Deepweb and Cybercrime

It's Not All About TOR

Vincenzo Ciancaglini,
Marco Balduzzi,
Max Goncharov, and
Robert McArdle

Forward-Looking Threat
Research Team

Contents

Abstract

The term "deepweb" is used to denote a class of content on the Internet which, for different technical reasons, is not indexed by search engines. Among the different strategies in place to bypass search engine crawlers, the most efficient for malicious actors are so-called "darknets." Darknets refer to a class of networks that aim to guarantee anonymous and untraceable access to Web content and anonymity for a site.

While deepweb has often been uniquely associated with The Onion Router (TOR), in this paper, we introduce several other networks that guarantee anonymous and untraceable access— the most renowned darknets (i.e., TOR, I2P, and Freenet) and alternative top-level domains (TLDs), also called "rogue TLDs." We analyzed how malicious actors use these networks to exchange goods and examined the marketplaces available in the deepweb, along with the goods offered. Due to a large variety of goods available in these marketplaces, we focused on those that sparked the most interest from cybercriminals and compared their prices with the same class of merchandise found in traditional Internet underground forums, mostly Russian. Finally, we introduced some of the techniques that researchers can use to more proactively monitor these so-called hidden parts of the Internet.

Introduction

The term "deepweb" has been introduced over the past few years to denote Internet content that search engines do not reach, particularly:

- **Dynamic web pages:** Pages dynamically generated on the HTTP request.

- **Blocked sites:** Sites that explicitly prohibit a crawler to go and retrieve their content by using, CAPTCHAs, pragma no-cache HTTP headers, or ROBOTS.TXT entries, for instance.

- **Unlinked sites:** Pages not linked to any other page, preventing a Web crawler from potentially reaching them.

- **Private sites:** Pages that require registration and log-in/password authentication.

- **Non-HTML/Contextual/Scripted content:** Content encoded in a different format, accessed via Javascript or Flash, or are context dependent (i.e., a specific IP range or browsing history entry).

- **Limited-access networks:** Content on sites that are not accessible from the public Internet infrastructure.

The last point has two identified types of limitation that constitute two independent categories, namely:

- Sites with domain names registered on an alternative Domain Name System (DNS) root (i.e., rogue TLDs). These are sites whose hostnames have been registered using a registrar independent from the Internet Corporation for Assigned Names and Numbers (ICANN).[1]

  Standard domain names follow a naming hierarchy coordinated by the ICANN, which is responsible for defining standard TLDs (e.g., .com, .edu, .gov, etc.) and coordinates domain name assignment. Consequently, standard DNSs are synchronized according to the name hierarchy defined by the ICANN and can resolve all domain names assigned within the ICANN space.

  One can, however, connect to specific DNS servers that manage additional namespaces not recognized by the ICANN, allowing the registration of domain names that do not follow ICANN rules such as a nonstandard TLD. While resolving these domain names requires the use of specific DNS servers, their use can present some advantages in the form of an easier and, sometimes, untraceable way to register new domain names.

- **Darknets and alternative routing infrastructures:** Sites hosted on an infrastructure that require a specific software to reach the content provider. Examples of such systems are TOR's hidden services or sites hosted on the Invisible Internet Project (I2P) network. These sites are generally identified as well by a nonstandard domain name that requires using the same software to be resolved to a routable endpoint.

It is worth noticing that, while as of now crawling of such sites does not happen, it is not due to a technical limitation. Crawlers could resolve an alternative DNS name by connecting to one of the specific DNS servers publicly available and the TOR and I2P software act as SOCKS proxy, making it possible for a crawler to access the said content. The only noticeable leakage of information from darknets to a search engine happens thanks to gateway services such as tor2web, which offers a clearnet domain to directly access content hosted on hidden services.[2]

---

1   http://www.icann.org/
2   http://www.tor2web.org/

Overview of Existing Deepweb Networks

To date, three main networks are used to grant anonymity on both the client and server side—TOR, I2P, and Freenet.

Note that the latter two have not yet reached the same adoption that TOR has but present desirable technical features that could lead them to become viable alternatives in the near future (e.g., should the TOR network become too unreliable for users).[3]

TOR

The TOR network was originally developed by the U.S. Naval Research Laboratory and first introduced in 2002. It allows for anonymous communications by exploiting a network of volunteer nodes (i.e., more than 3,000 to date) responsible for routing encrypted requests so that the traffic can be concealed from network surveillance tools.[4]

To take advantage of the TOR network, a user needs to install software that acts as a SOCKS proxy. The TOR software conceals communications to a server on the Internet by selecting a number of random relay nodes to form a circuit. Before entering the network, every request is recursively encrypted using the public key of each selected node. Then, by bouncing from one relay to the next, every layer of encryption is lifted off for the next relay, until an exit node is reached and the unencrypted request can then travel to its destination.

Adopting this mult-layered encryption mechanism has the following advantages:

- A server that receives a request coming from the TOR network will see it as being issued by the last node in the TOR circuit (i.e., the exit node) but there is no straightforward way to trace a request back to its origin.

- Every node within the circuit only knows the previous and next hop for a request but cannot decipher the content nor find out its final destination.

- The only TOR node that can view the unencrypted request is the exit node but even this does not know the origin of the request, only the previous hop in the circuit.

---

3   As already witnessed in the FreedomHosting seizure case: http://nakedsecurity.sophos.com/2013/08/05/freedom-hosting-arrest-and-takedown-linked-to-tor-privacy-compromise/.
4   https://metrics.torproject.org/

In recent versions of the TOR protocol, a new functionality has been introduced to allow entire sites to be hosted on TOR nodes, making them untraceable. The services that run within the TOR network are known as "hidden services." The approach works by storing the contact information to reach a hidden service in the form of a rendezvous node that will act as an intermediary and an encryption key in a Distributed Hash Table (DHT). The DHT acts as a form of distributed DNS, resolving a .onion hostname into the contact information necessary to establish a connection to the hidden service. In this case, both the client and the server IP addresses are concealed to any third party that is trying to analyze or block the traffic. Their real locations are even concealed from each other.

