Mining the Dark Web

Drugs and fake ids

Andres Baravalle, Mauro Sanchez Lopez, Sin Wee Lee
School of Architecture, Computing and Engineering
University of East London
London, United Kingdom
{a.baravalle, u1340677, s.w.lee}@uel.ac.uk

Abstract — **In the last years, governmental bodies have been futilely trying to fight against dark web marketplaces. Shortly after the closing of "The Silk Road" by the FBI and Europol in 2013, new successors have been established. Through the combination of cryptocurrencies and nonstandard communication protocols and tools, agents can anonymously trade in a marketplace for illegal items without leaving any record. This paper presents a research carried out to gain insights on the products and services sold within one of the larger marketplaces for drugs, fake ids and weapons on the Internet, Agora. Our work sheds a light on the nature of the market; there is a clear preponderance of drugs, which accounts for nearly 80% of the total items on sale. The ready availability of counterfeit documents, while they make up for a much smaller percentage of the market, raises worries. Finally, the role of organized crime within Agora is discussed and presented.**

Keywords—etl processing, data wrangling, authomation, dark web, security, tor, bitcoins

I. Introduction

In the current era of the internet of things, a Dark Web has been developing under the surface.

The Dark Web usually relies on the combination of crypto currencies such as bitcoins and anonymized access as the foundations in creating a market place for dealing illegal drugs, weapons and other illegal contrabands.

In recent years, the Dark Web has been in extreme scrutiny and investigations from legal authorities around the globe.

The clamp down of these have come with mixed success. Whilst many of these websites have been successfully shut down; others resurface or re-migrate soon after. The most significant case is the Silk Road, which was shut down in October 2013 and then resurface as Silk Road 2.0 in a month later. Then, soon after the shut down of Silk Road 2.0 in 2014, many of the vendors just simple migrate to other marketplaces such as Evolution and Agora.

Although a number of prominent web sites in the Dark Web have been shut down thanks to intervention from police forces, the main characteristics of these Dark Web sites are still unknown and we didn't know enough of how these websites operate.

The term Dark Web was first introduced in the year 2000s and has been in use both in the media and in academia since then. Most research done are focusing in identifying extremist views and identifying extremist groups. Only recently there interest has started to surge in others areas, including drug trafficking [1], looking at alias classification and authorship attribution.

In this paper the focus will be on the analysis of the Dark Web marketplaces as a phenomenon and on an in-depth analysis of Agora, a marketplace for drugs, counterfeit documents and ids that we monitored for our research.

The structure of this paper is as follows: sections 2 describes the context of our research and the related work. Then, in section 3, we present a detailed descriptions about the data collection procedures and the experiment setup. The experimental results are discussed in section 4. Finally, conclusions are presented about the research.

II. Under the surface: Deep Web and Dark Web

Research on the "size" of the Internet shows that its size (in term of hosts) has reached 1.05 billion hosts in early 2016 (http://ftp.isc.org/www/survey/reports/current/); about 3.5 billion users have now access to the Internet.

Due to the constant growth of the contents available and users, a specific family of service providers has arisen in order to try to index the information and allow users to reach the content within the internet through these search engines [2].

This collection of resources indexed by search engines and made publicly available falls under the category of "surface web". However, regardless of the effort done by these search engines in order to index more content and in the most useful way to the final user, some of the contents available on the internet are yet not indexed.

On the top layer of the web, we have the indexed, surface web in which we navigate daily. Under the surface web, we have the deep web, content underneath the surface and not indexed by the search engines. Bergman [3] estimated the deep web to be 400 to 550 times larger than the content on the surface.

Under the deep web, we can find the dark web, the back alley of the Internet [4]. While “deep web” and “dark web” are sometimes used interchangeably (especially by the media, but also by other authors), will use the above distinction for this paper.

