

A Framework for Dark Web Threat Intelligence Analysis

Xuan Zhang, Criminal Investigation Department of Shandong Police College, Jinan, China

KP Chow, Department of Computer Science University of Hong Kong, Hong Kong, China

ABSTRACT

This article describes how the Dark Web is usually considered the dark side of the World Wide Web. Cyber criminals usually use specialized tools, e.g. TOR, to access the hidden services inside the Dark Web anonymously. Law enforcement officers have difficulty tracing the identity of these cyber criminals using traditional network investigation techniques that are based on IP addresses. The information available in the Dark Web, which includes BitCoin wallets, email addresses, hyperlinks, images and user behavior profiles, can be used for further analysis, such as a correlation analysis. Present within this article is a threat intelligence analysis framework to help analyze the crimes and criminals in the Dark Web and the framework is realized by the implementation of the Dark Web Threat Intelligence Analysis (DWTIA) Platform.

KEYWORDS

Dark Web, Data Mining, Threat Intelligence, TOR

INTRODUCTION

The World Wide Web (WWW) is much bigger than what people see today. Existing search engines, e.g. Google and Baidu, can only search approximately 5% of the whole WWW. Besides those searchable contents, there are a lot more resources and data that are available on the Internet and such places are usually known as Deep Web and Dark Web (Pagliery, 2014). Deep Web usually refers to resources and data that are available on the Internet, but are not accessible with normal web browsers and hyperlinks. According to some statistics, the part of WWW that are accessible by normal web browsers (also known as Surface Web) contain approximately 4 billion web sites, while the Deep Web contains several times more web sites than the Surface Web. A portion of Deep Web that is widely used for criminal activities, such as drugs dealing, child pornography, weapons selling, etc., is known as the Dark Web. Items for sales in the Dark Web include stolen email accounts and credit card numbers, personal identity information and medical information, fake identities, design drawings, malware, systems vulnerabilities, child pornography, drugs, weapons, and hire to kill services. Most of the sales items are illegal (Vogt, 2017).

Cyber criminals usually use specialized tools, e.g. the TOR, to access the hidden services inside the Dark Web anonymously. Law enforcement officers have difficulty to trace the identity of these cyber criminals using traditional network investigation techniques that are based on IP addresses. Therefore, specialized intelligence analysis techniques are needed to trace cyber criminals in the Dark Web. Law enforcement agencies all over the world are trying to trace the identity of users that access the Dark Web using TOR and progress is very limited. Silk Road, an e-commerce platform in the Dark Web, was launched in February 2011 selling illegal items. Due to the support of the hidden

DOI: 10.4018/IJDCF.2018100108

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

services protocol and TOR, Silk Road was able to hide its identity from law enforcement agencies. Silk Road was taken down by FBI in October 2013 and Silk Road 2.0 was taken down later. In 2015, after the discovery of the child pornography website Playpen in the Dark Web, FBI used Network Investigative Technique (NIT) tools to trace hidden users behind the encrypted and anonymous TOR network (Condliffe, 2016). FBI eventually found more than 1,300 “real” IP addresses, of which 137 users were charged with crimes (Osborne, 2014). However, these two cases also caused a lot of controversy. Does the use of hacking techniques to trace network users that are using anonymous tools compliance with laws and regulations? In April of 2016, the Supreme Court of US approved a change to the existing Rule 41 that would allow US federal judges to issue search warrants to use NIT to hack computer anywhere (Moore et al., 2016).

As the increase in popularity of the Dark Web by normal web users, how to conduct cybercrime investigation in the Dark Web in a legal manner becomes a challenge to law enforcement officers. How to identify the anonymous web surfer in TOR? How to identify the e-commercial sites that are using hidden services in TOR? All these are new challenges to today’s law enforcement agencies. With tools like FBI’s NIT, it has to rely on system’s vulnerabilities even it is allowed under the legal framework. By collecting and analyzing large volume of data and information from the Dark Web may be a possible alternative to assist law enforcement officers to combat cyber criminals. In this paper, we present a framework which investigators can analyze data and information from the Dark Web, which includes BitCoin wallets, email addresses, hyperlinks, images and user behavior profiles. This deep analysis can help investigators to have a better understanding of potential crimes and behavior of the criminals. The proposed framework is realized in the design of the Dark Web Threat Intelligence Analysis Platform.

