

Hacking Tool Identification in Penetration Testing

Da-Yu Kao^a, Yun-Ya Chen^a, Fuching Tsai^b

^aDepartment of Information Management, Central Police University, Taiwan

^bDepartment of Criminal Investigation, Central Police University, Taiwan

^bCorresponding Author: fcttai@mail.cpu.edu.tw

Abstract—With the rapid development of information technology, cybercrimes are getting increasingly rampant. The popularity of various toolkits facilitates the cyber-attackers to intrude on online systems and access the vast streams of sensitive data. Prevention is better than cure. Penetration testing (PT) not only defends the illegal attacks but also identifies potential vulnerabilities. It is capable of distinguishing network traffic on the massive amount of collected data from the following three phases: data reduction, data analysis, and data classification. This paper analyzes the PT network traffic and observes their tool marks.

Keywords—Forensic Analysis, Penetration Test, Cybercrime Investigation, Cyber-attack, Tool Marks

I. INTRODUCTION

Due to the convenience of the Internet, people can socialize with each other, do business, and learn something new. Many problems of information security are arising one after another. Some systems may face data breach issues and that ultimately challenges the cybercrime investigators. Cybercrime is defined as crimes directed against a computer, where the machine itself holds the sherds of evidence, or where the equipment is used to commit the offense [10]. Investigators should be prepared with the knowledge, skills, and abilities to cope with cybercrime activities.

Everything has some drawbacks, and so does the computer [2]. Illegal activities conducted by cybercriminals are often motivated by financial gain or fun [11]. The cybercrime issue emerged as information technology has become more popular and less expensive than before. When vulnerable data transferred through networks, some criminals looked to access computer systems or exploited it for their illegal purposes [9]. Hacking behaviors are in a row to compromise or abuse their sensitive data. Many laws have been enacted to prevent people from becoming victims of hackers. Nevertheless, that has not stopped criminals from committing hacking crimes. Law enforcement agencies (LEAs) should do their utmost to look for feasible approaches, find out evidence, and bring cybercriminals to justice.

Choosing a trusted partner to conduct penetration testing (PT) is itself a sensitive matter. PT simulates the activity conducted by the hacker and points out vulnerabilities [7]. It tests for known and unknown technical vulnerabilities in network systems. On the whole, this process is used to help operations and applications free from any weaknesses, threats, risks that may cause a loss. PT strengthens the analysis model of cybersecurity experts and enriches an analyst's

understanding of an adversary's tactics, techniques, and procedures. It follows a series of phases from the early reconnaissance stages to the exfiltration of data. The idea is to check for potential vulnerabilities to an external hacking attempt by simulating an attack as a malicious attacker. It aims to identify, tackle, analyze, and attack the target [1].

The structure of this paper is organized as follows. Section 3 describes the experiment environment. Information extraction and its analysis are described in Section 3. Section 4 presents the pattern findings and information analysis. Section 5 concludes the paper.

II. EXPERIMENT ENVIRONMENT

This paper focuses on the forensic analysis of network packets from penetration test toolkits and deals with various tool marks. Information collected can be various and complex. Taking raw data as input and presenting the data in a human-readable format help users understand the relational information quickly. These experiments were carried out in the computer classroom of Central Police University on November 9th, 23rd, 30th, and December 7th, 2018. This PT is under a controlled environment to simulate hacking behavior. Packet sniffers refer to the analysis of network traffic. In Table 1, the data of PT toolkits is analyzed from the following elements: date, IP address, cs-method, and sc-status.

TABLE 1. The software in the experiment

NO.	Software	Date	IP Address
1	Zenmap	20181109	163.25.10.107
		20181123	163.25.10.118
		20181207	163.25.10.107
2	Nmap	20181109	163.25.10.101
3	Netsparker	20181109	163.25.10.115
		20181123	
4	Acunetix Web Vulnerability Scanner	20181109	163.25.10.118
		20181130	
5	Arachni	20181109	163.25.10.117
6	Superscan	20181123	163.25.10.114
7	GFI LanGuard	20181109	163.25.10.120
8	Google Hacking	20181109	163.25.10.112
		20181123	
9	Paros	20181109	163.25.10.111
10	DNSrept	20181109	163.25.10.106
11	Netcraft	20181109	163.25.10.113
12	FOCA	20181109	163.25.10.119
		20181130	
13	NetworkView	20181109	163.25.10.102
14	Nessus	20181109	163.25.10.109
15	Vooki	20181123	163.25.10.117

Packet sniffers can capture network traffic, assemble data, and display them in an easy-to-read format. Hackers may use packet sniffers to gain unauthorized access or confidential data through the network. We use Wireshark to sniff packets in this experiment. Wireshark is an open-source and network analysis packet analyzer on Linux and Windows. It captures real-time data and displays it in a human-readable format. It contains detailed data about network protocols, decryption, and packet information. The experiment environment is illustrated below.