A. *Dark web*

We can define the Dark Web as "a collection of websites that are publicly visible, but hide the IP addresses of the servers that run them" [5]. These web sites can be visited by users, but it is hard to identify where they are hosted and who hosts them, as they are hidden behind encryption protocols – typically either Tor (The Onion Routing) or I2P (Invisible Internet Project).

While the expression "dark web" as we intend it today is relatively recent, the concepts around dark web have been under investigation since the early 2000s.

The concept for example comes up in several works by Chen, H. et el. around a "Terrorist Knowledge Portal" (cited in [6]; the papers are not indexed, but the slides for a presentation are available [7]), linking it to "the use of the Internet by terrorist and extremist groups".

B. *Timeline of the Dark Web marketplaces*

It was through a forum post (https://bitcointalk.org/index.php?topic=3984.msg57080) on March 2011 that the first large dark web hosted market place was publicly announced. The Silk Road was presented as an anonymous online market, still under development in hope to receive feedback from the bitcoin community.

While starting with drugs, the staff was aiming to offer weapons and other products that may be difficult to find on the surface web market places.

With the combined efforts of FBI and Interpol, the site was seized on February 2013, ending with the arrest of Ross William Ulbricht, who was charged with engaging in a money laundering and narcotics trafficking conspiracy as well as computer hacking [8].

Other dark web sites started to fill the market niche left by The Silk Road. Some of these include “Evolution”, “Hydra” and the “Silk Road 2.0". However, combined efforts of Europol and FBI would seize the vast majority of them during “operation Onymous” on November 2014 [9].

Yet some of this platforms remained after operation Onymous: “Evolution” and “Agora”. However, a few months later, Evolution was allegedly subject of an inside scam where 130.000 Bitcoins were allegedly stolen, following by the closure of their services (https://www.reddit.com/r/DarkNetMarkets/comments/2zeuxo/complaintwarning_evolution_admins_exit_scamming/).

With no competition left, Agora became the “king of the Dark Net” [10]. Agora has changed host and domain name several times in an attempt to avoid cyber-crime law enforcers over its almost two years of existence. agorahooawayyfoe.onion, one of the instances of this marketplace is the subject of this study.

As of today, the site has been shut down by their staff since September 2015, with the motivation cited as a number of vulnerabilities exposed regarding the anonymization process of the TOR network (https://www.reddit.com/r/AgMarketplace/comments/3idznd/agora_to_pause_operations/).

C. *Agora*

Agora is (or possibly was) a popular “dark web” marketplace. Agora does have characteristics that are similar to other black market operations [11] and it’s based on behaviour that doesn’t comply with an organisational set of rules.

It is fairly obvious that entities that fail to comply with the set of rules do so in order to obtain some sort of benefit.

Generally speaking, Agora was selling both products and services, with a minimal set of rules. At the time of our research the only items that couldn’t be sold were body parts, and the only service that was forbidden to sell was assassination.

It is possible that the very limited set of rules might have been some sort of market positioning statement - in practice the key areas of business within the market where drugs, fake ids and weapons (but only for a limited time).

As for all black market operations, operations on Agora were not taxed, neither directly nor indirectly, and Agora offered sellers the possibility for sellers to place products that could not be typically sold legally.

The key aspects of Agora are largely similar to the ones of other illegal operations: protection of the identity for the members, exchange of money, illicit profits.

III. Experimental Environment

A. *Architecture*

The first proof-of-concept for our spider was developed with a few lines of code to simulate a human authentication on the market.

The on-line marketplace was first of all invite-only - so access to the market place required some digging for an invite and several sessions on the web site were required to be able to replicate human-like sessions to proceed with the data collection.

The application used for collection has been built on a classic LAMP stack for data collection – and a variety of languages for data analysis.

The miner was developed using command line PHP (and the cURL library) and an object oriented approach, using MySQL as a backend.

The analysis of the data has been carried with several tools - including Weka and ad-hoc Java and Python scripts.

