Poster: Knowledge Seeking on The Shadow Brokers

Seung Ho Na*
KAIST
harry.na@kaist.ac.kr

Kwanwoo Kim*
KAIST
kw2128@kaist.ac.kr

Seungwon Shin
KAIST
claude@kaist.ac.kr

ABSTRACT

The Shadow Brokers (TSB) are an infamous group of hackers responsible for major cybercrime incidents. There are currently few studies on TSB, and to prevent their attacks in the future, we believe it is necessary to study them beforehand. This study constructs a relation graph of all the entities concerning TSB using identifiers in the Web. We introduce a systematic approach to finding relations among entities using a case study based on identifiers and clearness of relations. Our investigation covers data from both the Surface Web and Dark Web, with our Dark Web data consisting of over 40 million Dark Web webpages. We have uncovered many hacking forums, hacking groups, and individuals having a relation with TSB using our method. The relation graph of TSB will become a stepping stone in developing a knowledge base of TSB.

CCS CONCEPTS

• **Applied computing → Investigation techniques**; • **Social and professional topics → Computer crime**;

KEYWORDS

Bitcoin, Dark Web, Systematic Investigation, The Shadow Brokers

ACM Reference Format:
Seung Ho Na, Kwanwoo Kim, and Seungwon Shin. 2018. Poster: Knowledge Seeking on The Shadow Brokers. In *2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18), October 15–19, 2018, Toronto, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3243734.3278512

1 INTRODUCTION

The Internet is a widely useful invention that has merged into the everyday life of people, worldwide. The increasing usage of the Internet, however, has introduced a platform for a new set of crime: cybercrime. Any criminal activity that involves a computer and a network such as the Internet can be classified as cybercrime. A worldwide cybercrime incident example is WannaCry, a ransomware outbreak that occurred in mid-2017. WannaCry utilized leaked hacking tools originated from the National Security Agency (NSA). The culprit known to be behind hacking the NSA is a hacker group called The Shadow Brokers (TSB) [6].

TSB has been known as an active hacker group that hacks software exploits and sensitive information from organizations such as NSA and distributes them online for a price. In addition, they are responsible for major ransomeware attacks such as WannaCry [8] and Petya variant NotPetya [9], and pose as a global security threat. Despite the attention on TSB, there are few studies on the group itself and their behaviors in the cyber space are still unclear.

In this paper, we intend to figure out the relationships that TSB has with other entities and ultimately, systematizing an automatic knowledge base constructor to understand TSB's behaviors. Specifically, we try to comprehend which domains in the Surface/Dark Web are connected with TSB, because we believe that this connectivity can reveal TSB's operations, intentions, and finally overall behaviors. To do this, we construct a relationship graph of all related entities (i.e., linked Surface or Dark web sites) concerning TSB using identifiers and found in the Surface web and Dark web. Most likely due to the group's shady nature, TSB commonly accepts payment in cryptocurrency, and the most common cryptocurrency used is Bitcoin. Bitcoins can be exchanged anonymously but because the blockchain mechanism is established on a peer-to-peer network, transactions of addresses can be looked up by anyone [5]. By looking at TSB's Bitcoin wallet addresses that have been reported by private intelligence agency Tactical Rabbit[7], the interactions with other addresses can be seen, denoting relations with those addresses. Using data from the surface web and dark web, identifiers of possible TSB concerned entities can be found and linked, starting with the above mentioned addresses.

To conduct this research, we first collect Dark web sites by leveraging existing Dark Web indexing services (e.g., Ahmia [1]), and obtain more than 40 millions of Dark Web pages ranging from more than 40 thousand domains. Then, we track how Bitcoin payments for TSB are operated in the Dark Web, since Bitcoin payments in the Dark Web are likely to be linked with cyber crimes. In addition, we employ Google to trace these payments in the Surface Web. In our preliminary evaluation, we disclose that several hacking groups/forums are strongly related with TSB. Among them were famous Dark web forums such as Hell Forum and Ex0du$. We are currently extending this work to find more related information with TSB and finally understand TSB's overall behaviors in the cyber space.