I2P

I2P was designed as an anonymous peer-to-peer (P2P) distributed communication layer that can run any traditional Internet service. It has been developed since 2003 as an evolution of the Freenet network, which aims to allow for several services to run on top besides HTTP. While TOR was initially conceived to enable anonymity when connecting to an Internet service (i.e., WWW) and was only later extended to general hidden services, I2P's exclusive goal is to provide a way for users to host services (e.g., IRC, Web, mail, and bittorrent) in a stealthy way.

TOR's main principle is creating circuits (i.e., encrypted paths through a random set of nodes to reach either an exit node that serves as a proxy or to a rendezvous point that acts as an intermediary to communicate with a hidden service). I2P, on the other hand, introduces virtual tunnels. Every node in an I2P network is a router. It creates and maintains a pool of inbound and outbound virtual paths. For example, if node A wants to send a message to node B, it routes the message to one of its outbound tunnels together with the information needed to reach one of B's inbound tunnels.

The information about inbound tunnels is stored, much like in TOR, in a DHT that serves as a decentralized network database. This way, no central point of failure exists. Every communication is encrypted using multiple layers: point-to-point encryption between sender and receiver, transport encryption between routers in the network, and end-to-end encryption in tunnels. Note that, while TOR uses an encryption scheme called "onion routing," the encrypted routing used in I2P is known as "garlic routing."

The hidden sites hosted in the I2P network, also called "eepsites," such as torrent trackers or anonymous email servers can be identified by either a hash value or a domain name featuring the .i2p TLD.

Freenet

Freenet has been around since 2000 and can be considered the predecessor of I2P. Unlike I2P though, it implements a pure DHT in the form of an unstructured overlay network. This it means that each node is responsible for a subset of the resources available in the network and serves them collaboratively when it receives a request. Furthermore, nodes maintain a list of neighboring nodes, usually known and trusted neighbors, to increase security. This is also known as the "small world principle." Nodes and data are identified by a key, usually represented with a hash value. When looking for a resource, a request will travel across all of a node's neighbors in order of preference (i.e., to the nodes whose key is closest to the resource key).

Because of the adopted approach, Freenet is more suitable to serving static content such as static sites and does not cope well with dynamically generated web pages or other forms of Internet services (e.g., IRC, mail, etc).

Compared with I2P and TOR, Freenet offers less flexibility in terms of hosted services, being limited to serving only static content without, for example, server-side scripting. The range of services that can be implemented on top is smaller. This, however, does not mean that Freenet cannot be a suitable platform to host simple marketplaces or exchange information related to malicious activities.

Alternative Domain Roots

Alternative domain roots, also known as "rogue TLDs," refer to a class of networks that use DNS entities that are not under the control of the ICANN, in contrast with traditional .dot/.net/.org domains. Domains registered within a rogue TLD require the use of dedicated name servers. On the other hand, depending on the institution that runs the DNS root, registering a domain name may be less problematic for a malicious actor, as in the case of .bit domains, for which domain registration follows a P2P paradigm. In short, this means that any new domain name registration rather than being handled by a central authority is autonomously propagated in a P2P network made of all the .bit DNS servers until every server is aware of the newly registered domain.

While alternative DNS domains do not offer particular forms of anonymity and TOR does, they present some clear advantages for malicious actors, notably a certain protection against domain sinkholing, better flexibility domain management, and, to date, the possibility of "escaping" search engine crawling. While it would be technically possible for a search engine to crawl a site on an alternative DNS (e.g., simply using one of its DNS servers), it does not normally happen and, if it does, the results are not shown to users that do not have the right DNS servers configured.

At the time of writing, we identified the following alternative DNS roots as active:

- **Namecoin:** Responsible for the .bit TLD. It is based on a P2P infrastructure working on the same principle as bitcoins. A client willing to access the domains can either run a dedicated DNS client or refer to one of the gateway DNS servers available online. More information on the .bit TLD will be featured in a Trend Micro research paper in the coming months.[5]

- **Cesidian root:** An alternative DNS run by a private Italian citizen that offers TLDs such as .cw, .ispsp, .5w, and .6w.[6] It was born to support Mr. Tallini's para-political vision who is also the self-proclaimed governor of the United Micronations Multioceanic Arcipelago (UMMOA).[7] Folkloristic aspects aside, the cesidian root counts a network of more than 30 DNS servers all over the world, running on IPv4 and IPv6.

- **Namespace.us:** This organization offers 482 alternative TLDs such as .academy, .big, and .manifesto. It has been on the market since 1996, when it was founded to extend the limited number (at that time) of available TLDs and offered a faster process for domain registration as well as other domain-related services.[8] Having failed in the late 1990s to have its TLDs integrated into the DNS root zone, it remains an alternative provider of domain names to date, offering its own DNS servers that resolve both its TLDs as well as ICANNs official ones.[9]

---

5   http://dot-bit.org/Main_Page
6   http://cesidianroot.net/
7   http://www.foxnews.com/tech/2012/03/02/cesidian-root-bizarre-peek-at-world-wide-weird/
8   http://www.namspace.us; http://swhois.net/
9   https://namespace.us/about.php

- **OpenNIC:** This project consists of a network of DNS servers run by hobbists and volunteers that aim to offer a DNS infrastructure that is neutral, independent from governments and organizations, democratic, and free for everyone.[10] Anyone can offer a computer to be used as a tier-2 DNS server with the sole condition of respecting a strict policy concerning its security, performance, and anonymization.[11] Besides offering a network of DNS servers for the standard ICANN DNS root, this DNS provider also offers an alternative namespace of 14 TLDs and supports the four alternative TLDs of NewNations, an organization that provides domain roots for certain political entities such as Tibetan or Kurdish people.[12]

Cybercrime in the TOR Network

This section describes the malicious commercial activities identified in the deepweb, particularly marketplaces and goods cybercriminals exchange.