Libraries such as Pandas, Numpy, NLTK and MatPlotLib have been used for the analysis, integrated within a Jupyter notebook, enabling the team to explore and interact with the code and results of the analysis.

Both the MySQL database and the Jupyter notebook are being hosted using a popular cloud hosting service allowing the team to collaborate remotely in real time.

Finally, in order to promote a homogeneous development environment between all members of the team and the hosted service, we have integrated all our software dependencies on a Docker container running Ubuntu Linux with an installation of Jupyter.

All the code for both the data collection spider and the data analysis tools can be found in Github, respectively at https://github.com/zaharovs/collector and https://github.com/mauromsl/dark_web_datamining

B. Data Collection

Navigation on Agora is anonymous as Agora is hosted on the Tor network; from our analysis, Agora also used a combination of classic discretionary access control techniques for authentication [12], combined with techniques to discourage web scraping.

Protection of their business model in general, and specifically assets is something that Agora's team very much considered, but the techniques used by the team were neither advanced nor seemed to show awareness of the developments of the last few years.

There is extensive research on techniques to discourage web scraping; the most common ones include:

1. Turing tests
2. User-agent identification
3. Throttling of HTTPD requests
4. Obfuscation
5. Data tainting
6. Injecting markers
7. Network traffic analysis

Within the Agora team, there was clear effort in trying to minimize the impact of deep web extraction. Turing tests and user-agent identification were implemented at the time we started our work, and network traffic analysis was most likely introduced later on while we were working on the data collection.

Turing tests to differentiate real users from software robots are common in the industry [13], and Agora did use optical character recognition CAPTCHAS. This was the single protection that slowed the most our spider; while the spider was running largely automatically, the CAPTCHA challenge was answered manually (and had to be repeated several times per day due as sessions had limited persistence).

Agora also implemented some sort of user-agent identification techniques, but they were fairly basic (user agent header and referrer) and did not implement more advanced HTTP header sniffing (e.g. looking for headers that a normal browser would send, as Accept-Encoding, but that a spider may not send).

Agora didn't implement either more advanced techniques as obfuscation (e.g. generation of text with JavaScript), data tainting [14] or even account audits, which overall facilitated our spidering.

Injecting markers is normally used to identify intellectual property that is taken and re-published; given the business operation that Agora was running it is unlikely that it was too relevant.

Routine site maintenance and self imposed throttling (we were trying to avoid detection) did mean that our spidering was not as fast as we would have liked – and we managed to have only a limited number of data collection points.

In time, the web site administrators might have realized that data mining was in progress as extra layers of protection were added: geolocation, session expiration and session management were added after we started the monitoring and before the closure.

The rationale for session expiration and session management checks is self-evident: sessions were frequently expired automatically, which meant having to restart the Tor connection, re-authenticate and deal again with the Turing test.

When it comes to geolocation checks (and bans) we can just hypothesize that Agora was trying to block TOR connections from exit points where traffic level was too high. This was countered again by restarting the Tor connection.

Towards the end of our data collection, the site was typically unavailable every day for hours, which again required changes in the code and in the way we were collecting data.

That said, Agora's network traffic analysis was fairly weak; Agora was an invite only web site, and invites were not easy to find. Two accounts only were used to mine the data and no particular actions have been attempted to audit (and/or block) user connections.

The main data collected was the list of products (including both product description and images) and vendors in Agora.

IV. Results and Analysis

A. Overview

A total of 30.680 records were collected during our study, starting from 22nd July 2015 until 3rd September 2015.

When looking at the supply side of the market, most of the products categories are drugs such as cocaine, MDMA or heroin. The "other" category seems to be the largest of them in terms of number of listings (this doesn't mean it is in terms of value).