THE DARK WEB

To access the Dark Web, specialized tools are needed, such as TOR (The Onion Router), I2P and Freenet. All these tools support anonymous web serving. The most popular one is TOR. The principle behind TOR is communication between a user and the server will go through many routers and all communications are encrypted. Moreover, communication between any 2 routers uses different encryption key. Therefore, no one is able to trace where the real user and the server are. In this manner, anonymous web serving can be guaranteed.

The TOR Router

In 1990, the US Naval Institute has begun studying the Onion Routing network project. The American Defense Advanced Research Institute takes over onion routing. Syverson and some scientists began designing the alpha version of Tor, which use for routing onions technology. And then the relevant source code for the Tor project was leaked from the US Naval Research Institute to the Internet. Mathewso and five other scientists restarted the Tor project and built tor network nodes on servers around the world in 2006.

Under normal circumstances, a connection from a user to a web server in TOR involves 3 parties: Entry Guard router, Middle router(s), and Exit router (Cox, 2016).

The Entry Guard router is the entry point to the TOR network. When a TOR router has been existed for a period of time with stable connection and of sufficient bandwidth, it will be selected to be an Entry Guard router. When a user accesses the TOR using an application, e.g. the TOR Browser, he will be connected to one of the Entry Guard router.

Middle routers are intermediate hops of the TOR network. They are responsible to relay the traffics from the Entry Guard routers to the Exit routers. There is no direct connection from the Entry Guard router to the Exit router, and therefore either party is unable to determine the identity of the other party.

Exit routers are routers on the boundary of the TOR network. They are responsible to deliver the messages to the target server as specified by the user. TOR, stands for The Onion Router, means the request from the user is encrypted multiple times, layer by layer, like an onion. Only the Exit router has the original decrypted request from the user. Every Middle router is responsible to remove a layer of encryption from the encrypted message, like pill a layer from an onion. Please be noted that the raw data of the request from the user will be decrypted at the Exit router, and therefore the Exit router will know the content of the request from the user, but unable to determine the actual identity of the user. If the request of the user contains plain information, e.g. HTTP request or FTP request, the Exit router will have full knowledge of the request.

The Hidden Services

TOR not only supports anonymous web serving, but also supports hidden services. With hidden services, the identity of the web server can provide services in an anonymous manner, and no one is able to identify the real location of the web server. The hidden services web server will use the top level domain.onion. In TOR, requests to hidden services with the top level domain.onion will be routed to the corresponding web server automatically.

With the support of Tor2Web, users can add the suffixes .cab or .to to the original hidden services domain name, and access the services without using the TOR browser. With the OpenData service by the Tor2Web, many statistics on accessing the Dark Web through Tor2Web can be collected. These statistics are the starting point for our Dark Web Threat Intelligence Analysis Framework (Ford, 2016).

Dark Web Market

There are a lot of e-commerce platforms in the Dark Web, such as AlphaBay, CryptoMarket, Hansa, etc. These platforms provide different types of services, such as hire to kill and hackers hiring, and items to sell, such as drugs, weapons, child pornography, stolen accounts, fake passports. Many of these transactions will go through the hidden services protocol, of which the user and the service can be hidden from each other. The platforms also provide some types of credit rating mechanisms, so that the buyers and the sellers can have an estimate on the “credits” of the other party. These can improve the confidence of the buyer and the seller of a transaction. To ensure anonymous, BitCoin is the most commonly used payment methods in the Dark Web. The Dark Web also provides an encrypted delivery address to the parties. When the buyer completes the payment of a transaction, the seller will deliver the items to the encrypted address to complete the transaction. The buyer can also put down his/her comment about the seller. The process is quite similar to the transaction in normal e-commerce sites in the Surface Web.

As the transaction in Dark Web can be completed in an anonymous manner and the whole process is opaque, these types of transaction may not be that stable and may lead to unrecoverable loss. For example, when the buyer has completed the payment, the seller decides not to conduct any more selling in the Dark Web and deletes all his identities. The buyer will not be able to trace the seller.

Based on the Dark Web technology and the market in the Dark Web, our proposed Dark Web Threat Intelligence Analysis Framework is based on the following techniques:

- **Honeypot Router:** A well-known technique by law enforcement agencies, such as FBI, to establish honey Entry Guard router to collect user information when someone enter the TOR network
- **Virtual Identity Profiling:** Correlate user identities in the Dark Web and user identities in the Surface Web, so as to identify the real user identities
- **Web Site Vulnerabilities:** Using vulnerabilities in the web sites to identify the real IP addresses, MAC addresses, etc. of the web servers
- **Traffic Analysis:** Monitor the traffics of the Exit routers in order to collect the plain contents from the users, with the goal to identify the real users

- **Behavior Analysis:** Based on the user behavior on navigating the Dark Web using the TOR browser, so as to obtain a user behavior profile.

