

Deep Web - Some Lessons to Learn

Goran Popović

PMG PRODUKTI d.o.o., Croatia

Ivan Strugar

Faculty of Economics & Business – Zagreb, University of Zagreb, Croatia

Abstract

Internet is a huge resource of information, but only the small portion of it is visible to regular users. More than 90 per cent of the web is hidden and to be able to see it, users need to use specific software like Tor. Currently there are more than 2.5 million daily users using Tor to surf the “hidden web”. Ever since Bitcoin and other virtual currencies became popular, more and more users started using Tor for illegal activities. These activities include exchanging illegal goods such as drugs, weapons, illegal data, and stolen credit cards information. The aim of this paper is to analyse the content of anonymous marketplace in one short period of time (72 hours) according to different products offered, scale of performed transactions, origin of the seller, and customer evaluation of the seller. For the needs of this paper, specialised crawling software was developed and as a result, 35 gigabytes of data was downloaded from one of the many black markets used by Tor users. The study shows that in that period, more than 5000 different products worth 1.8 million dollars were listed on the market. The research includes 2500 transactions each worth on average 51 dollars. With the advances in the technology and a growing number of Internet users, further growth of illegal activities on the “dark web” can be expected. It is interesting to analyse how the transactions are closed on the “dark web” since there are no legal institutions or any legal frame for business. These transactions are based on trust, confidence and customer evaluation on social networks.

Keywords: Web, Deep Web, e-commerce

JEL classification: L86, O33

Introduction

Deep web is specific subset of the net, and it is referred to the Web content that is not indexed by search engines and therefore invisible to majority of the people. “The dark Web is a part of the deep Web that has been intentionally hidden and is inaccessible through standard Web browsers” (Chertoff, Simon 2015). Tor is the most popular software and network provides “Dark Web” users near-complete anonymity and ability to safely run server under pseudonym. It has offered creating hidden services since 2004 by encrypting data packets and sending them through several network nodes, called onion routers. Anonymous environment can be attractive in case of rising Internet censorship, in publication dissident's news and sensitive, confident or controversial documents or topics for some societies. Platforms and services which promote freedom of speech are Wikileaks, GlobalLeaks and Strongbox. But in the same time anonymous service attract users for illegal activities such as identity theft, controlled substance trading (Silk Road) and illegal financial transactions.

Traditional search engine like Google can see only about 0.03 percent of the information that is available (Bergman 2001). The rest consists of data that cannot be located with a simple Google search and is called “Deep Web”. There are large

number of "Deep Web" researches related with different aspects of research, but the largest part of them is focused on different illegal activities, as international terrorism (Jihadist) phenomena (Chen Hsinchun, 2012) or focused to security issues (Chertoff, Simon 2015). Large list of different studies related with deep Web is Research at University of Luxembourg analyse Tor hidden services (Biryukov, et. al. 2013). They collected 40 thousands unique onion addresses classifying their content. The conclusion is that the number of hidden services related to illegal activities was equal to the number of resources focusing on legal activities. Conclusion of one another research proves that the reasons for Silk Road use accord with broader online commerce trends (range, quality, convenience, ratings), its appeal to drug purchasers is moderated by country-specific deterrents and market characteristics. (Barratt, Ferris, Winstock, 2014)

The aim of this research is to analyse the content of anonymous marketplace in one short period of time according to different products offered, scale of performed transactions, origin of the seller, and customer evaluation of the seller. It is interesting to analyse how the transactions are closed on the "dark web" since there are no legal institutions or any legal frame for business. These transactions are based on trust, confidence and social networks.

How do we collect and analyse the data in detail is explained in methodology. In results we categorize collected data and analyse supply and demand. The different aspects of the Deep Web commerce are discussed in discussion part of the paper.

Methodology

For the purpose of this paper, specialised software, referred in some papers as a robot or crawler was developed. A crawler is a program that automatically collects Web pages to create a local index and/or a local collection of web pages (Junghoo C., et. al., 1999). Crawler's responsibility regarding this paper was to download as much data as possible from one of the black markets hidden in a Tor network within 72 hours. Crawled marketplace is called Nucleus and it is available only through Tor network.

Developed crawler is designed to work completely autonomously and is used for both fetching and processing data. It is able to handle slow and breaking connections which happen often in Tor network. A typical job consists of fetching URLs retrieved from the job queue, saving them and processing them. Processing starts with extracting URLs from the fetched webpages which are then saved to the job queue and later fetched. Crawler is also able to process and extract useful data for this research in a streaming and batch mode. Streaming process is happening after extracting URLs. This processed data is saved to the data warehouse. Webpage resources such as html, images and scripts were saved to the distributed storage and later used for batch processing.