2 METHODS

Direct transactions are the most certain relations between two Bitcoin wallets. The initial step of our investigation was to examine The Shadow Brokers (TSB) direct transactions of the seeded main Bitcoin wallet addresses posted from Tactical Rabbit. We considered all direct deposits from those addresses as identifiers of entities concerning TSB. However, many of the linked addresses were disposable Bitcoin wallets having a very small number of transactions, which is a main characteristic of mixing service wallets. Once a transaction enters a mixing service, addresses become very difficult to track.[4] To filter out the addresses involved with mixing services,

*co-authors of this study

CCS '18, October 15–19, 2018, Toronto, ON, Canada

ACM ISBN 978-1-4503-5693-0/18/10.
https://doi.org/10.1145/3243734.3278512

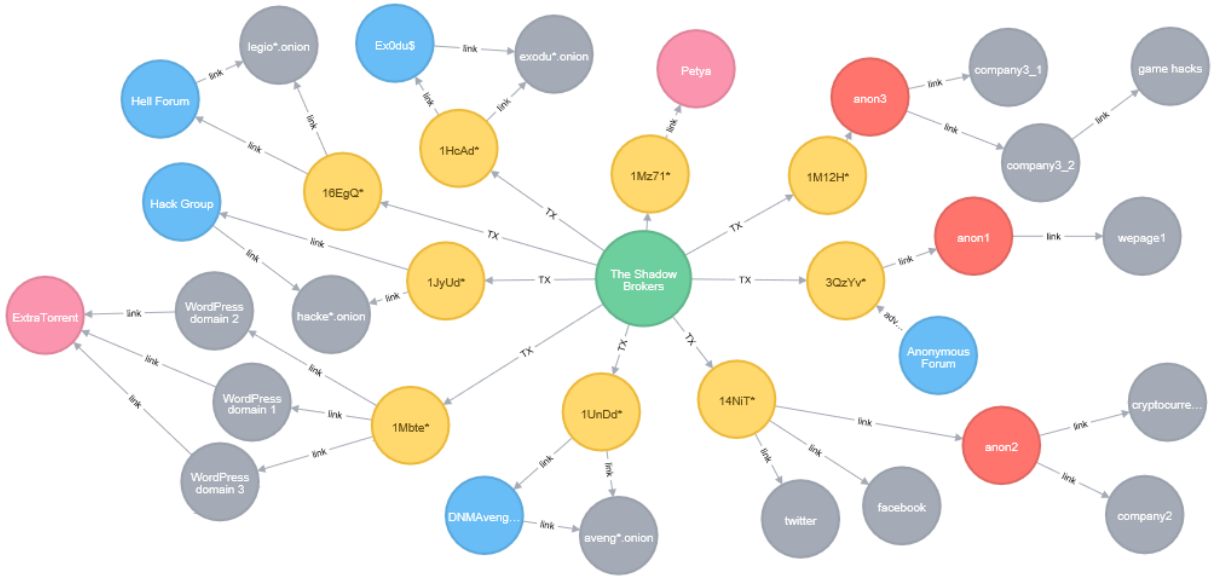


Figure 1: Visualized Entity Relation Graph. This figure shows a portion of the relations we have uncovered. Each censored label refers to an identifier with sensitive information. The actual information is not shown on the graph due to ethical issues. Identifiers of the same class are shown in the same color (e.g. forums are blue, Bitcoin addresses are yellow). All nodes that are connected together by a 'link' form an identifier chain, representing an entity. Petya variant NotPetya is labelled as Petya.

we narrowed down the pool of addresses to those that have also appeared in our Dark Web repository. In addition to the anonymity of the Dark Web, disposable addresses used in mixing services or temporary transactions are not advertised on the Dark Web, making our repository a valid filter. We then search both the Surface Web and Dark Web on these addresses to discover identifiers of corresponding entities.

