WHAP: Web-Hacking Profiling Using Case-Based Reasoning

Mee Lan Han*, Hee Chan Han*, Ah Reum Kang†, Byung Il Kwak*, Aziz Mohaisen† and Huy Kang Kim*
*Korea University, South Korea, †SUNY Buffalo, USA
Emails: {blosst, huer2783, kwacka12, cenda}@korea.ac.kr, Emails: {ahreumka, mohaisen}@buffalo.edu

***Abstract*—As in the real world's criminal investigation, cyber criminal profiling is important to attribute cyber attacks. Every cyber crime committed by the same hacker or hacking group has unique characteristics such as attack purpose, attack methods, and target's profile. Therefore, a complete analysis of the hacker's activities can give investigators hard evidence to attribute attacks and unveil criminals. In this paper, we implemented WHAP, a profiling system that uses Case-Based Reasoning (CBR). We verified WHAP's usefulness by analyzing large scale of web defacement cases including North Korean hacker's attacks against South Korea, and unveiling a relationship between those attacks and another set of attacks against Sony Pictures Entertainment.**

I. Introduction

In the recent five years, South Korea suffered from various nation-wide cyber attacks that were presumably associated with the North Korean cyber army, known as the Dark Seoul Group, as chronologized in Figure1 [1]. The severest attack happened on March 20, 2013, and as a result, South Korea's major broadcasting companies and top four banks were hacked and taken down. Interestingly, the malware used in and found in the aftermath of these attacks has several common strings and routines that can be an important clue to their origin and identity. North Korean cyber army's attacks become the utmost threat. The same type of attacks is not limited to South Korea but included targets in United States initiated by the North Korean malicious actors, including the notorious Sony Pictures Entertainment hacking case (2014).

Those attacks belong broadly to the class of Advanced Persistent Threats (APTs), which have become a significant source of the threat. These attacks are sophisticated in nature, thus it is hard to capture hard evidence needed to detect the malicious actors behind them. However, we speculate that 1) if one can find the most similar hacking cases from historical data to a case of interest, contextualizing and annotating such APT attacks would be made easier, and 2) if one has a well-defined profile vector that describes a hacking case, he can also efficiently search similar cases. To this end, we developed WHAP, a hacking profiling system. WHAP consists of a hacking case database, a hacking case crawling system, a similarity measure, and a case-based reasoning system, and is used to facilitate cyber crime investigation.

II. WHAP: System Design and Implementation

The overall architectural view of the system is illustrated in Figure2. In the following, we describe the system using an instance of implementation. First, we built a large hacking case database which includes 212,093 web-hacking cases that happened during the past 15 years. We collected the hacking cases from Zone-H.org [2]. Second, we designed the case vector according to the 5W1H formula (Who, When, Where, What, How, and Why) to depict the attack case intuitively and informatively. Among many features found in the attack cases, and after a thorough review by trained security analysts, we chose the most significant features with the frequently found in web-hacking cases as highlighted in Table I.

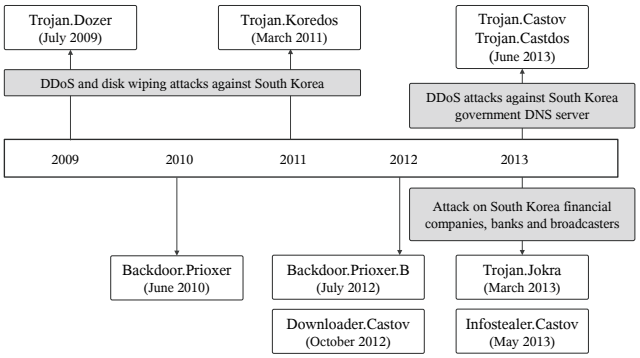

Fig. 1. Recent attacks by North Korean cyber army

TABLE I. Case vector design for WHAP

Case type	Selected features
Web-hacking	attack date, notify, target IP address, domain name, OS, web server, domain type, thanks to, hacker group, e-mail address, social media, messenger, country code, encoding, font, sound, link site, hacker's message

Third, we developed a measure to estimate the similarity between case vectors. Table II shows examples of how to calculate the similarity score between the non-numeric features.

The similarity score between two cases is the similarity sum among all of their features (weights are set evenly, then updated based on features' importance):

$$\text{Similarity Score } (S) = \sum_{i=1}^{n} (Feature_i \times Weight_i)$$

We defined the extent of similarity between cases C_i and C_j as 0, 1 and a numeral value from 0 to 1. 0 means that case C_i and case C_j are unrelated while 1 means that case C_i and case C_j are identical. S $(0 < S < 1)$ is the similarity between case C_i and case C_j. If S is closer to 1, the case C_i is more similar to case C_j [3]. Based on this similarity measure, we applied a Case-Based Reasoning algorithm [4] to search the most similar attack cases utilizing the aforementioned case database.

