Dark Web Threat Intelligence and Market Analysis

Huanhuan Yu, Ying Yang, Lina Yang and Guichun Zhu

ABSTRACT

It is described that the dark network is usually the dark side of the World Wide Web. At the same time, there are many threatening crime information in the dark network. It is a hidden part of the Internet and has gradually become a place for illegal activities, including the drug market, child pornography, gun sales, Financial fraud, assassination of employment, etc. Most criminals use specialized tools to access dark networks, such as TOR, which enable them to access networks anonymously. As a result, it has brought serious difficulties to research and law enforcement personnel. To achieve active cyber threat intelligence in cybersecurity research, this paper proposes a hidden network threat intelligence analysis framework, which is primarily used to identify the sellers of the dark network market and identify the sellers who have the greatest impact on these markets.

1. INTRODUCTION

The dark network is mainly a network that cannot be accessed by the existing browsers. At present, the browsers we use, such as Google and Baidu, can browse and visit web pages, which may be only about 4% of the entire World Wide Web. The pages that we can't access are called deep and dark.
Both the dark and deep nets are actually translated from English. Conceptually, a dark network is a network that needs to be connected through special software, special authorization, or special settings for computers; the concept of deep network refers to non-surface network content on the Internet that cannot be indexed by standard search engines. In short, the dark network is a subset of the deep network. Most of the dark web sites are illegal, but not all of the deep web sites are illegal. The deep network is not intentionally hidden. On the contrary, the deep network does not set anonymity, while the dark network is completely anonymous, which ensures the anonymity of visitors and service providers [1] [2].

Therefore, all kinds of illegal information exchanges flooded the dark network, especially after 2011, due to the rise of bitcoin technology, the dark network finally evolved from an anonymous information exchange to an anonymous value exchange stage. This subversive change, along with Silk Road The establishment (Silk Road: can be understood as a bitcoin-based dark e-commerce platform) has set off a climax of illegal transactions.

The biggest markets of the Darknet have been destroyed by the FBI in recent years, such as Silk Road and AlphaBay. Therefore, the dark market in the dark market has converged in 2017, not as popular as in previous years. How to identify visitors in the dark network and how to identify hidden services in the electronic morning is a challenge for researchers and law enforcement [3] [4]. Therefore, this paper proposes a dark-net threat framework for analyzing dark-net data, including buyer information and seller information, which can help researchers better understand the potential criminal behavior of criminals.

2. LITERATURE REVIEW

2.1 Tor Network Topology

Although there are many tools for accessing the Darknet, TOR is still one of the most popular tools. Tor uses a dedicated tor browser, and the Tor browser launches the Tor process in the background and connects to the network through it. Once the program is disconnected, the Tor browser automatically deletes privacy-sensitive data such as cookies and browsing history. The Tor browser is currently available in 16 languages, including Simplified Chinese (but no Traditional Chinese). The Tor browser itself provides SOCKS proxy services, and some applications have been able to use the Tor network [7] [8].

TOR and web server involve 3 parties: entry Guard router, intermediate router, and exit router. The Entry Guard is the first node in the circuit that accepts incoming traffic. The intermediate router is the intermediate hop node of the TOR that passes data to the next node. Exiting the router is the last node responsible for uploading data to the Internet [6]. When requesting access to a website, the web request arrives at the ingress node through multiple layers of encryption, the ingress node is passed to a randomly selected intermediate node that is placed around the world, and then the next node is selected, when the network request arrives at the egress node, all The encryption layer is cancelled, the last hop is encrypted, and the unencrypted request is finally sent to the web server [5]. TOR work flow chart is shown in Figure 1.
2.2 Illegal and Criminal Content

TOR's anonymity provides a secure platform for trading on the dark network. These buyers and sellers have the opportunity to conduct illegal transactions and reduce concerns about law enforcement officials. In the dark network, it is possible to find various illegal criminal products, sell drugs, and sell guns. Since the information on the dark network is opaque, all its transactions are done anonymously, so its transactions are unstable and risky, which may result in unpredictable losses in the transaction. For example, after the transaction is completed, the seller has completely disappeared, and some websites require high registration fees [9] [10].

2.3 Dark Web Market

In the early darknet market, the Silk Road was all the rage and it finally disappeared. What followed was the emergence of various markets. The Dream market, as a stable market, has been in operation since 2013. There are frauds and depressions in the dark network, which may cause an e-commerce platform to close quickly. The Wall Street market has fewer products, but it has more categories, including "security and hosting." There are many types of traded goods in the dark network, such as drugs, guns, fake passports, child pornography, information disclosure, etc. Users market their products on the market to attract customers to buy their goods.

