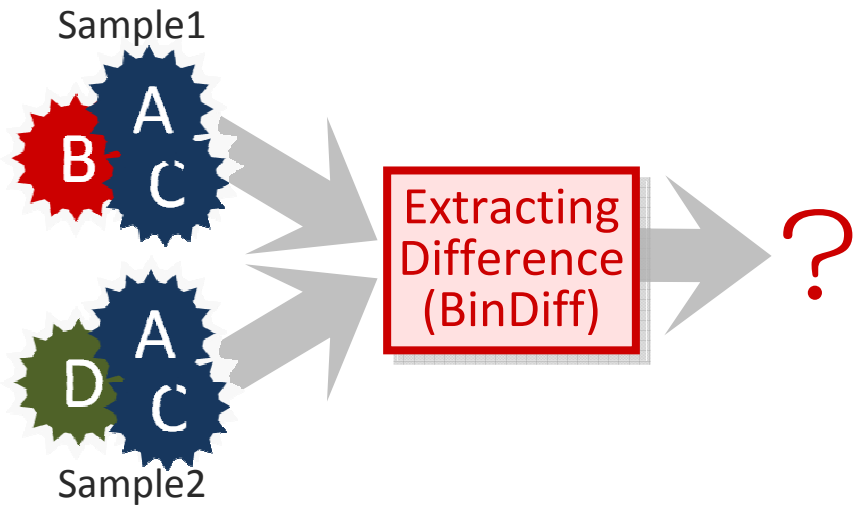
情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ “SpyEye” malware

- which leaks passwords and credit card information

Sample	Number of Function	MD5 Hash
Sample1	523	9D2A48BE1A553984A4FDA1A88ED4F8EE
Sample2	139	D64CA15261C53279A7288616B3CB1A92

- Compare two malware programs and extract the difference



- Result of comparison

Function	Number of function
Common functions in sample1 and 2	53 (53/523=10.1%)
Only sample1	470
Only sample2	58