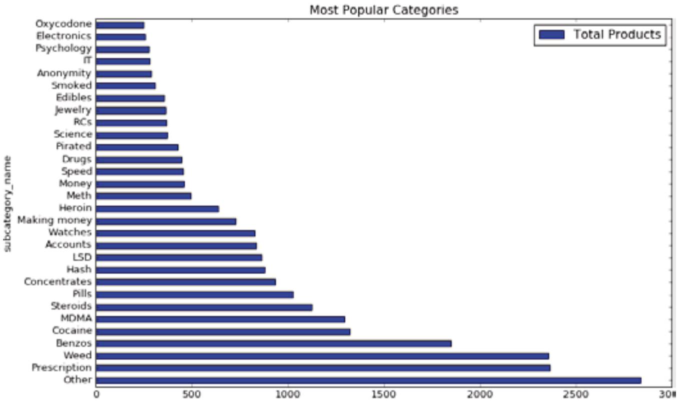

Figure 1 Most popular product categories

Grouping drug-related items together it is evident the preponderance of drugs within the marketplace – nearly 80% of the size of the market in terms of products on sale.

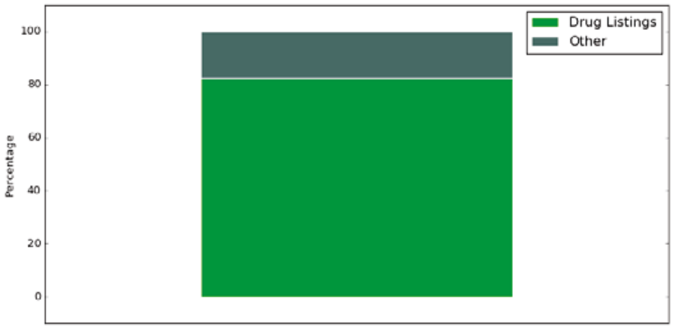

Figure 2 Drug listings vs other items

B. Geographical distribution

Looking at our data, US leads the product supply – with UK a distant second.

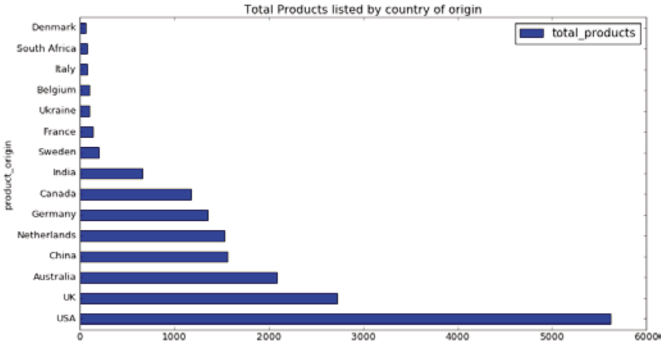

Figure 3 Geographical distribution

When analyzing the data more in depth, we can see how the drugs market is dominated by suppliers from US and UK, while sellers from China lives up to the stereotype and focus on watches and clothing (most likely counterfeit products).

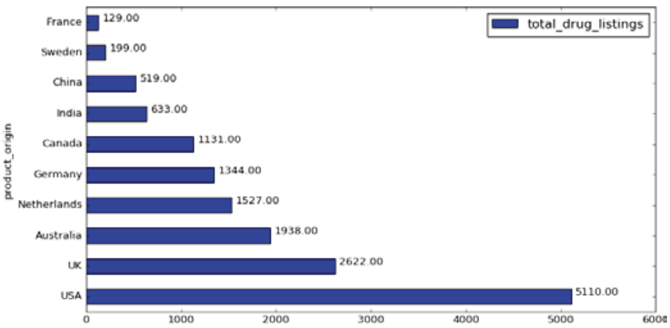

Figure 4 Geographical distribution for drugs

C. Drugs

The total market for drugs within Agora was massive, in all ways – number of products on sale, number of sellers operating in the country and total size of the market.

Looking at the data, the highest number of sellers are concentrated within USA, UK and Australia – while the top countries by market size are Germany, USA and Netherlands respectively.