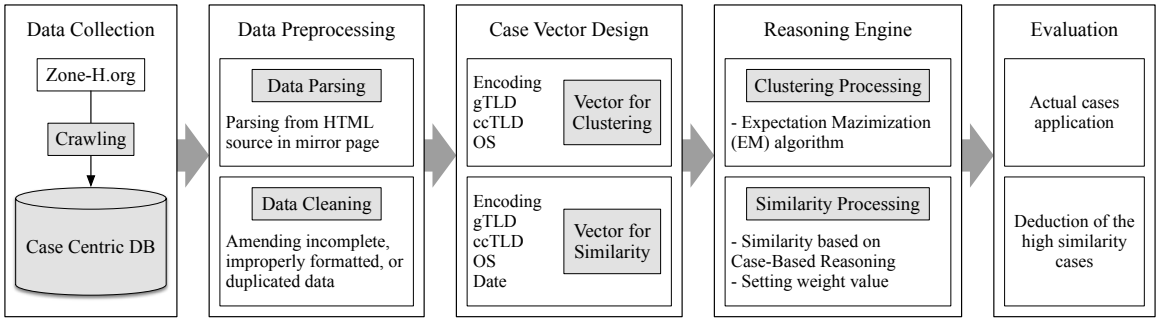

Fig. 2. Overview of WHAP, the hacking profiling system

TABLE II. Similarity score among attack IP address, domain name, date with condition of IP address and attack domain

Condition of IP address	S
if the same (e.g. 143.248.1.1 & 143.248.1.1)	1
if class A, B and C are matched (e.g. 143.248.1.1 & 143.248.1.8)	0.75
if class A and B are matched (e.g. 143.248.1.1 & 143.248.3.4)	0.5
only class A is matched (e.g. 143.248.1.1 & 143.13.2.8)	0.25
no common class (e.g. 143.248.1.1 & 163.13.2.5)	0
Condition of attack domain	S
an identical domain	1
service name is matched, one of the gTLD and ccTLD is matched	0.8
gTLD and ccTLD are matched	0.3
service name is matched	0.1
ccTLD is matched	0.1
gTLD is matched	0.1
non-identical domain	0

III. Analysis of the Dark Seoul Attacks

As a result of the Dark Seoul cyber-attack on March 20, 2013, the groupware homepage of LGU$^+$ and Korean Broadcasting System's homepage were defaced. The attacker left a unique image and messages on the defaced websites. The significant characteristics found in the websites are the Calaveras (skull) image in LGU$^+$ website and "HASTATI" strings in KBS website. All clues in the two cases have common characteristics: 1) the image is frequently found on European sites, 2) "HASTATI" is a military word related to Roman Empire army, and 3) the message left is written in the same encoding system as European languages based on the Latin language system, where most Korean web pages use another encoding system.

Fig. 3. LGU$^+$ groupware case (left) and KBS case (right)

We calculated all similarities between the past cases. Among the 212,093 collected cases of web-hackings, the similarity scores of the randomly selected two cases typically are distributed around 0.3 points. We took the distribution of the similarity score using the central limit theorem, which is the process repeated ten thousand times after calculating the mean value by summarizing a set of ten of the similarity results of the randomly selected two cases. Interestingly, we found that the Sony Pictures Entertainment and Dark Seoul cases are very similar as shown in Figure4.

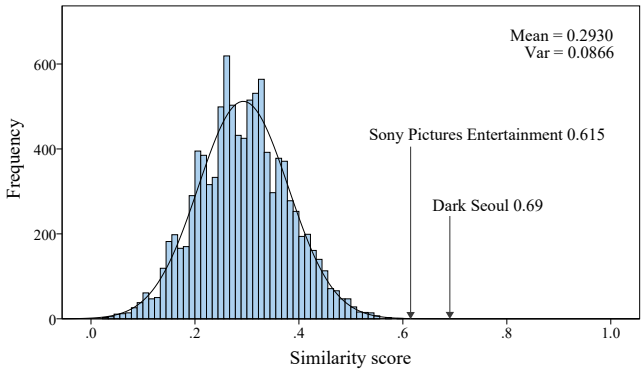

Fig. 4. Similarity score distribution of two random cases

IV. Conclusion

Intelligence derived from WHAP cannot guarantee a full accuracy, but highlights hidden similarities between various cases and assists investigator in timely decision-making. It contributes by providing insight into the confidential and undercover network of cyber-crime as well, especially when there is a lack of information. WHAP makes the analysis easier and reduce the time required in searching possible suspects.

Acknowledgment

This work was supported by the ICT R&D Program of MSIP/IITP.[14-912-06-002, The Development of Script-based Cyber Attack Protection Technology]

References

[1] Symantec Security Response. "Four Years of DarkSeoul Cyberattacks against South Korea Continue on Anniversary of Korean War," http://www.symantec.com/connect/blogs/four-years-darkseoul-cyberattacks-against-south-korea-continue-anniversary-korean-war, 2013.

[2] Zone-H.org, http://zone-h.org/.

[3] H. K. Kim, K. H Im, and S. C. Park, "DSS for Computer Security Incident Response Applying CBR and Collaborative Response," Expert Systems with Applications, vol. 37, pp. 852-870, 2010.

[4] D. B. Leake, "Case-Based Reasoning: Experiences, Lessons and Future Directions," MIT press, 1996.