Data extracted from the black market consisted of information about offered products (product name, image, description, sellers username, shipping regions, shipping pricing and product prices in dollars), information about sellers (username, region/country, products listed, date of last activity) and feedbacks which were used to analyse transactions (buyers anonymised users, grade, feedback, quantity and date). Illegal markets are not displaying list of successful transactions and therefore there is no obvious way to extract and analyse transactions. To extract transactions, the concept of all C2C markets has to be exploited. Marketplaces use feedbacks as a way of fighting against spammers and cheaters. For the feedback system to be effective it has to contain details about the product, including name,

price, quantity, details about seller, buyer and time of transaction, all of which can be downloaded and used to analyse transactions on the markets in the Deep Web.

Results

Data was collected in the period from 23 to 26 March 2015. In that period, more than 35 gigabytes of raw http data was collected, including html pages, images, scripts and stylesheets.

Table 1

Products per categories

Category	Number of products
Drugs	6152
Guides & Tutorials	970
Digital Goods	586
Service	313
Fraud related	288
Counterfeits	171
Erotica	132
Drug Paraphernalia	113
Weapons	69
Electronics	62
Lab Supplies	17
Miscellaneous	13
Jewellery	10
Total	8896

Source: Data downloaded from Nucleus web(2015)

Table 1 shows products listed on the Nucleus web market based on their categories. When crawling started, there were 8896 products listed of which the majority were drugs, illegal guides and tutorials. As it can be seen, on illegal markets, even weapons, lab supplies and illegal services can be acquired.

Table 2

Supply analysis

Name	Value
Products	5211
Market value	1.814.522\$
Listing price - mean	348\$
Listing price – st. dev.	1.880\$
Sellers	934
Listings per seller - mean	5,58
Listings per seller – st. dev.	10,14

Source: Data downloaded from Nucleus web(2015)

Table 2 show statistics regarding downloaded products. Crawler fetched and processed 5211 items worth 1.814.522 dollars. Due to the needs of anonymity, the available quantity of each item is not displayed. Therefore, the true market value is a lot bigger and can be only estimated. On average each product was worth 348\$.

These results should be observed with caution because of a large standard deviation which is a result of 872 free listings and 334 listings with price larger than 999 dollars. Alongside with drugs sold for personal usage, there are also listings containing larger amounts of illegal drugs intended for wholesale greatly influencing the average listing price. On the market, 934 sellers listed their products and services, each seller having on average 5,58 active listings.

Table 3
Demand analysis

Name	Value
Transactions	2457
Market value	126.214\$
Transaction value - mean	51,37\$
Transaction value – st. dev.	166\$
Sellers	321
Transactions per seller - mean	7,65\$
Transactions per seller – st. dev.	12,84\$

Source: Data downloaded from Nucleus web (2015)

Table 3 show statistics regarding sold products retrieved from feedback system. In total, crawler fetched 2457 transactions worth 126.214 dollars. Average transaction was worth 51,32 dollars but that number should be taken with caution because of high standard deviation which is a result of 498 free transactions and 20 transactions worth more than 500 dollars. Biggest transaction was worth 4916 dollars.

Discussion

To create successful illegal market on the web the programmers needed virtual currencies. They copied already proven concept of C2C market places, changed payment methods from credit cards and payments systems like PayPal to virtual currencies and planted these markets in a safe and anonymous environment of Tor hidden network. When physical items are bought they are transferred by postage.

Sellers even provide the origin of the product and multiple postage options allowing users to decide. This information can be later used to analyse geographic location of sellers and can also be used for future analysing of buyer's location.

Virtual currencies brought anonymity to the most crucial part of the exchange, payment. Using virtual currencies had many positive effects. By allowing users to pay with virtual currencies, they've attracted a lot of users from all around the world especially those with a large amount of Bitcoins with only a few choices to spend them.

Marketplaces like E-bay often ignore smaller market filled with users who would love to use service but are unable to pay because of the restrictions of their countries. A great example is Croatia, whose citizens until recent times were unable to buy and sell on E-bay because PayPal services could not make payment to bank or debit card accounts in Croatia (Panian Z., et. al., 2011). Virtual currencies also have no fees, or have a very small fee which allows them to do micro transactions. These transactions are impossible on classic marketplaces because of large fixed transaction fee. Instead of acting as a mediator they've created a transparent system where buyers and sellers grade each other and both handle the responsibility

for risk. By only connecting sellers and customers, the speed and ease of exchange is growing.