Market size (bitcoins)	Drug number selling in country	Selling country	Number of vendors
17094.63	1177	Germany	74
13279.65	3985	USA	388
6500.61	1477	Netherlands	72
4204.84	1652	Australia	138
2583.053	1811	UK	137
2446.51	1013	Canada	71
1516.203	236	China	13
797.98	13	Argentina	2
785.681	5	Peru	1
507.39	99	Belgium	6

Table 1 Top 10 countries by market size

Looking at the inventory value for the sellers on Agora, is it immediately evident that it's not small criminality. We are talking about entities as RADICALRX with over 10 million dollars of product on sale on Agora over the time of our study. This is hardly teenagers in basements – the scale is the one of organized crime.

Products on sale (bitcoins)	Alias	Countries of operation	Categories
13028.72	RADICALRX	Germany, Sweden	Hydromorphone, Oxycodone, Fentanyl, Meth
2729.37	HonestCocaine	USA	Cocaine
1958.49	p3nd8s	USA	Meth, Prescription, Heroin, Cocaine, DMT, LSD, MDMA, Fentanyl

1786.10	Clandestine_Pharmaceutics	USA	2C
1291.57	drugbrothers	Peru, Germany, Netherlands	Cocaine, Speed, MDMA, Weed
1249.78	DutchWholesale	Netherlands, Australia	Cocaine, MDMA, Shake/trim, Ketamine
1003.51	dutchelite	Netherlands	Cocaine, MDMA
917.61	zunidog	Germany, Netherlands	Prescription, Heroin, Ketamine, Cocaine, Oxycodone, Speed, Pills, MDMA, Weed, Fentanyl
910.48	alchemycd	China	Concentrates, MDMA, Synthetics
747.75	SanaDi&#118;ersion	Argentina	Cocaine

Table 2 Top 10 sellers, sorted by value of products on sale

D. Counterfeit documents

The total market size of counterfeit documents on Agora is 3747.85 BitCoins. Documents of any type could be sourced on Agora, making it a viable option for international terrorist groups and criminals in general.

Differently from the drugs market, the counterfeit documents market seems to be more concentrated – with less vendor operating in the niche.

The top vendor, alias plasticA, has put on sale ids for 3433.61 bitcoins, about 1.8 million British pounds.

When looking at the data on passports, we can see how there are two main types of passports on sale: physical documents and scans.

Physical documents are counterfeit passports on sale; the vendors claim that they will be accepted by the authorities as real for being exact copies with all the safety features being replicated. Scans are just scanned copies of real passports – possibly to be used for identity fraud.

During our research, 84 scans/photos of passports were on sale, and 12 physical passports.

A UK passport can be bought for as cheap as £752. Scanned passports are available for as litte as £7, and can be bought in bulks, with a wide range of countries to choose from.

Counterfeit identity cards can be bought for as cheap as £142 for an European id card and even cheaper for US state id cards, with prices ranging between £25 and £92.

The total number of listings for passports and counterfeit identity cards is 65, but some vendors claim to be able to produce any number of them, personalising all the details of the counterfeit document.

Driving licenses are also on sale; the vast majority of them are US driving licenses. Prices for those licenses range between £51 pounds up to £300. Prices for European driving license were slightly more expensive, up to £419 but more impressively, in one of the listings, the vendor claimed that the license sold would be registered officially.

Market size (bitcoins)	**Counterfeit ids on sale**	**Counterfeit ID's seller name**	**Selling countries**	**Postage countries**
3433.6050338799996	5	plasticA	USA	Worldwide,
4.3575266699999995	11	i&#79;racle	USA	
4.17117371	5	TemplatesAndFakes	USA	USA,
3.59959234	2	harveynorman	Australia	
2.56928505	3	Cherymoya	USA	
2.46805955	8	DrRopata	New Zealand	,worldwide
2.2997546	13	xOneStopShoPx	USA	
1.42178443	2	Keira666	Croatia,Spain	
1.1994291099999999	3	dmvnationsupply	USA	Global,
0.74509266	2	stoeprand	Netherlands	
0.34257134	1	ProfessorPlastic	USA	USA
0.23979993	1	elangaba	USA	

Figure 5 Top 10 counterfeit document vendors, sorted by market size

E. Organized crime

As part of our research, we wanted to gain some insight on the size of the vendors operating on the site.