A. Hardware

- ASUS computer Inc.

B. Software

1) OS: Windows Ver. 10

2) PT Toolkits in Table 1

3) Analysis toolkits

- Wireshark ver. Win32 2.6.3
- Excel
- i2 Analyst's Notebook 8.5.5

III. INFORMATION EXTRACTION AND ITS ANALYSIS

The universal application of data science is developing rapidly to cope with vast amounts of evidence in a limited time efficiently. Various analytical approaches to data reduction, analysis, and classification are applied in forensic science. Table 2 indicates the following three phases in the experiment: Data Reduction, Data Analysis, and Data Classification.

TABLE 2. Analytical Methodology

Phase	Toolkit	Procedure
I. Data Reduction	Excel	a. Import CSV Files into Excel b. Simplify the Quantity of Information c. Erase Missing Data
II. Data Analysis	I2 Analyst's Notebook	a. Import Excel into I2 b. Create Relational Linkage c. Merge the Tool Entity d. Remove Common Contents
III. Data Classification		a. Multi-Methods b. TRACE Method c. Conventional Methods d. Common Behavior

A. Data Reduction

Data extracted from network traffic may be massive and contain a large amount of redundancy and noise. Data reduction increases storage efficiency and transforms information into a corrected, ordered, and simplified form. Minimized data size is tiny in volume and comparatively original. We can mitigate the data handling costs and will reduce the analysis time [3]. Researchers are capable of filtering useful information for a detailed payload analysis.

1) Import CSV Files into Excel: Some Comma-Separated Values (CSV) files are created from network traffic to provide the tool mark data for Excel analysis. Excel is a spreadsheet program to record and analyze numerical data. It is a collection of columns and rows that forms a table. Alphabetical letters are usually assigned to columns, while numbers are typically attached to rows. The columns presented in Excel show the following fields: Date, Time, s-ip, cs-method, cs-uri-stem, cs-uriquery, s-port, cs-user, c-ip, cs(User-Agent), cs(Referer), sc-status, sc-substatus, sc-win32status, and time-taken. To identify the characteristics, the researchers add the tool name into the first column for better interpretation.

2) Simplify the Quantity of Information: The original data in Excel contains 16 fields. Researchers use Excel filtering function to create a critical data set. With the process of simplifying the vast volume of textual data, the columns are reduced to three fields, including tool, cs-method, and sc-status.

3) Erase Missing Data: Traffic captured and recorded in files is not always complete. While some data is missing, it may be unsuitable for identifying the tool mark. We erase some missing data while processing the data sets.

B. Data Analysis

Data analysis discovers connections and patterns among elements in large quantities of data [8]. I2 Analyst's notebook is a visual analysis tool that helps users turn data into useful information. It can visualize connected networks and display geographic views to discover hidden linkages. It can help users identify criminals. Using I2 Analyst's Notebook to observe the correlation between different columns helps us find some unique characteristics of tools effectively and efficiently.

1) Import Excel into I2 Analyst's Notebook: Researchers define columns, select rows, assign columns, and finally import detailed data into I2 Analyst's Notebook. There are many different fields in log files. Each log contains multiple pieces of data stored as text. Parsing splits data into groups of information [5]. The goal of log parsing is to recognize and group them in a meaningful way.

a) sc-status: The HTTP status code replies to the client. Sc-status can be divided into five classes, including 1XX, 2XX, 3XX, 4XX, and 5XX [4].

b) cs-method: Cs-method means The HTTP method performed by the client, inclusive of GET, POST, HEAD, OPTIONS, PROPFIND, and TRACE [4].