Conclusion

Anonymous environment can be attractive in case of rising Internet censorship, in publication dissident's news and sensitive, confident or controversial documents or topics for some societies. But in the same time anonymous service attract users for illegal activities, which are based on trust, confidence and social networks.

Drug dealers and illegal suppliers are moving to Internet and it is only a matter of time before the majority of illegal traffic moves to the "Deep Web". It has become a normal thing for people to buy illegal goods; pay for them from the safety of their homes and in several days gets them on their address.

To be able to stay online and hidden from police, programmers had to upgrade current concepts of marketplaces and some their ideas could be taken and implemented in current public marketplaces. Bitcoin allows marketplaces to be a lot faster, cheaper and accessible to more people. Using currencies lowers seller's expenses and allows users from all around the world to buy and sell goods. They also enable the micro transactions. Those transactions allow buying and selling goods for really small prices which were previously impossible because of fixed fees.

Social networks are the base for building trust and relationship between the parties in environment where no official customer protection or goods return policy exists.

There are number of reasons why "Deep Web" and anonymity should be supported but in the same time there is lot of reasons why illegal activities should be carefully monitored.

Limitations of the study lays in fact that for a more comprehensive analysis of illegal trade on hidden web, scale of the research should be extended. Nucleus, crawled market place is only one of the many illegal marketplaces and future researches should include several other markets even the specialised ones like markets for weapons. The crawling process was quite short, as it lasted for only three days and for a more comprehensive analysis, crawling should last a longer period. For the future research it will be interesting to relate and compare for one country or region official e-commerce results with Deep Web commerce.

References

1. Barratt, M. J., Ferris, J. A., Winstock A. R. (2014), "Use of Silk Road, the online drug marketplace, in the United Kingdom, Australia and the United States", Article first published online: 12 FEB 2014, DOI: 10.1111/add.12470, Vol. 109 No. 5, pp. 774-783, available at: <http://onlinelibrary.wiley.com/doi/10.1111/add.12470/abstract;jsessionid=6A2CF2050C78A36AA657BFB24F5D6C35.f01f03> (accessed May 1st 2015)
2. Bergman M. K, (2001), "White Paper: The Deep Web: Surfacing Hidden Value", available at: <http://quod.lib.umich.edu/cgi/t/text/idx/j/jep/3336451.0007.104/%E2%80%93white-paper-the-deep-web-surfacing-hidden-value?rgn=main;view=fulltext> (accessed May 1st 2015)
3. Biryukov A., Pustogarov I., and Weinmann R. P. (2013), "Content and popularity analysis of Tor hidden services" July 29, 2013, University of Luxembourg, available at: <http://cryptome.org/2013/09/tor-analysis-hidden-services.pdf> (accessed May 7th 2015)
4. Chertoff, M., Simon T. (2015), "The Impact of the Dark Web on Internet Governance and Cyber Security", Chatham House, Centre for International Governance Innovation and the Royal Institute for International Affairs, 2015, available at:

https://www.cigionline.org/sites/default/files/gcig_paper_no6.pdf (accessed May 5th 2015)

5. Junghoo, C., Garcia-Molina, H. (1999), „The Evolution of the Web and Implications for an Incremental Crawler“, Department of Computer Science Stanford, CA 94305.
6. Panian, Z., Korizma, B. (2011), „Electronic Auction Market in Croatia“, World Academy of Science, Engineering and Technology, Vol. 5, pp. 11-22.

About the authors

Goran Popović is currently employed as a programmer at the PMG Produkti d.o.o. in Varaždin, Croatia. He received his bachelor's degree in economics (major Business Economics) in 2013 and graduated (major Managerial Informatics) in 2014 at the Faculty of Economics & Business – Zagreb where he is currently enrolled in a doctoral program. His interests include cloud infrastructure, big data, data mining and distributed systems. The author can be contacted at goranpopovic@gmail.com

Ivan Strugar is a Professor at the Department of Informatics at the Faculty of Economics & Business. He graduated at the Faculty of Economics & Business – Zagreb, where he received his Ph.D. degree in Social Sciences, Humanities and Theology in the field of Economics in 1997. He is lecturer at the Faculty of Economics teaching information technology courses at undergraduate, graduate, post - graduate and doctoral level at the various Universities in Croatia, Slovenia and Bosnia and Herzegovina. His interests include business information systems application in accounting, banking and tourism, cloud computing, collaboration and document management systems, Web, intranet and e-learning development and application. Author can be contacted at istrugar@efzg.hr