The proposed framework is realized in the implementation of the Dark Web Threat Intelligence Analysis Platform. We will give a brief description of the Platform in the next section.

DARK WEB THREAT INTELLIGENCE ANALYSIS PLATFORM

System Overview

The Dark Web Threat Intelligence Analysis Platform (DWTIA Platform) is designed to process large volume of information. It combines information collected from the Dark Web and the Surface Web, together with potential criminal activities, to identify the actual criminal. The architecture of DWTIA Platform is as shown in Figure 1.

To interface with typical threat intelligence platform, the DWTIA Platform adopts STIX 2.0 standard for data acquisition and sharing. Through the Platform, investigators can predict and combat crime on the Dark Web. The key data collected and processed by the Platform are:

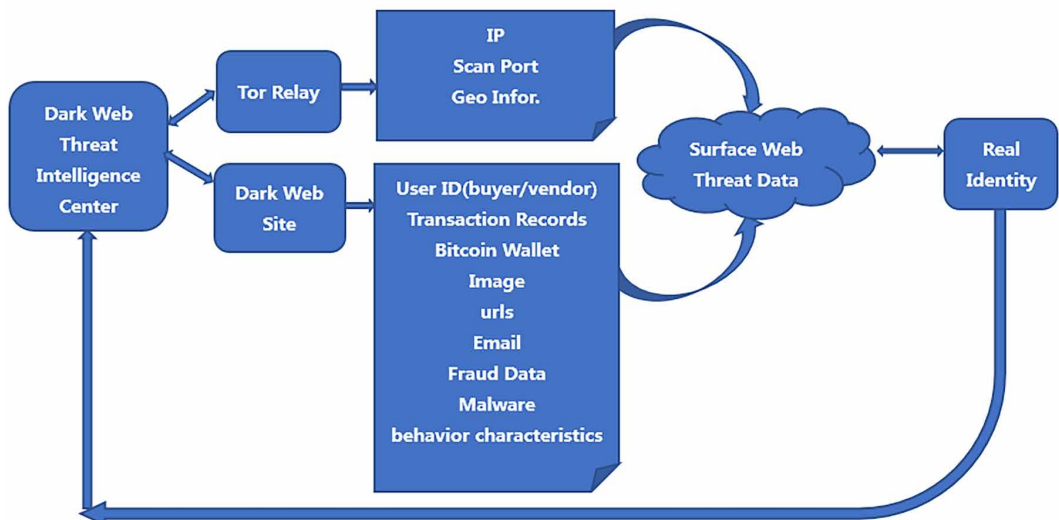
TOR Web Site Dynamic List: Through Tor2web's OpenData service, the Platform will detect the trend traffic flow of websites inside the Dark Web.

TOR Relay List: TOR Network Status allows monitoring of TOR exit nodes, and also provides a list of Exit nodes. The Platform will detect potential threats based on IP addresses, host names, locations and other information of TOR relays.

Critical Information: Based on the TOR web site list, we can extract the critical information, which includes User ID (buyer/vendor), transaction records, Bitcoin wallet, images, URLs, emails, fraud data, and malware.

Correlation Analysis: Using machine learning methods, the Platform correlates the Surface Web data with the critical information in the Dark Web.

Figure 1. Dark Web Threat Intelligence Analysis Platform



Threat Intelligence Center Feedback: The TOR websites list, relay list and critical information will be feedback to the Threat Intelligence Center.

Methods

The Platform consists of the following processing modules to collect and analyze the data:

Data Acquisition Module: This module manages the list of dynamic web sites in the Dark Web, and crawl data from these web sites.

Indexing Module: A multidimensional index is constructed to allow efficient access to the collected data.

Data Analysis Module: Machine learning methods are implemented to analyze the data, so as to identify the real user identities.

Data Visualization Module: The module supports visualization of multi-dimensional and multi-level data.

IMPLEMENTATION

We have implemented a proof-of-concept system that includes the main functions of dynamic dark web site list, TOR exit routers list, illegal transaction analysis etc. as Figure 2 shows.