Despite the fact that all of the aforementioned systems have the potential to support illegal trades of every sort, to date, the only network that seems to have gained some traction for underground marketplaces is TOR.

The reason behind this may be linked to the fact that TOR is proportionally more mature and more developed than the competition and has been endorsed by organizations such as the Electronic Frontier Foundation as the first choice among anti-censorship tools, putting it under the spotlight recently.

TOR Marketplace Overview

The TOR network features two major marketplaces, along with two others, which are no longer active but worth mentioning. It also has plenty of small sites that offer individual services.

Each marketplace features a fully operative e-commerce solution with different sections, shopping carts, checkout management, and payment and escrow services. They all support crypto-currencies such as bitcoins and litecoins.

---

10 http://www.opennicproject.org/
11 http://www.opennicproject.org/opennic-policies/dns-operation-policy/
12 http://www.new-nations.net/

Silk Road is probably the most notorious of all, having been extensively featured by the press over the last couple of years.[13] It catalogs goods into different sections (see Figure 1) and provides seller ratings and guides for buyers on how to securely purchase items. It is, so far, the only marketplace that has been extensively analyzed by researchers. In fact, a recent paper from the Carnegie-Mellon University shows that in 2012, it had an estimated income of US$22 million and its number of users doubled in under six months.[14] As it turned out, however, this was massively lower than the actual number.
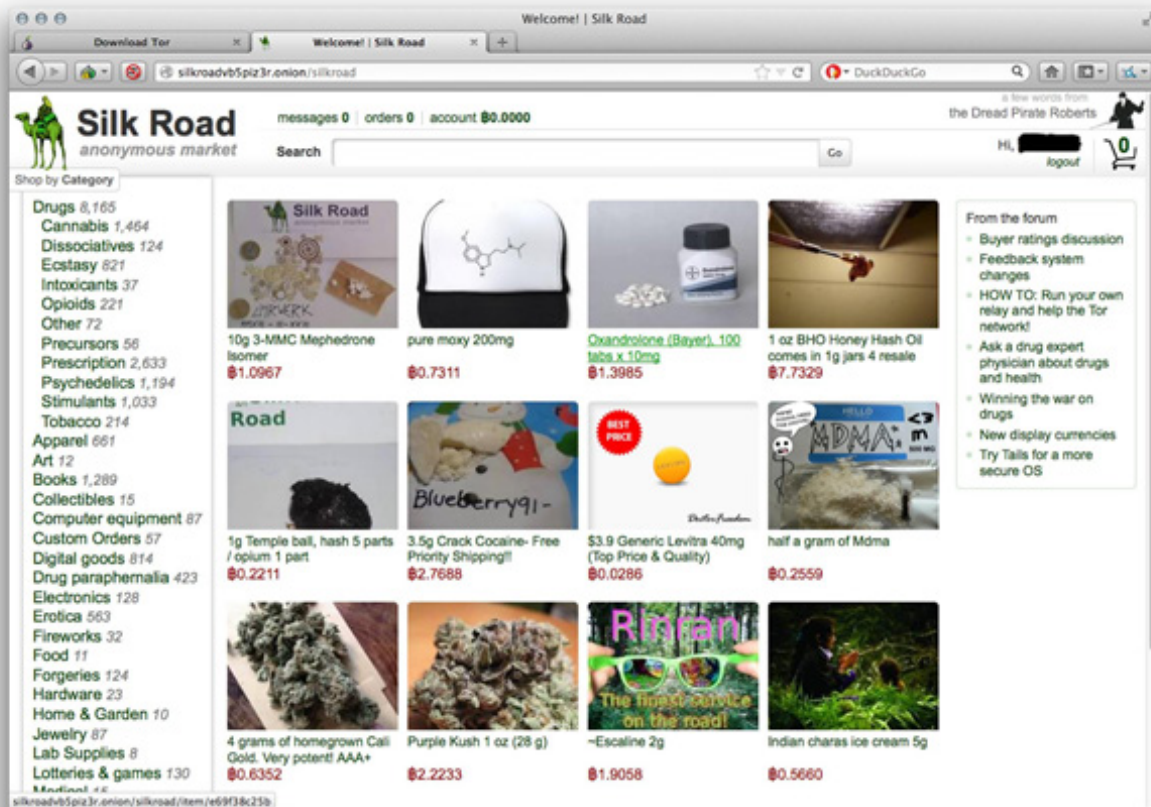
Figure 1: Silk Road main page

13 https://silkroadvb5piz3r.onion; http://www.theatlantic.com/technology/archive/2011/06/libertarian-dream-a-site-where-you-buy-drugs-with-digital-dollars/239776/; http://www.gq-magazine.co.uk/comment/articles/2013-02/07/silk-road-online-drugs-guns-black-market/viewall; http://www.forbes.com/sites/andygreenberg/2012/08/06/black-market-drug-site-silk-road-booming-22-million-in-annual-mostly-illegal-sales/
14 http://arxiv.org/pdf/1207.7139v1.pdf

As of October 2, 2013, Silk Road is no longer active. Ross William Ulbricht who stands accused of being "Dread Pirate Roberts," the owner and main administrator of the marketplace, was arrested by the Federal Bureau of Investigation (FBI) at a public library in San Francisco on Tuesday, October 1. The complaint filed against Mr.Ulbricht gives several more details about the marketplace's operations and accuses him of narcotics trafficking as well as computer hacking and money laundering conspiracy.[15] Mr.Ulbricht is also being accused of soliciting the murder-for-hire of another Silk Road user who was threatening to release the identities of thousands of the site's users.