3. FRAMEWORK DESIGN

The framework design of this paper has three main components; data collection, asset analysis, and framework functions. The details of each component are shown in Figure 2 and are described in the following sections.
3.1 Data Collection

After determining the dark web forum and market for research purposes, web crawlers can be used to automatically collect websites, using a custom web crawler to specify a starting seed site. Web crawlers are a common technique used in various Internet data collection projects. The crawler will automatically download the webpage it encounters while tracking the hyperlinks encountered and constantly discovering new webpages. Successfully downloaded cybercrime data can then be processed and archived for long-term storage. The crawler data flow chart is shown in Figure 3.
3.2 Asset Analysis

3.2.1 SUBJECT CLASSIFICATION AND EXTRACTION

When web crawlers traverse the forum, the pages they download must be processed to extract the information of interest. This article uses a regular expression text parser program to accomplish this task (somewhat similar to the program used in snowballs). Using a text parser to identify the HTML code pattern corresponding to the data of interest, used to extract the product name, category, description, shipping_options, shipping departure, shipping destination, price, and payment method and seller included in the web page (seller_name, member_since (date), pgp key, seller's description, feedback rating, and forum ID, forum name, author's name, etc.

3.2.2 SOCIAL NETWORK ANALYSIS AND VISUALIZATION

This paper uses social network analysis methods to analyze and identify seller suppliers. The social network analysis method is a graph-based approach for analyzing social relationships and their impact on individual behavior and organizational structure.

Here, the paper analyzes the extracted data using social network analysis methods. The graph theory is used to generate the bipartite graph, and the relationship between the product and the seller gang is represented by nodes, and the two-way network of the seller and the product is created. This paper uses the community SLPA algorithm to identify the product suppliers in the dark network. Divide all data sets into two disjoint sets, verify whether there are similar sellers in the disjoint sets; calculate the supplier's similarity matrix; product sellers have multiplicities, depending on the product category and the number of corresponding products shared between them. To infer the implicit connection between them and identify the seller suppliers in the dark network.

3.2.3 FRAME FUNCTION

The framework function module is mainly for the main functions provided by the user; mainly includes search, classification, browsing; interactive seller dashboard; interactive product list; and forum information list.

4. FRAME FUNCTION

The following sections provide descriptions and screenshots of the above framework features. It is mainly a visual interface, which is convenient for users to find browsing data.

4.1 Framework Description and Overview

The main login page of this article's framework contains three main sections that make it easy for users to access the collection of items in the framework. The first is
the asset repository for forums and markets. They can all be sorted by product name, seller nickname, price, etc. The second part is the ID of the forum, the name of the forum, the name of the author, and so on. Third, the visualization of social network analysis methods can describe the relationship between buyers and sellers. The main login page also contains an information page.

4.2 Searching

Figure 4 shows how users search for data on the dark market and forums. Users can search and sort in any provided metadata field product name, category, product description. The ID on the forum, the name of the forum, the name of the author, etc.

4.3 Category Browsing

Figure 5 shows that the user can further browse the commodity by clicking on the product to view other information about the product, and can also browse selectively by category, some information that you need to view. Figure 6 shows some forum information.

![Figure 4. Threat frame.](image)

![Figure 5. Market data information.](image)
5. CONCLUSIONS

In this paper, a hidden threat intelligence framework is proposed. The main functions include: product search, browsing and downloading, classification, and social network analysis. To help identify the dark network seller. There are many promising directions for future expansion. First, you can implement commodity comparisons with other features, including more robust similarity functions or enhanced similarity visualization. In addition, it includes real-time collection and analysis from the forum through continuous crawling, analysis and updates.

REFERENCES

2. Victor Benjamin, Hsinchun Chen. “Securing Cyberspace: Identifying Key Actors in Hacker Communities” Department of Management Information Systems The University of Arizona Tucson, AZ 85721, USA.
5. Eric Nunes, Ahmad Diab, Andrew Gunn, Ericsson Marin , Vineet Mishra, Vivin Paliath, John Robertson, Jana Shakarian, Amanda Thart, Paulo Shakarian “Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence” Arizona State University Tempe, AZ 85281, USA.
6. Meenakshi Sharma, Department Of Computer Science, Sscet Badhani, Pathankot, India, Supriya” Deep Webpage Classification and Extraction (DWCE)” Department Of Computer Science, Sscet Badhani, Pathankot, India.