We wanted to try to understand first of all how concentrated was the supply within the different vendors, and then if there were any existing patterns that would manifest that the supply was operated by well-coordinated organizations instead of individuals.

By looking at the total value of the listings for each of the vendors, we hoped to understand how concentrated the supply is.

The plot below displays just the largest 25 vendors and it can be already be seen how the suppliers market value is not evenly distributed. In modern economic studies, this phenomenon often occurs in oligopolistic markets where the large majority of this supply is concentrated into a small number of providers. This phenomenon is commonly referred as "kinked supply curve" [15].

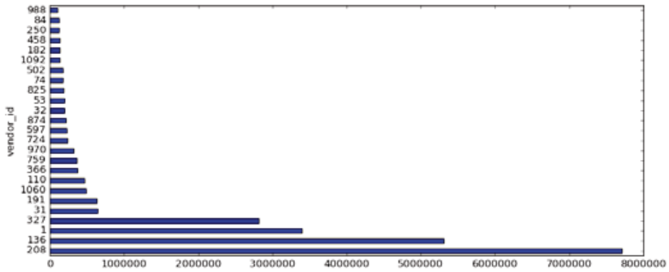

Figure 6 Top 25 vendors by market value

Looking more in depth at the data, we can see how over 90% of the market is dominated by the largest 10% vendors.

This becomes even more prevalent within the drugs market (which copes 80% of the total market). Having such concentration of the value within a small vendors is a factor that points us to the fact that the agents behind these markets are not small independent vendors but rather larger criminal organizations. In regards to the counterfeit passports and national ids market, the dominance is not as wide as in the drugs market but still a respectable 40% of the market.

It's worth nothing also that in some cases the quantity of merchandise on sale on one individual transaction was such that would suggest the presence of organized crime. When looking at the hashish category, for example, the mean amount on sale is 47g, with a median of 10g, but with some sellers selling up to 1 kg at the time.

Finally, our research indicates that there was some use of sockpuppets – identities used for deception – within Agora. We decided to try to identify sockpuppets by looking at vendors consistently using exactly the same images for their products – on the grounds that vendors using the same images were likely to be connected.

The analaysis did show a number of sockpuppets – but the amount was fairly limited – with no more than 2 sockpuppets per vendor.

V. Conclusions

When we started this work, we wanted to try to lift the curtain on dark web markets. We didn't have any idea of its nature, of its size, of the role that organized crime seems to be playing within the market.

There are serious issues that emerge from our research.

Over 170691.12 BitCoins (about £26 million) of merchandise where on sale on the period under examination. Over 30,000 products were on sale; 1233 sellers participated in the market, spread across 20 countries, with the largest number located in the US and UK.

Drugs, ids and also weapons were readily available in a trans-national marketplace, just one click away and anonymously.

When it comes to counterfeit documents, it is relevant to mention that any EU ID card would allow the potential buyer to travel through any country in the EU, open bank accounts and in general create a new identity for himself/herself.

While we didn't manage to collect any data on weapons as they were removed from the market early on, it was originally possible to buy also weapons – and to have them delivered through multiple packages, disassembled.

Black market services are working very cautiously, implementing security measures and hacker avoidance updates regularly. They are largely dominated by organized crime, and they keep resurfacing regardless of the efforts made to shut them down.