The FBI said that it also seized approximately US$3.6 million worth of Bitcoins. As all Bitcoin transactions are public, we can simply observe this transaction in the Bitcoin blockchain.[16] Bitcoin is a highly volatile currency and, as such, its value dropped in light of this takedown but it will most likely swiftly recover.

According to the FBI, in the two-and-a-half years of its existence, the site generated sales amounting to over 9.5 million Bitcoins and collected commissions on those sales of over 600,000 Bitcoins. At the time the complaint was filed, this equated to approximately US$1.2 billion in sales and US$80 million in commission.

---

15 http://www.scribd.com/doc/172768269/Ulbricht-Criminal-Complaint
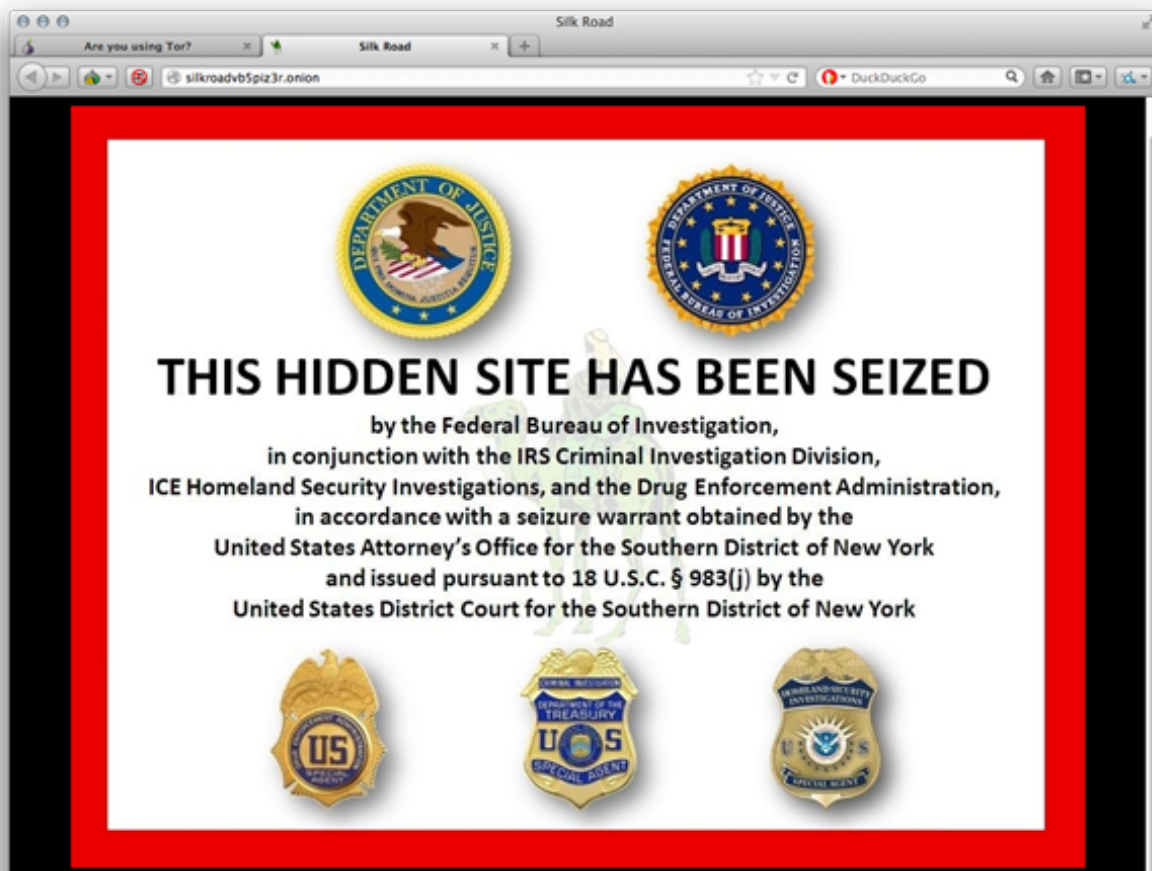16 http://blockchain.info/address/1F1tAaz5x1HUXrCNLbtMDqcw6o5GNn4xqX

Figure 2: Silk Road main page after the takedown

Atlantis had, until recently, emerged as a fierce competitor of Silk Road, offering the same features but with a more convenient seller commission and support for multiple currencies.[17] This forum was shut down as well on September 20, 2012.[18] According to a message posted on the group's Facebook page, this shutdown was due to security reasons outside of its control:

> "We have some terrible news. Regrettably, it has come time for Atlantis to close its doors. Due to security reasons outside of our control, we have no choice but to cease operation of the Atlantis Market marketplace. Believe us when we say we wouldn't be doing this if it weren't 100% necessary.
>
> Due to the urgency, we are allowing all users to withdrawal all their coins for one week before the site and forum are shut down permanently. Please remove all of your coins, these will not be recoverable after one week from now. Anything remaining in your accounts will be donated to a drug-related charity of our choosing."

17 https://atlantisrky4es5q.onion/
18 https://www.facebook.com/AtlantisMarket/posts/421945931244529

We do not exactly know why Atlantis was taken offline but what is clear is that these deepweb marketplaces are coming under increased pressure from authorities.