Dynamic Dark Web Site List

Dark Web intelligence starts from the list of web sites in the Dark Web. The DWTIA Platform maintains a dynamic list of web sites in the Dark Web which will be used as starting points for intelligence gathering. The Platform uses the data from the Tor2web OpenData Project, which contains monitored traffics in the Dark Web. There are about 1 million requests per day in the Dark Web, which involves about 2,000 web sites. Figure 3 shows the top 20 web sites in the Dark Web in 9 April 2017.

By monitor the web sites in the Dark Web and the corresponding access patterns, the Platform can determine how active a web site is. Figure 3 shows the changes in access pattern of the Dark Web market Hansa (hansamk2bizhmib4.onion) in 10 days from 1 April to 10 April. Web sites with active access patterns should be the focus of further analysis.

TOR Exit Routers List

By monitoring TOR Exit routers, we can effectively prevent network attacks and malware distribution (Hardesty, 2015). There are approximately 8000 TOR Exit routers. The web site <http://torstatus.blutmagie.de/> keeps the list of TOR Exit routers, which includes the IP addresses and the host names of some Exit routers. Figure 5 is the Exist Relay map.

Illegal Transaction Analysis

Based on the list of web sites in the Dark Web, we deployed the Dark Web crawler, OnionScan, to collect data of more than 8,000 web sites. By data mining and correlation analysis, we are able to reveal deep relationship of buyer and vendor. In some cases, we can even trace where the real user and the server are.

Correlation analysis of user data of Dark Web forum and the Surface Web data provide clues and evidence for further investigation. The data association rules include attributes such as User ID (buyer/vendor), transaction records, Bitcoin wallet, images, URLs, emails, fraud data, and other elements. For example, 'Karmacticals' is a vendor in the Dark Web market named malware

Hansa. Through the analysis of the data link to other web sites, we found the same User ID appeared in other trading platforms, such as Dream, Valhala, Alphabay, Traderoute etc. We can then analyze the relationship of illegal transactions from different sites.

Figure 2. DWTIA Platform Interface

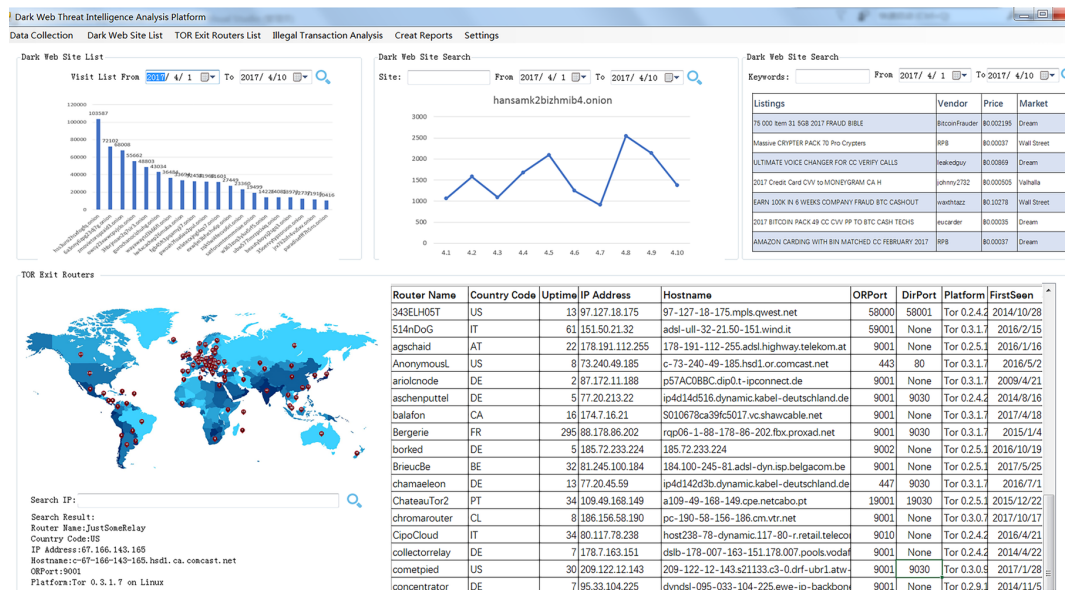


Figure 3. Dark Web Site Top20(2017.4.9)