2) Create Relational Linkage: In the process of importing data into I2 Analyst's Notebook, researchers select critical columns to display in the relational linkage. The linkage analysis may provide evidential insights. The linkage analysis of PT tools focuses on sc-status and cs-method.

a) Tools and Status: Figure 1 illustrates a relational coupling on tools, sc-status, and their interactive counts. There are many counts from status 200, 302, and 404.

b) Tools and Methods: Figure 2 demonstrates a relational linkage on tools, cs-method, and their interactive counts. GET and POST connect much more counts than other methods.

3) Merge the Tool Entity: Figure 3 merges the tool entity in Figure 1 and Figure 2. The linkage combination helps us illustrate a complete and comprehensive connection among these three fields.

4) Remove Common Contents: The original version of the combined linkage on tools, sc-status, and cs-method is very complicated. Researchers remove some typical content and simplify their class characteristics in Figure 3. Researchers eliminate the following symptoms to clarify the linkage relationship. That makes their marks much clear.

a) Remove Common Symptoms (200, GET, and POST): Sc-status 200 means that the request has been successfully received and accepted. Cs-method GET indicates that the user transmits to Web Server with a query string. Cs-method POST shows that the user sends to Web Server with a query string. These contents are common characteristics of many tools.

b) Remove Other Symptoms (302 and 404): Sc-status 302 demands users to conduct temporary movement. Sc-status 404 indicates that the requested resources are not found in the server. These two statuses are common to several tools.

C. Data Classification

Data captured from network traffic is enormous and various. Through the process of simplifying and filtering, linkage presented in I2 Analyst's Notebook may still be somewhat complicated. Data classification helps researchers identify specific content quickly and clearly. In Figure 3, researchers exclude some common symptoms and classify them into the following four types: Multi-Methods, TRACE Method, Conventional Methods, and Common Behaviors.

1) Multi-Methods in Figure 3a: Tools such as Zenmap, Nmap, and Netsparker have their unique methods and status. Their characteristics are clear and easy to be detected. The recognizabilities of Zenmap and Netsparker are too noisy and it is easy for defenders to find their activities.

2) TRACE Method in Figure 3b: This category consists of Netsparker, Acunetix Web Vulnerability Scanner, Arachni, and Superscan. The TRACE method performs a loopback test among the servers marked by the specified resource.

3) Conventional Methods in Figure 3c: GFI LanGuard, Google Hacking, Paros, DNSreport, Netcraft, and FOCA are not connected to other unique methods, exclusive of traditional methods such as GET or POST. Through pre-processing, tools in this category are linked to sc-status 206, 301, 304, 403, or 500.

4) Common Behaviors in Figure 3d: NetworkView, Nessus, and Vooki are categorized into common behaviors. These three tools have no other unique correlation to an unconventional approach or status.