Static analysis of this function can be done effeciently

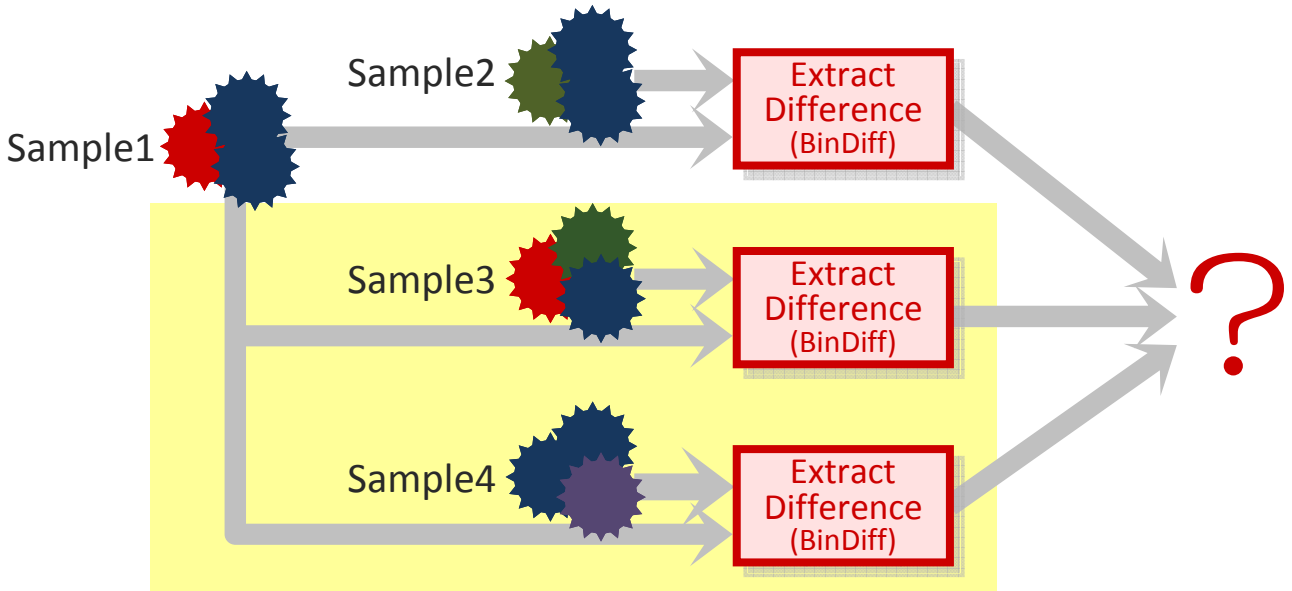
Result 2

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ Several “SpyEye” malware

- Add malware sample3 and 4 to database

Sample	Number of function	MD5 Hash	Role
Sample1	523	9D2A48BE1A553984A4FDA1A88ED4F8EE	analysis target
Sample2	139	D64CA15261C53279A7288616B3CB1A92	in the database
Sample3	609	DF04C2CD2B5F7E471CB0435FDB9B3014	in the database
Sample4	218	42DACFBE2E5AF0C43D17356CA76F0271	in the database



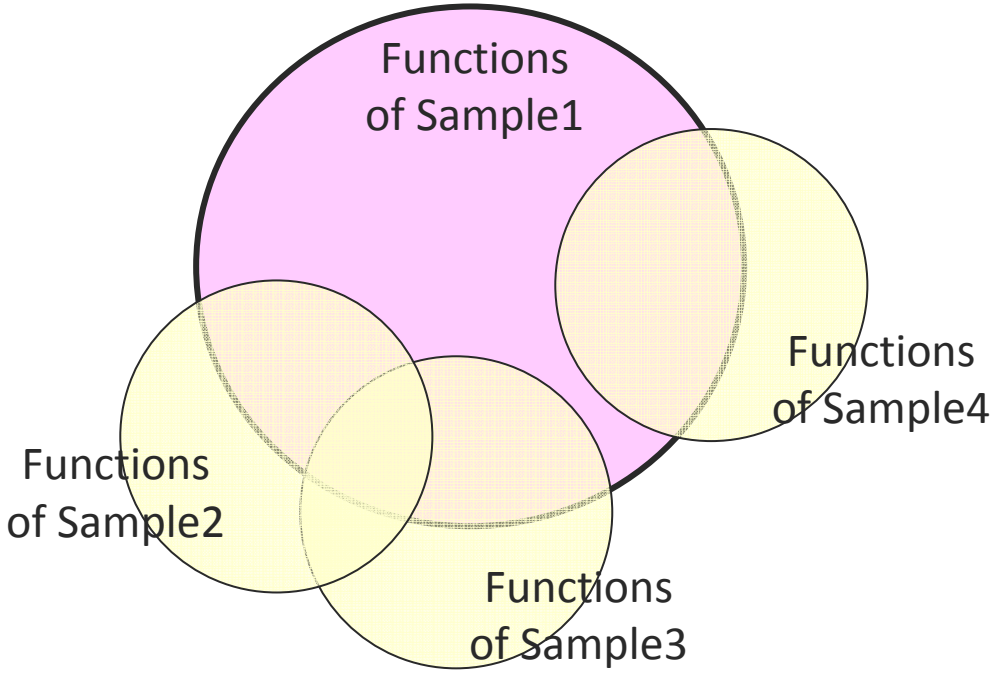
Result 2

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

♣ Result of comparison

	Sample2	Sample3	Sample4	Sample4, 5, 6
Common functions in sample3	53	78	85	135 (135/523=25.8%)

- Using multiple malware programs, number of common functions are improved



In summary

情報セキュリティ大学院大学
INSTITUTE of INFORMATION SECURITY

- We proposed new architecture which makes static analysis more efficient
- One of the key components in this system is a similarity analysis function which compares disassembly code of the target malware with already known malware in the database
- We think cloud system is useful to construct the malware database to share the analysis result all over the world