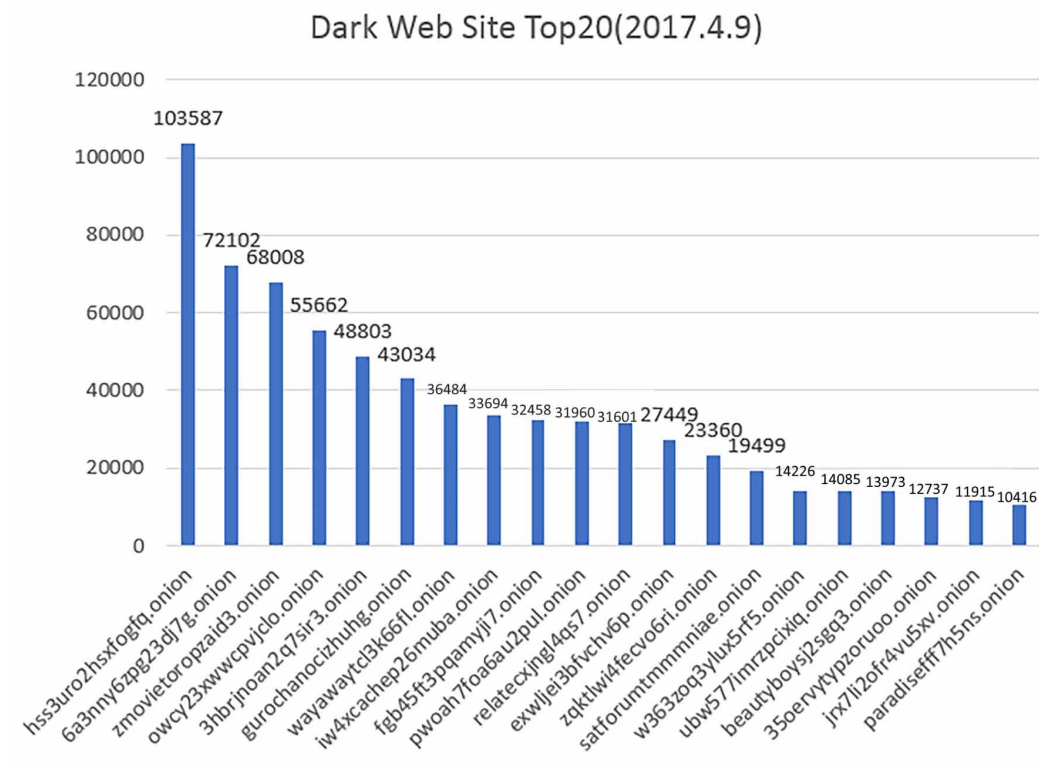


Figure 4. Access records of Hansa (2017.4.1-2017.4.10)