With these two heavyweights gone, more attention will focus on the likes of Black Market Reloaded (see Figure 3), which features a goods distribution less focused on drugs (i.e., but nonetheless very prominent) and more centered on digital goods and services; and Sheep Marketplace (see Figure 4), which, despite offering a much lower number and variety of goods, is the only one that has a preview site showing the goods for sale but not allowing any transaction, accessible from outside the TOR network and indexed by search engines.[19]
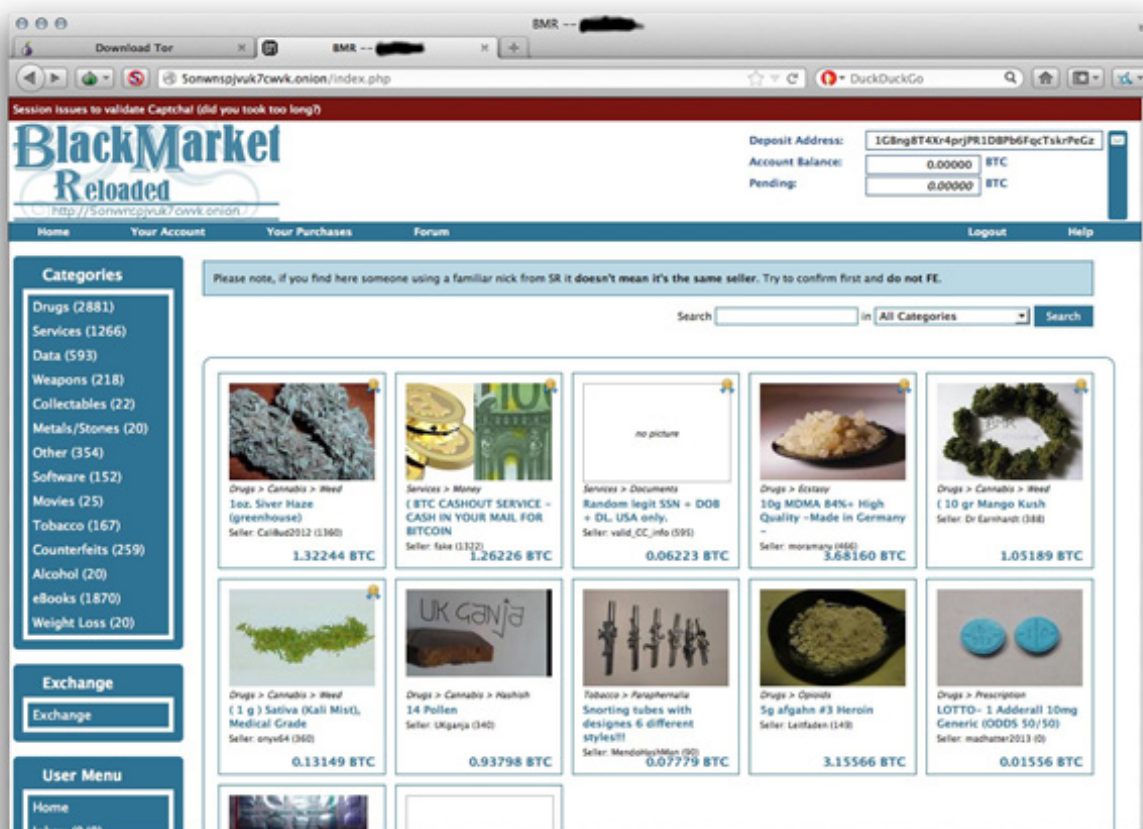

Figure 3: Black Market Reloaded main page

---

19 https://5onwnspjvuk7cwvk.onion; https://sheep5u64fi457aw.onion; http://sheepmarketplace.com
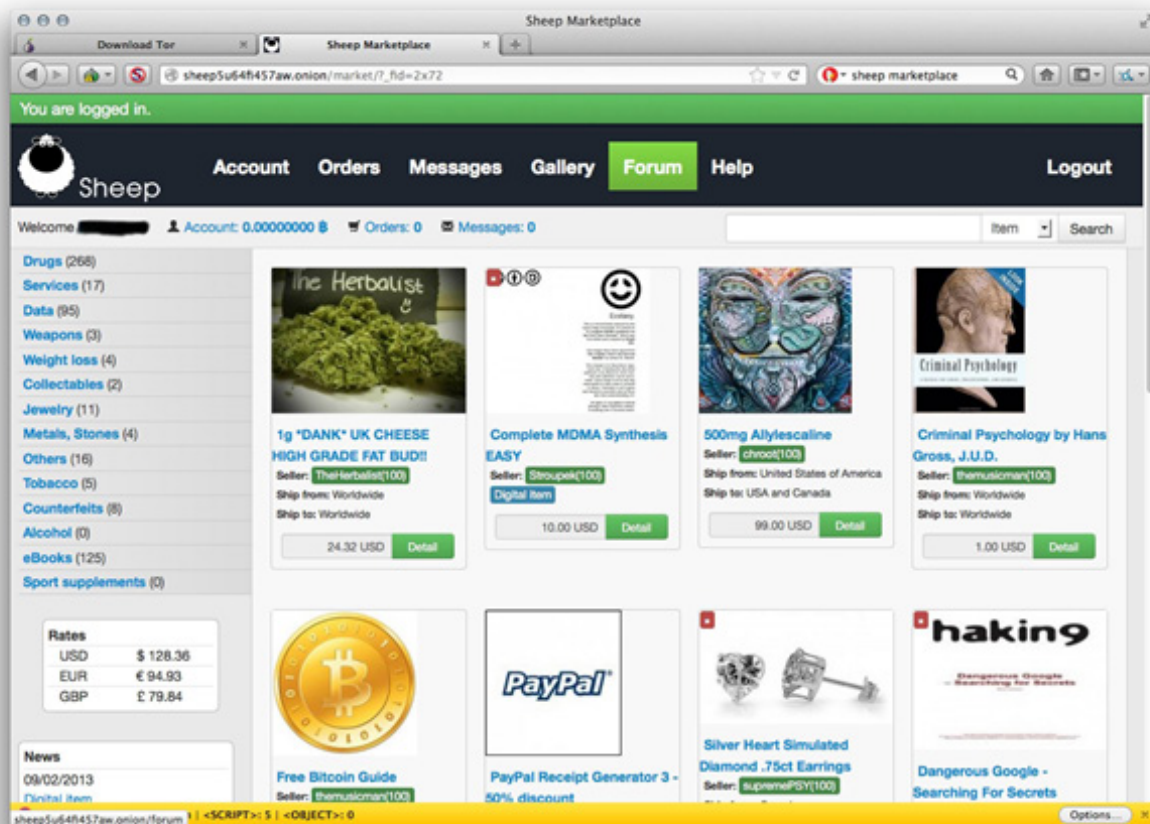
Figure 4: Sheep Marketplace main page