An identifier, in a sense, would be a feature of an entity; it is a *pointer* to an entity. Identifiers pointing to equivalent entities would be linked together. These links would form an identifier chain, which would denote a single entity. An entity is an independent presence, may it be an individual or group that is pointed at by an identifier. A strong identifier points to a single entity while a weak identifier would point to multiple. Strong identifiers are favored because they more accurately depict and characterize an entity. Depending on the identifiers, relations among entities become explainable, or clear. In order to link new identifiers and find relations among entities concerned with TSB, an in-depth investigation is required. In this study we manually distinguished relations of entities and identifier links because of the need for a higher level of interpretation of information. We devised a case work for extracting the relations among entities:

(1) When there is a strong identifier with clear relations
(2) When there is a strong identifier with unclear relations
(3) When there are only weak identifiers

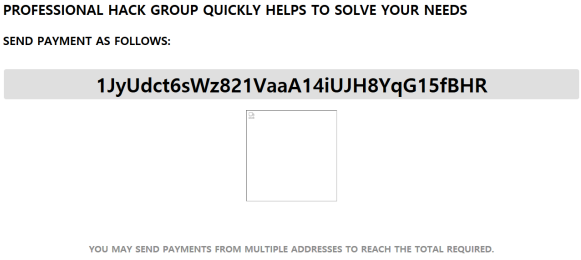
The first case is basically what we are aiming to do. The second case will be explained with an illustration. If a bitcoin address is found as the official address of a Bitcoin donation onion domain, we assume the keywords found with the address are new identifiers to the same entity, collecting more features. For example, name of domain, onion address, and e-mail address found with donation wallet are new identifiers and linked together. By using many features from the identifier chain, the motivation of a relation among entities can be found, becoming a clear relation. When search results of an address are related to multiple entities, the problem is complicated. We assume that there is a similarity in the entities of the weak identifier that would lead to a common entity and start a comparison analysis. When a common entity emerges, we change the identifier into a strong identifier and repeat the above process. The examples for each case will be explained in section 3.

3 CASE STUDY

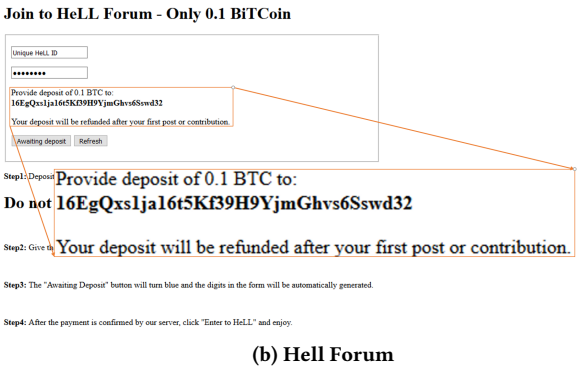
Figure 1 shows a portion of our relation graph of the entities concerned with the Shadow Brokers(TSB). There are many payments linked to forums/groups in the Dark Web, as expected. One familiar identifier we see on the relation graph is that of NotPetya. There were also quite a number of individual entities that had identifiers of game hacks, anonymous forums, and cryptocurrency. We will now explain some example cases of how we achieved these relations.

3.1 Strong Identifier with Clear Relations

One transaction from the list of TSB addresses was a payment of 0.1 BTC to *'1JyUd*'*. The search result from our Dark Web repository uncovered it to be a webpage of Professional Hack Group's domain [2] as shown in Figure 2a. The webpage clearly indicates that the address is for receiving payment of hacking services, which means that TSB purchased services from Professional Hack Group. Similarly, TSB made a transaction sending 0.1 BTC to Bitcoin address *'16EgQ*'*. Figure 2b is the registration page of 'Hell Forum'[3], and can be seen that the address is used to collect membership



(a) Professional Hack Group



(b) Hell Forum

Figure 2: Webpages from various sources and identifiers

fees. The membership payment matches the transaction, so we could reasonably conclude that TSB has signed up as a member of Hell Forum. In these cases, because it can be known that TSB had made transactions with hacking groups/forums for "purchases" and "membership", the relations are very clear.