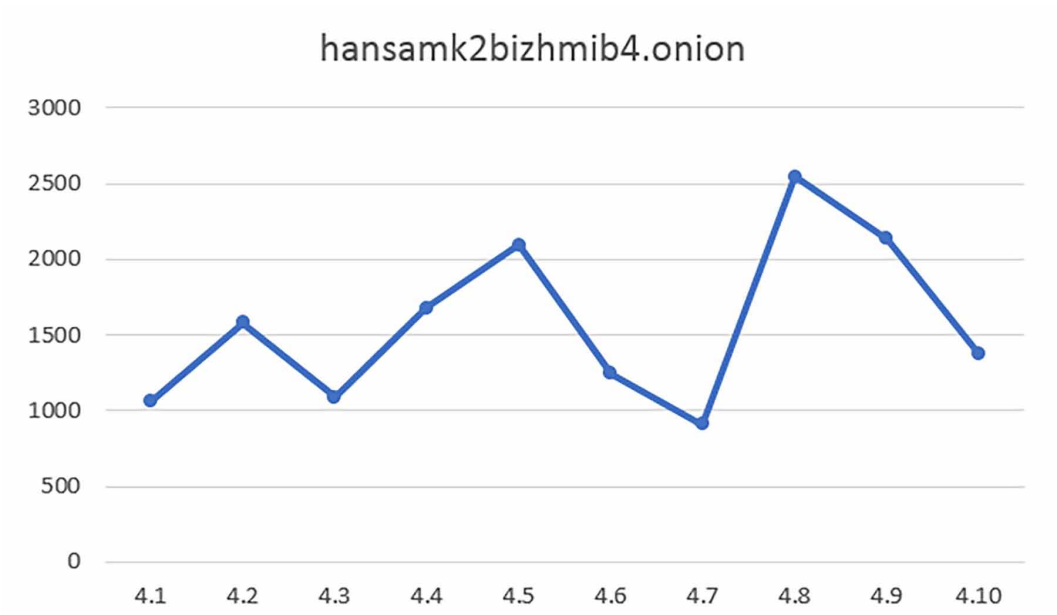


Figure 5. Exit Relay Map



The flow chart (Figure 7) shows the key steps in the analysis process, which includes data acquisition, data storage, data analysis, and visualization of results. In data acquisition, the crawler collects data according to the TOR web site dynamic list. Data are then stored and indexed

Figure 6. Illegal Trading Analysis

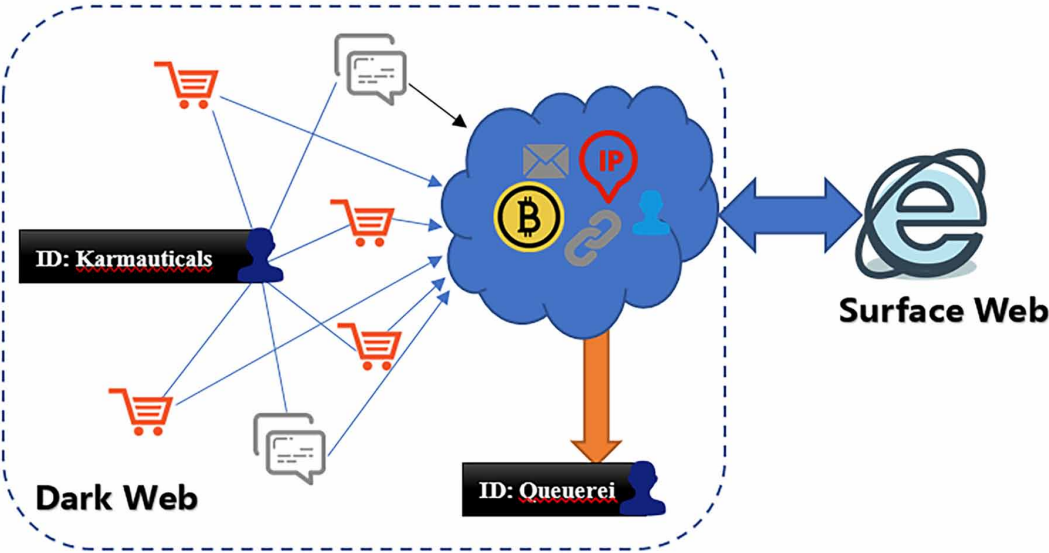
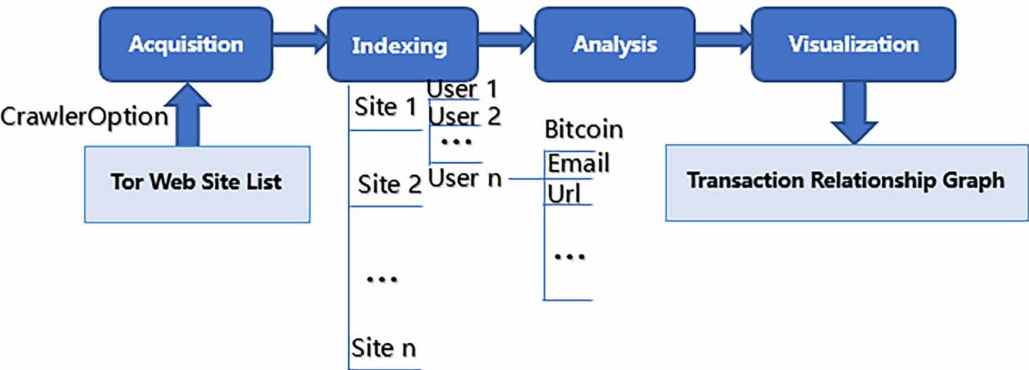