TOR Private Offerings

Besides the aforementioned major marketplaces, we identified two categories of sites that allow anonymous trading. The first category has underground message boards (e.g., Underground Market Boards 2.0) where people can post and read generic classifieds regarding any sort of good or service.[20] The rest have privately maintained sites that offer specific types of goods. Some of these consist of a mere presentation page with prices and contact information for anonymous orders and inquiries while others provide a full order and payment management system to automate orders.

While the range of goods offered in these sites is fairly vast and spans pretty much over every form of item suitable for illegal activities (e.g., drugs, guns, hired assassins, etc.), we will only focus on those related to cybercrime since the rest have been already covered in previous research.

20 http://nb5df7xeas3zl3sf.onion/

Table 1 summarizes some of the goods found on these sites with their prices. For credit card-related prices, unless otherwise specified, the reported prices are per unit for a fully functional card provided with all the data required to do transactions (e.g., credit card holder name, expiration date, authorization code, etc.).[21]

Table 1: Prices of Different Types of Goods				
Site name	Address	Type of good	Cost	Normalized Cost (US$)
CloneCard	http://kpmp444tubeirwan.onion/board/int/src/1368387371226.jpg	EU/US credit cards	1 BTC	US$126
Mister V	http://wd5pbd4odd7jmm46.onion/	EU credit cards	€40–80	US$54–100
CC-Planet Fullz	http://tr36btffdmdmavbi.onion	EU/US credit cards	UA$40	US$54
CC 4 ALL	http://qhkt6cqo2dfs2llt.onion/	EU/US credit cards	€25–35	US$33–47
Cloned credit cards	http://mxdcyv6gjs3tvt5u.onion/products.html	EU/US credit cards	€40	US$54
NSD CC Store	http://4vq45ioqq5cx7u32.onion	EU/US credit cards	US$10	US$10
Carders Planet	http://wihwaoykcdzabadd.onion/	EU/US credit cards	US$60–150	US$60–150
HakPal	http://pcdyurvcdiz66qjo.onion/	PayPal accounts	1 BTC for US$1,000	US$126 for US$1,000
Onion identity	http://abbujjh5vqtq77wg.onion/	Fake IDs/passports	€1,000–1,150 (ID) €2,500–4,000 (passport)	US$1,352–1,555 (ID) US$3,380–5,400 (passport)
U.S. citizenship	http://ayjkg6ombrsahbx2.onion/silkroad/home	U.S. citizenship	US$10,000	US$10,000
U.S. fake driver's licenses	http://en35tuzqmn4lofbk.onion/	Fake U.S. driver's license	US$200	US$200
U.K. passports	http://vfqnd6mieccqyiit.onion/	U.K. passports	£2,500	US$4,000
Guttemberg prints	http://kpmp444tubeirwan.onion/board/int/src/1366833727802.jpg	Counterfeit money	1/2 of the monetary value	1/2 of the monetary value
High-quality Euro replicas	http://y3fpieiezy2sin4a.onion/	Counterfeit Euro banknotes	€500 for 2,500 CEUR €1,000 for 3,000 CEUR €1,900 for 6,000 CEUR	US$676 for 2,500 CEUR US$1,352 for 3,000 CEUR US$2,570 for 6,000 CEUR

21 Prices were normalized in U.S. dollars at October 3, 2013 exchange rates: €1 = US$1.3522; £1 = US$1.6077; 1 BTC = US$126. CEUR = counterfeit Euro, CUSD = counterfeit U.S. dollar.

Table 1: Prices of Different Types of Goods				
Site name	Address	Type of good	Cost	Normalized Cost (US$)
Counterfeit U.S. dollars	http://qkj4drtgvpm7eecl.onion/	Counterfeit U.S. banknotes	US$600 for 2,500 CUSD US$2,000 for 5,000 CUSD	US$600 for 2,500 CUSD US$2,000 for 5,000 CUSD
Rent-a-Hacker	http://2ogmrlfzdthnwkez.onion/	Hacking services	€200–500	US$270–676
TOR Web developer	http://qizriixqwmeq4p5b.onion/	Web development	1 BTC per hour	US$126 per hour

Columns 4 (cost) and 5 (normalized cost in US$) show the prices when the services were found with their original values on the site. We normalized the prices in U.S. dollar amounts to provide comparable figures. Overall, we noticed the following price ranges:

- Credit cards can be purchased from US$10 (NSD CC Store) to US$150 (Carders Planet).

- PayPal accounts go for US$126 (1 BTC) for a US$1,000 account (HackPal).