IV. PATTERN FINDINGS AND INFORMATION ANALYSIS

A. Pattern Findings in Experiment Results

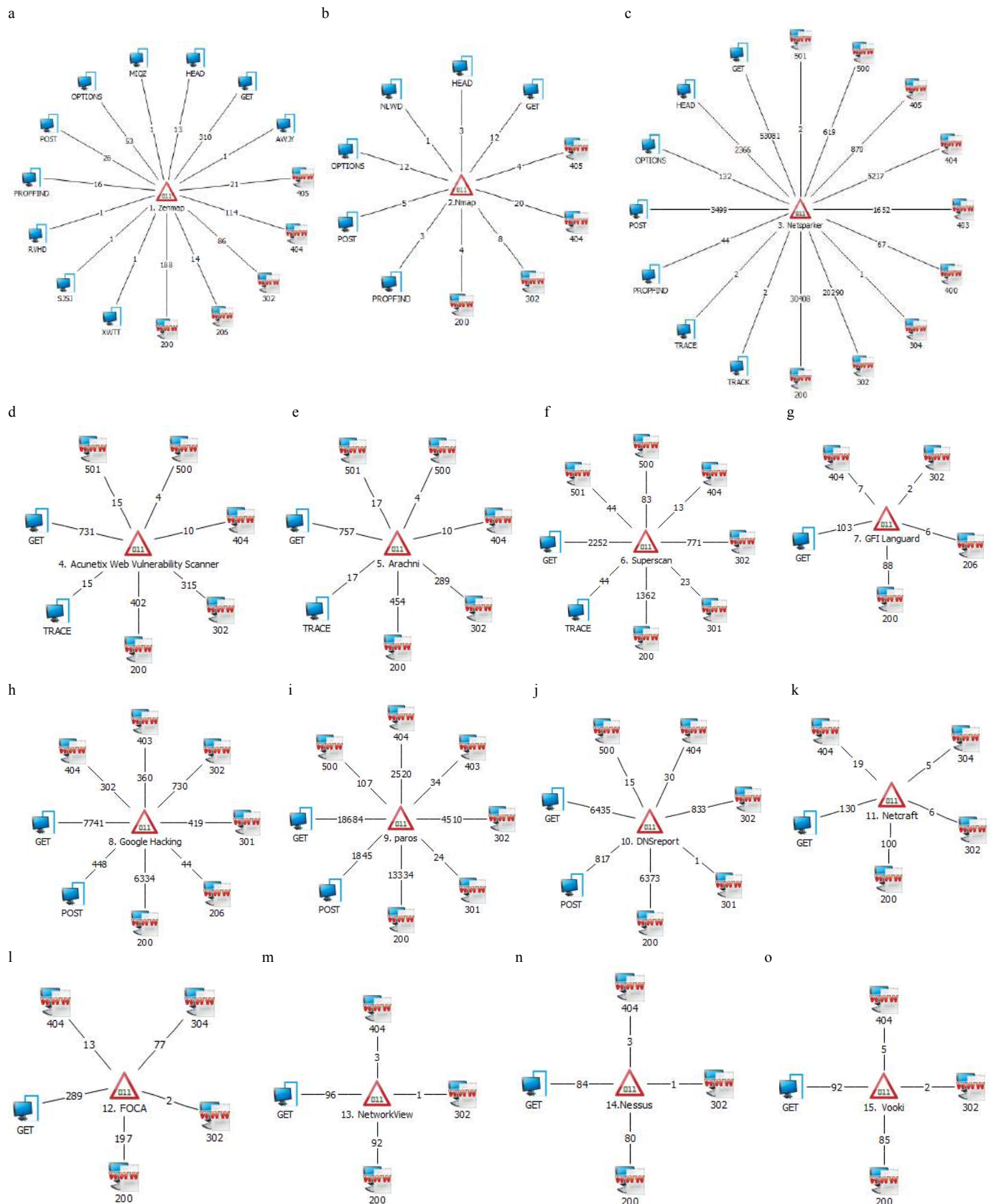
Figure 4 illustrates a separate linkage of different tools and their relational fields. It reduces potentially huge information to a small, understandable set of graphically supported elements. Each presents its connection. The goal is to uncover facts and provide information to assist in the cyber-attack response.

1) Several Defined Methods not Appearing: Hypertext Transfer Protocol (HTTP) defines a set of request methods to provide resources and perform specific operations, including GET, HEAD, POST, PUT, DELETE, CONNECT, OPTIONS, and TRACE. For example, the defined method DELETE does not appear in the packets captured in this experiment. DELETE means delete specific resources. If the method is performed, it may lead to the missing data, further causing the victim to notice the hacking behavior and take **actions** in time.

2) Some Undefined Methods Appear in Zenmap: Data captured from Zenmap indicates the following five undefined methods: XWTT, RVHD, MIQZ, AWJY, and SJSJ. The sc-substatus in these methods is zero. It means that these HTTP requests could not be made because the server could not resolve them.

3) Some Regular Tool Marks: Tools like NetworkView, Nessus, and Vooki are only having related tool marks, such as GET, POST. Those characteristics are so regular that they are not easy to be detected.

4) More Experiments is Necessary: A little data is not that persuasive. When data becomes sufficient, the inference gets more comprehensive. Complete tool marks improve the judgment.



B. Information Analysis for Case Reconstruction

Investigators need techniques for dealing with various criminal scenarios. Determining where hacking tool patterns exist from is a sincere consideration in the context of creating case reconstruction.

Investigating cybercrimes requires a basic skill set and experience to identify where the evidence is and to track back to a possible source.

1) Cross-Referencing

Cross-referencing refers to related information elsewhere in the same material. The intention of putting it all together is to demonstrate how online evidence can identify potential suspects for multiple cybercrimes. It is of utmost importance because it forms a network structure of relations existing between different parts of data.

a) Log Analysis: Log analysis is an essential part of forensics. While we analyze an incident, it is imperative to collect appropriate records from various types of logs.

b) Tool Marks: Tools leave distinctive marks. A tool mark is an impression left by the contact of a software. Tool mark identification determines if it was issued by a particular tool. This identification can establish a link between a tool mark and the software that created it. There are two types of evidence: class or individual characteristics. While individual characteristics narrow down the evidence to a single or particular source, class characteristics place a specific bit of evidence into a group of objects [6]. According to the tool marks, analysts need to know how to identify the relationship between the software and the criminal.