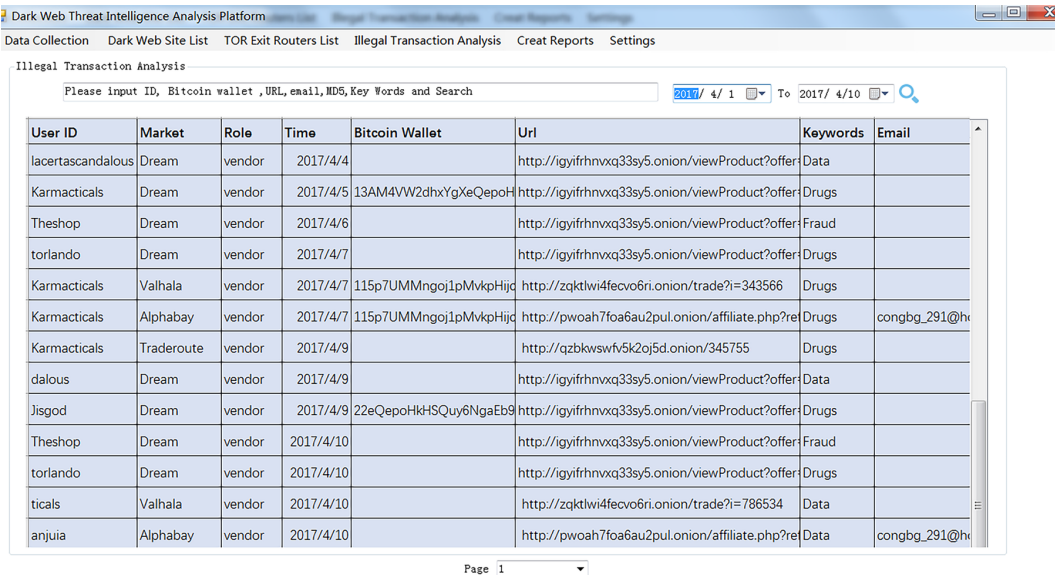
Figure 7. Steps of Analysis Process



according to where they were crawled. The stored data includes the user ID and its associated email, Bitcoin wallet and URL. All these data are processed at the analysis step.

Figure 8 shows the interface of illegal transaction analysis. The crawler program helps us collect all the illegal trade records from several dark web market platforms. The data items include User ID, Market, Role, Time, Bitcoin Wallet, URL, Keywords, Email, MD5 of malware files etc. And we can use these data items to search for data. For example, if we input “Kamactices” into the search bar, all relevant information will be listed. We can use search results to mine this user’s registration information on different platforms. We can see that the user “Kamactices” has registered the same ID on different platform as Dream, Valhala, Alaphbay, Traderoute etc. He also left several bitcoin wallet address and Email on some platforms. We can also associate another registered account “anjuia”

Figure 8. Illegal Transaction Analysis Interface



Dark Web Threat Intelligence Analysis Platform

Data Collection Dark Web Site List TOR Exit Routers List Illegal Transaction Analysis Creat Reports Settings

Illegal Transaction Analysis

Please input ID, Bitcoin wallet , URL, email, MD5, Key Words and Search

2017/ 4/ 1 To 2017/ 4/10

User ID	Market	Role	Time	Bitcoin Wallet	Url	Keywords	Email
lacertascandalous	Dream	vendor	2017/4/4		http://igyifrhnxq33sy5.onion/viewProduct?offer	Data	
Karmacticals	Dream	vendor	2017/4/5	13AM4VW2dhxYgXeQepoH	http://igyifrhnxq33sy5.onion/viewProduct?offer	Drugs	
Theshop	Dream	vendor	2017/4/6		http://igyifrhnxq33sy5.onion/viewProduct?offer	Fraud	
torlando	Dream	vendor	2017/4/7		http://igyifrhnxq33sy5.onion/viewProduct?offer	Drugs	
Karmacticals	Valhala	vendor	2017/4/7	115p7UMMngoj1pMvxpHjcd	http://zqktlwi4fecvo6ri.onion/trade?i=343566	Drugs	
Karmacticals	Alphabay	vendor	2017/4/7	115p7UMMngoj1pMvxpHjcd	http://pwoah7foa6au2pul.onion/affiliate.php?ref	Drugs	congbg_291@h
Karmacticals	Traderoute	vendor	2017/4/9		http://qzbkswsfv5k2oj5d.onion/345755	Drugs	
dalous	Dream	vendor	2017/4/9		http://igyifrhnxq33sy5.onion/viewProduct?offer	Data	
Jisgod	Dream	vendor	2017/4/9	22eQepoHkHSQuy6NgaEb9	http://igyifrhnxq33sy5.onion/viewProduct?offer	Drugs	
Theshop	Dream	vendor	2017/4/10		http://igyifrhnxq33sy5.onion/viewProduct?offer	Fraud	
torlando	Dream	vendor	2017/4/10		http://igyifrhnxq33sy5.onion/viewProduct?offer	Drugs	
ticals	Valhala	vendor	2017/4/10		http://zqktlwi4fecvo6ri.onion/trade?i=786534	Data	
anjua	Alphabay	vendor	2017/4/10		http://pwoah7foa6au2pul.onion/affiliate.php?ref	Data	congbg_291@h

Page 1

with the same Email address. At last, we will get three bitcoin wallet addresses, one Email address and another account registered by the same person of user “Kamactices”. Further on, with big data on the Internet, we can correlate more data.