3.2 Strong Identifier with Unclear Relations

From the transactions, there was a strong identifier Bitcoin address *'14NiT*'*. The address was followed with other strong identifiers: the owner's name and social network links which become new identifiers. However, given the identifiers, it was difficult to fathom what the motivation of the transaction was, making it an unclear relation. In other words, we lacked the background data of the entity, which calls for more identifiers. Additional search using the given identifiers lead to a version control system account owned by the entity. We found that the entity had close ties to an alternative coin and a Bitcoin funding company. From these additional features we may not be able to deduce the exact motivation, but see more value and possibilities in the relation.

3.3 Weak Identifiers

Bitcoin address *'1Mbte*'* searches from the Surface Web lead to multiple '.torrent' file download pages. Surprisingly, each webpage belonged to domains of many companies that should have no connection with torrent files. In addition, the companies themselves were not related in any means with each other, so they were independent entities. As mentioned in section 2 however, the identifier should point to a single entity, so we assumed that there must be some feature involved with hacking held in common. From those searched pages, the common keyword 'extraTorrent' appeared repeatedly, and each domain had a "HELLO WORLD!" post from WordPress. The common feature was the fact that the companies used WordPress for their homepages, and an entity was able to hack WordPress based homepages. We could conclude the existence of a hacking group behind the identifier 'extraTorrent', and that they hacked the homepages using an exploit of WordPress. This inference induces the relation between TSB and a hacking group that releases torrent files. We now have an entity derived from the weak identifier with a clear relation.

4 DISCUSSION

In this study, we verified the feasibility of investigation on the relations of entities with The Shadow Brokers (TSB) using Web sources. We successfully discovered entities with direct financial transactions with TSB. Although some processes were done manually, the systematic method is the linchpin to the knowledge base system for automatic investigation. We discovered that a majority of entities in our relation graph are hacking forums or groups, and the second largest portion are individuals and companies having to do with Bitcoin and cryptocurrency technology. These mentioned fields might be the interests of TSB, worthy of looking out for in investigations. This knowledge profiling of TSB concerned entities could be a new starting point of our future study.

Future work. To reach the ultimate goal of building a knowledge base of TSB, we aim to develop a classifier that extracts the identifiers pointing to the same entity with machine learning and natural language processing techniques. This makes automatically linking identifiers in structural or semantic relations possible on the vast data available on the Web. In addition to the automation of linking identifiers, we intend to introduce variables on the strength of identifiers and clearness of relations using latent parameters with probabilistic models. Our manual tagging used on this study can be used as a ground truth. By expanding our investigation pool using automatic tools to develop, we will be able to build a knowledge base of TSB.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No.2016-0-00078, Cloud based Security Intelligence Technology Development for the Customized Security Service Provisioning)

REFERENCES

[1] ahmia.fi. [n. d.]. ahmia, hidden service search engine. https://ahmia.fi
[2] hackerrljqhmq6jb.onion. 2017. Professional Hack Group. Retrieved April 14, 2017 from http://hackerrljqhmq6jb.onion/checkout.php
[3] legionhiden4dqh4.onion. 2017. Hell Forum. Retrieved April 11, 2017 from http://legionhiden4dqh4.onion/reg.php
[4] Malte Moser. 2013. Anonymity of bitcoin transactions. (2013).
[5] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. (2008).
[6] Ellen Nakashima and Craig Timberg. 2017. NSA officials worried about the day its potent hacking tool would get loose. Then it did. *The Washington Post* (May 2017). https://www.washingtonpost.com/business/technology/nsa-officials-worried-about-the-day-its-potent-hacking-tool-would-get-loose-then-it-did/2017/05/16/50670b16-3978-11e7-a058-ddbb23c75d82_story.html?noredirect=on&utm_term=.7bcaa55da133
[7] tacticalrabbit.com. 2017. shadowBroker. Retrieved August 21, 2018 from https://tacticalrabbit.com/wp-content/uploads/2017/05/shadowBroker01_final.pdf
[8] US-CERT. 2018. Indicators Associated With WannaCry Ransomware. https://www.us-cert.gov/ncas/alerts/TA17-132A
[9] US-CERT. 2018. Petya Ransomware. https://www.us-cert.gov/ncas/alerts/TA17-181A