2) Case Reconstruction

In court, the results of a forensic reconstruction can provide additional support or refute a particular chain of events. An investigator will collect the proofs acquired from the digital crime scene and present these in court. The goal of the crime reconstruction is to perform testing in a forensically sound manner so that the results may support or refute a hypothesis regarding a particular sequence of events.

V. CONCLUSIONS

A forensic analyst should always keep up with the latest technologies. LEAs also need to regularly update their knowledge, skills, and abilities for investigating cybercrimes. Various separated digital data seems unspectacular and not that persuasive. If analysts can combine specific clues and tool marks, put it all together, and take advantage of data reduction, analysis, and classification, it would be better to find out their connections. Even when criminals intend to change or tamper single data, the analyst can clarify their relationship through cross-referencing or case reconstruction.

ACKNOWLEDGMENT

This research was partially supported by the Executive Yuan of the Republic of China under the Grants Forward-looking Infrastructure Development Program (Digital Infrastructure-Information Security Project-109).

REFERENCES

- [1] Abawajy, J.H., Mukherjee, S., Thampi, S.M., Ruiz-Martínez, A. (Eds.), *Security in Computing and Communications*, pp. 438-452, 2015.
- [2] Al-Alawi, A. I., "Cybercrimes, Computer Forensics and their Impact in Business Climate: Bahrain Status," *Research Journal of Business Management*, Vol. 8, pp. 139-156, 2014.
- [3] Bayuk, J., "Cyber Forensics: Understanding Information Security Investigations," Springer's Forensic Laboratory Science Series, Vol. 59, Humana Press, pp.59-101, 2010.
- [4] Casey, E., *Handbook of Digital Forensics and Investigation*, Elsevier Inc., pp. 21-208, 2010.
- [5] Cuesta, H. and Kumar, S., *Practical Data Analysis (2nd Edition)*, Packt Publishing, pp. 1-25, 2016.
- [6] Douglas, P.L., *Forensics For Dummies (2nd Edition)*, pp. 41-42, 2019.
- [7] E-Council, *Ethical Hacking and Countermeasures – Attack Phases (2nd Edition)*, E-Council Press, pp. 25-162, 2016.
- [8] Hoscheidt, M. M. and Eichner, E. F., "Legal and Political Measures Address Cybercrime," *UFRGS Model United Nations*, Vol. 2, pp. 445-477, 2014.
- [9] Khadam, N., "Insight to Cybercrime," *Taipei University Law Review*, Vol. 29, No. 1, pp. 55-80, 2012.
- [10] Waschke, M., *Personal Cybersecurity: How to Avoid and Recover from Cybercrime*, Apress Business, pp. 221-229, 2017.
- [11] Yadav, T. and Rao, A. M., "Technical Aspects of Cyber Kill Chain," The Third International Symposium on Security in Computing and Communications (SSCC'15) Proceedings, Kochi, India, pp. 438-450, 2015.



Da-Yu Kao is an Associate Professor at the Department of Information Management, College of Police Science and Technology, Central Police University, Taiwan. He is responsible for various recruitment efforts and training programs for Taiwan civil servants, police officers, or ICT technicians. He has an extensive background in law enforcement and a keen interest in information security, ICT governance, technology-based investigation, cyber forensics, human resource development, and public sector globalization. He was a detective and forensic police officer at Taiwan's Criminal Investigation Bureau (under the National Police Administration). With a Master's degree in Information Management and a Ph.D. degree in Crime Prevention and Correction, he had led several investigations in cooperation with police agencies from other countries for the past 20 years. He can be reached at camel@mail.cpu.edu.tw.



Yun-Ya Chen is a student at the Department of Information Management, College of Police Science and Technology, Central Police University, Taiwan.



Fu-Ching Tsai is an Assistant Professor at the Department of Criminal Investigation, Central Police University, Taiwan. He received the M.S. and Ph.D. degrees in Institute of Information management from National Cheng Kung University, Taiwan, in 2005 and 2012, respectively. His research interests include data mining, text mining, digital forensics, social network analysis, and cyber criminology.