- Fake documents can cost from US$200 for a fake U.S. driver's license (USA Fake DL) to US$5,400 for a fake U.S. passport (Onion Identity), not to mention US$10,000 for U.S. citizenship (USA Citizenship).

- Rates for counterfeit money depend on the amount purchased and can go from US$0.24 per counterfeit dollar (US$600 to buy 2,500 fake dollars on Counterfeit USD) up to half the value of fake money desired (Guttember prints).

- We also found services for sale from US$126 per hour for a Web developer (TOR Web developer) to US$676 for various hacking services (e.g., botnets, social engineering, account credential stealing, etc.) (Rent-a-Hacker).

Comparison with Russian Underground Marketplaces

This section compares the prices reported above with those for the same goods sold in Russian underground forums.[22] Compared with sites hosted on the TOR network, Russian underground forums are reachable over the Internet (i.e., no need for darknet software nor a rogue TLD DNS server) but their membership can be limited to trusted individuals.

Our analysis revealed that for digital goods (e.g., credit card numbers, PayPal accounts, development services, and malware), underground forums seem to offer a bigger number of goods and transactions.[23] This can be explained by the bigger number of potential users since access does not require the use of additional darknet software but also by the lower level of anonymity the system offers. The increased anonymity afforded by the TOR network, while useful for sellers to avoid getting caught, is somewhat detrimental because it prevents an actor of a commercial transaction to build and maintain a reputation over time. Being hidden behind an anonymous nickname has the drawback of not being able to certify one's reputation, which is essential when dealing with sensible goods. As a practical example, TOR domain names suffer the issue of scam sites (i.e., sites that entirely replicate a marketplace), an issue made worse in the TOR network by lack of control in domain name registration and the "scrambled" nature of .onion domain names.

Table 2 shows sample prices found in underground forums for the same digital goods offered in the TOR network (shown in Table 1). Comparing them with the goods traded in TOR sites, it can be said that:

- Credit cards cost from US$2 (entry 4) to US$120 (entry 2), showing an average price much lower than that found in TOR sites (US$68.8 in TOR sites versus US$23.7 in underground forums).

- More stolen accounts and account information are sold in Russian underground forums than in TOR sites although their prices seem comparable (US$126 for a US$1,000 account in TOR sites versus US$100 for a US$1,000–2,000 account in underground forums).

- Other goods such as fake documents and counterfeit money seem to be lacking in the underground forum scenario or, at least, were much harder to find compared with the TOR space during our investigation.

22 http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-russian-underground-101.pdf
23 Also verified, on the other hand, by the lack of malware offerings seen in deepweb sites.

Table 2: Sample Russian Underground Offerings and Prices		
Address	**Type of Good**	**Cost**
http://forum.prologic.su/index.php?showtopic=7468	U.S. credit cards	US$2.5
http://xek.name/showthread.php?t=10519	U.S. credit cards EU credit cards	US$25–40 US$70–120
http://r00t.in/showthread.php?t=18510'	U.S. credit cards EU credit cards	US$2–3 US$10
http://brute.name/threads/8643/	U.S. credit cards EU credit cards	US$2–3 US$8–9
http://carding.cc/showthread.php?t=6030	Credit card scans	US$14
http://exploit.in/forum/index.php?showtopic=38917'	PayPal accounts	US$2–15
http://carding.cc/showthread.php?t=2548	PayPal accounts	US$10 for US$0–200 account US$20 for US$20–200 account US$50 for US$200–1,000 account US$100 for US$1,000–2,000 account US$150 for US$3,000–4,000 account
http://brute.name/threads/8643/	PayPal accounts	US$200 for US$2,000 account US$500 for US$8,000 account US$1,000 for US$15,000 account
http://www.xaker.name/forvb/showthread.php?t=21284	Russian passports	US$250

Monitoring the Deepweb

The deepweb, in general, and the TOR network, in particular, offer a secure platform for cybercriminals to support a vast amount of illegal activities—from anonymous marketplaces to secure means of communication to an untraceable and difficult to shutdown infrastructure to deploy malware and botnets.[24]

As such, it becomes more and more important for the security industry to be able to track and monitor the activities that take place in darknets, focusing today on TOR networks but possibly extending in the future to other technologies (i.e., I2P, above all).

Due to its design, however, monitoring the darknet proves to be challenging. To tackle it, our future work should focus on the following areas, several of which have already been implemented in our deepweb monitoring systems:

---

24 http://blog.trendmicro.com/trendlabs-security-intelligence/the-mysterious-mevade-malware/

- **Mapping the hidden services directory:** Both TOR and I2P use a domain database built upon a distributed system known as a "DHT." A DHT works by having nodes in the system collaboratively taking responsibility for storing and maintaining a subset of the database, which is in the form of a key-value store. Thanks to this distributed nature of the hidden services domain resolution, it is possible to deploy nodes in the DHT to monitor requests coming from a given domain.[25] By doing this, one can have a partial view over the domains database and inspect ongoing requests. Even though this does not allow one to trace who is trying to access a given service, it does offer a good statistical estimate of what new domains are gaining popularity. In addition, running more such nodes will give one a better statistical view of the overall requests on the network.

- **Customer data monitoring:** A security company could also benefit from analyzing customer Web data to look for connections to nonstandard domains. While this, depending on the level of logging at the customer side, may not prove that fruitful in tracking down connections to darknets, it may provide good insights on activities on sites hosted with rogue TLD domains. It is important to note that this can be carried out without monitoring customers themselves, the destinations of the Web requests (i.e., the darknet domains) should be of most interest, not who is connecting to them.