CONCLUSION

In this paper, we present a Dark Web Threat Intelligence Analysis Framework to help analyze the crimes and criminals in the Dark Web. The Framework is realized in the implementation of the Dark Web Threat Intelligence Analysis (DWTIA) Platform, which can help law enforcement agency to conduct criminal investigations in the Dark Web by gathering effective and relevant intelligence. The Platform supports dynamic web site list in the Dark Web and the TOR Exit Relay List. The Platform uses OnionScan Dark Web crawler to collect data of more than 8,000 sites in the Dark Web. By data mining and correlation analysis, we can reveal deep relationship of buyer and vendor. In some cases, we can even trace where the real criminals are. In the future, we will improve the data analysis model and develop automated tools to conduct efficient and effective analysis.

ACKNOWLEDGMENT

The research is supported by Shandong Social Science Fund. (Grant No.14CXWJ10)

The authors thank the China’s leading security intelligence service provider TianJi Partners to provide some of the data in this paper.

REFERENCES

- Condliffe, J. (2016). Judge Rules Evidence From FBI's Pedophile Tor Hack Is Invalid. *Gizmodo*. Retrieved from <http://gizmodo.com/judge-rules-evidence-from-fbis-pedophile-tor-hack-is-in-1772218525>
- Cox, J. (2016). FBI: Hacking Tool Only Targeted Child Porn Visitors. *Vice*. Retrieved from <https://motherboard.vice.com/read/fbi-hacking-tool-only-targeted-child-porn-visitors>
- Ford, M. (2016). The Supreme Court Expands FBI Hacking Powers. *The Atlantic*. Retrieved from <http://www.theatlantic.com/politics/archive/2016/04/supreme-court-fbi-hacking/480498/>
- Hardesty, L. (2015). Researchers mount successful attacks against Tor network—and show how to prevent them. *Phys.org*. Retrieved from <https://phys.org/news/2015-07-mount-successful-tor-networkand.html>
- Moore, D., & Rid, T. (2016). *Cryptopolitik and the Darknet, survival: global politics and strategy*. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/00396338.2016.1142085?needAccess=true>
- Osborne, C. (2014). Beyond Silk Road 2.0, over 400 'dark web' Tor sites seized by FBI. *ZDnet*. Retrieved from <http://www.zdnet.com/article/beyond-silk-road-2-0-over-400-dark-web-tor-sites-seized-by-fbi/>
- Pagliery, J. (2014). The Deep Web You Don't Know About. *CNN MONEY*. Retrieved from <http://money.cnn.com/2014/03/10/technology/deep-web/index.html>
- Vogt, S. D. (2017). The Digital Underworld: Combating Crime on the Dark Web in the Modern Era. *Santa Clara J. Int'l, L*, 15, 104.

Xuan Zhang is a Lecturer in the Investigation Department, at Shandong Police College. They became a lecturer of Shandong Police College after receiving a master's degree from Shandong University. Main research directions include network information security, digital forensics and cybercrime investigation. Zhang has finished four projects as Principal Investigator and participated a number of projects. Several relevant papers were published. The project "Research and Implementation of Traffic Accident Scene Representation and Image Distance" won the first prize in Shandong Province Science and Technology Progress. The project "Network Security Situational Awareness Simulation System" won second prize in Shandong province Public Security Science and Technology Progress. The project "Network Information Filtering System Based on Fuzzy Genetic Algorithm Research" won second prize of Shandong Public Security Science and Technology Progress.

KP Chow is an Associate Professor in the Department of Computer Science at The University of Hong Kong Associate Director, Center for Information Security and Cryptography Programme Director, MSc in Computer Science Associate Programme Director, MSc in Electronic Commerce and Internet Computing Visiting Professor, Liaoning Police Academy