AcknowledgmentS

The authors would like to thank Germans Zaharovs, who worked on this project during his research internship, and the University of East London, who funded the internship itself. Special thanks must also go Daniel Blackman, for the support and encouragement during the research, and for the help in the dissemination of the results.

1 References

[1] M. Splitters, F. Klaver, G. Koot and M. Van Staalduinen, "Authorship Analysis on Dark Marketplace Forums," in *roceeding of Intelligence and Security Informatics Conference (EISIC)*, Manchester, 2015.

[2] K. Bharat and A. Broder , "A technique for measuring the relative size and overlap of public Web search engines," *Computer Networks and ISDN Systems,* vol. 30, no. 1-7, pp. 379-388, 1998.

[3] M. Bergman, "White Paper: The Deep Web: Surfacing Hidden Value," *The Journal of Electronic,* vol. 7, no. 1, 2001.

[4] M. Eddy, "Inside the Dark Web," 04 02 2015. [Online]. Available: http://uk.pcmag.com/security/39461/guide/inside-the-dark-web. [Accessed 17 06 2016].

[5] M. Egan, "What is the Dark Web? How to access the Dark Web. What's the difference between the Dark Web and the Deep Web?," 2016 06 28. [Online]. Available: http://www.pcadvisor.co.uk/how-to/internet/what-is-dark-web-how-access-dark-web-deep-joc-beautfiulpeople-3593569/. [Accessed 17 06 2016].

[6] H. Oman, "Security Technology Progress: The 37th IEEE-AESS Carnahan Conference, Taiwan," *IEEE Aerospace and Electronic Systems Magazine,* vol. 19, no. 2, pp. 35-40, 2004.

[7] H. Chen, "The Terrorism Knowledge Portal: Advanced Methodologies for Collecting and Analyzing Information from the 'Dark Web' and Terrorism Research Resources," 08 2003. [Online]. Available: http://www.slideshare.net/suyu22/the-terrorism-knowledge-portal-advanced-methodologies-for-collecting-and-analyzing-information-from-the-dark-web-and-terrorism-research-resources. [Accessed 17 06 2016].

[8] A. Greenberg , "End Of The Silk Road: FBI Says It's Busted The Web's Biggest Anonymous Drug Black Market," 2 10 2013. [Online]. Available: http://www.forbes.com/sites/andygreenberg/2013/10/02/end-of-the-silk-road-fbi-busts-the-webs-biggest-anonymous-drug-black-market. [Accessed 16 6 2016].

[9] A. Greenberg, "Global Web Crackdown Arrests 17, Seizes Hundreds Of Dark Net Domains," 11 07 2014. [Online]. Available: https://www.wired.com/2014/11/operation-onymous-dark-web-arrests/. [Accessed 16 6 2016].

[10] A. Greenberg, "Drug Market 'Agora' Replaces the Silk Road as King of the Dark Net," 18 11 2015. [Online].

Available: http://www.wired.com/2014/09/agora-bigger-than-silk-road. [Accessed 17 06 2016].

[11] E. L. Feige , "Reflections on the Meaning and Measurement of Unobserved Economies: What Do We Really Know About the 'Shadow Economy'," *Journal of Tax Administration ,* vol. 2, no. 6, 2016.

[12] R. S. Sandhu and P. Samarati, "Access control: principle and practice," *IEEE Communications Magazine,* vol. 32, no. 9, 1994.

[13] A. Kolupaev and J. Ogijenko, "CAPTCHAs: Humans vs. Bots," *IEEE Security & Privacy,* vol. 6, no. 1, pp. 68-70, 2008.

[14] V. Bhagwan and T. Grandison, "Deactivation of Unwelcomed Deep Web Extraction Services through Random," in *Web Services, 2009. ICWS 2009. IEEE International Conference on*, Los Angeles, CA, 2009.

[15] C. Efroymson, "The Kinked Oligopoly Curve Reconsidered," *The Quarterly Journal of Economics,* vol. 69, no. 1, p. 119, 1995.