- **Social site monitoring:** Sites like Pastebin are often used to exchange contact information and addresses for new hidden services and, therefore, need to be kept under constant observation to spot message exchanges containing new deepweb domains.[26]

- **Hidden service monitoring:** Most hidden services to date tend to be highly volatile and go offline very often, maybe to come back online later under a new domain name. It is essential, therefore, to get a snapshot of every new site as soon as it is spotted, for later analysis or to monitor its online activity. When crawling hidden services under the assumption of ongoing malicious activities, one should bear in mind that, while crawling the clear Internet is usually an operation involving the retrieval of every resource related to a site; in the deepweb, this is not recommended due to the chance of automatically downloading illegal materials such as child exploitation materials, the simple possession of which is considered illegal in most countries worldwide.

- **Semantic analysis:** Once the data for a hidden service is retrieved, building a semantic database containing important information about a hidden site can help track future illegal activities on the site and associate them with malicious actors.

---

25 http://donncha.is/2013/05/trawling-tor-hidden-services/
26 http://www.pastebin.com/

- **Marketplace profiling:** Finally, another useful activity to focus on is profiling the transactions made on deepweb marketplaces to gather information about their sellers, users, and the kinds of goods exchanged, building up individual profiles over time.

Related Work

TOR and the deepweb, in general, has been known by the industry and the IT community for several years now. One of the first works that describes the deepweb is "Deep Content."[27] In this work, dated 2001, Bergman tries to quantify the hidden Internet by presenting the 60 known, largest deepweb sites. These contain about 750TB of data, roughly forty times the size of the known surface Web, and appear in a broad array of domains from science to law to images and commerce. The authors estimate the total number of records or documents within this group to be about 85 billion.

Given the remarkable size of the deepweb, Google itself has tried to surf its content, for example, by proposing a system to query for HTML pages and incorporate the results into a search engine index.[28] Others who attempted to crawl the deepweb were He, et al. and Kosmix.[29] Kosmix, in particular, used a new approach to information discovery on the web that significantly differed from a conventional Web search, called "federated search." Finally, in "Trawling for TOR Hidden Services: Detection, Measurement, Deanonymization," the authors exposed flaws both in the design and implementation of TOR's hidden services to measure the popularity of arbitrary hidden services.[30] Their approach allows for measuring the deepweb by de-anonymizing part of its supposed anonymous traffic.

More recently, security researchers are focusing their interest on deepweb as well by trying to uncover malicious use of the hidden Internet. Pierluigi, et al. describes Artemis, a project aimed at collecting open-source intelligence (OSINT) from the deepweb.[31] The authors exerted significant effort to investigate how cybercriminals use the deepweb for illicit activities.[32] For the sake of completeness, the same authors presented the deepweb in a more general form in "Diving in the Deep Web."[33]

---

27 http://grids.ucs.indiana.edu/courses/xinformatics/searchindik/deepwebwhitepaper.pdf
28 http://dl.acm.org/citation.cfm?id=1454163
29 http://www.inf.ufsc.br/~ronaldo/deepWeb/querying/Chang-dwsurvey-cacm07.pdf; http://citeseerx.ist.psu.
    edu/viewdoc/download?doi=10.1.1.151.9143&rep=rep1&type=pdf; http://citeseerx.ist.psu.edu/viewdoc/
    download?doi=10.1.1.152.8111&rep=rep1&type=pdf
30 http://www.ieee-security.org/TC/SP2013/papers/4977a080.pdf
31 http://resources.infosecinstitute.com/project-artemis-osint-activities-on-deep-web/
32 http://blog.trendmicro.com/trendlabs-security-intelligence/the-mysterious-mevade-malware/
33 http://resources.infosecinstitute.com/diving-in-the-deep-web/

Conclusion

The deepweb, particularly darknets such as TOR, represents a viable way for malicious actors to exchange goods, legally or illegally, in an anonymous fashion. In this paper, we conducted an analysis of different networks that guarantee anonymous and untraceable access to deepweb content. Our findings suggest that, at present, the main network that shows commercial activities for cybercriminals is TOR. While the deepweb has proven to be very functional for hosting botnets' command-and-control (C&C) servers and trading merchandise such as drugs and weapons, traditional cybercrime goods (i.e., malware and exploit kits) were less popular.

Sellers suffer from lack of reputation caused by increased anonymity. Somehow, being untraceable presents drawbacks for a seller who cannot easily establish a trust relationship with customers unless the marketplace allows for it.

However, the lack of observable activities in unconventional deepweb networks does not necessarily mean an actual lack of such. In fact, in agreement with the principle inspiring the deepweb, the activities are simply more difficult to spot and observe. Note that since a driving factor for marketplaces is critical mass, it is quite unlikely for them to long for such a high level of stealth unless the consequence, should they be discovered, is sufficiently severe (e.g., child exploitation imagery). In such cases, sites may only come online at specific times, have a brief window of trading, then disappear again, making them more difficult to investigate.

Recent revelations about wide-scale nation-state monitoring of the Internet and recent successful arrests of cybercriminals behind sites hosted in the deepweb are starting to produce other changes. It would not be surprising to see the criminal underbelly becoming more fragmented into alternative darknets or private networks, further complicating the job of investigators. For example, the recent shutdown of the Silk Road marketplace is a big blow for the underground trade of illegal items.

However, the deepweb has the potential to host an increasingly high number of malicious services and activities and, unfortunately, it will not be long before new large marketplaces emerge. As such, security researchers have to remain vigilant and find new ways to spot upcoming malicious services to deal with new phenomena the moment they appear. This is something that Trend Micro is proactively engaging in as part of its global mission to make the world safe for the exchange of digital information.

TREND **MICRO**™

Securing Your Journey to the Cloud

10101 N. De Anza Blvd.
Cupertino, CA 95014

U.S. toll free: 1 +800.228.5651
Phone: 1 +408.257.1500
Fax: 1 +408.257.2003