CIGI

CHATHAM HOUSE
The Royal Institute of International Affairs

Global Commission on Internet Governance

ourinternet.org

PAPER SERIES: NO. 20 — SEPTEMBER 2015

The Tor Dark Net

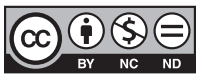
Gareth Owen and Nick Savage

THE TOR DARK NET

Gareth Owen and Nick Savage

CHATHAM HOUSE
The Royal Institute of International Affairs

Published by the Centre for International Governance Innovation and the Royal Institute of International Affairs.



CIGI

67 Erb Street West
Waterloo, Ontario N2L 6C2
Canada
tel +1 519 885 2444 fax +1 519 885 5450
www.cigionline.org

CHATHAM HOUSE
The Royal Institute of International Affairs

10 St James's Square
London, England SW1Y 4LE
United Kingdom
tel +44 (0)20 7957 5700 fax +44 (0)20 7957 5710
www.chathamhouse.org

TABLE OF CONTENTS

ABOUT THE GLOBAL COMMISSION ON INTERNET GOVERNANCE

The Global Commission on Internet Governance was established in January 2014 to articulate and advance a strategic vision for the future of Internet governance. The two-year project conducts and supports independent research on Internet-related dimensions of global public policy, culminating in an official commission report that will articulate concrete policy recommendations for the future of Internet governance. These recommendations will address concerns about the stability, interoperability, security and resilience of the Internet ecosystem.

Launched by two independent global think tanks, the Centre for International Governance Innovation (CIGI) and Chatham House, the Global Commission on Internet Governance will help educate the wider public on the most effective ways to promote Internet access, while simultaneously championing the principles of freedom of expression and the free flow of ideas over the Internet.

The Global Commission on Internet Governance will focus on four key themes:

- enhancing governance legitimacy — including regulatory approaches and standards;
- stimulating economic innovation and growth — including critical Internet resources, infrastructure and competition policy;
- ensuring human rights online — including establishing the principle of technological neutrality for human rights, privacy and free expression; and
- avoiding systemic risk — including establishing norms regarding state conduct, cybercrime cooperation and non-proliferation, confidence-building measures and disarmament issues.

The goal of the Global Commission on Internet Governance is two-fold. First, it will encourage globally inclusive public discussions on the future of Internet governance. Second, through its comprehensive policy-oriented report, and the subsequent promotion of this final report, the Global Commission on Internet Governance will communicate its findings with senior stakeholders at key Internet governance events.

www.ourinternet.org

ABOUT THE AUTHORS

Gareth Owen is a senior lecturer in the School of Computing at the University of Portsmouth. He holds a Ph.D. in computer science and has expertise in distributed computing systems, digital forensics and privacy-enhancing technologies. Before joining the university, he lectured at the universities of Kent and Greenwich in the United Kingdom.

Nick Savage is the head of the School of Computing at the University of Portsmouth. He was previously a principal lecturer in the School of Engineering at the University of Portsmouth, where he taught networking and security. He is a member of Working Group 3 for the European Commission's Network and Information Security Platform and has previously worked on projects funded by the Office of Communications and the Engineering and Physical Research Council. Nick holds a Ph.D. in telecommunications from the University of Portsmouth.

EXECUTIVE SUMMARY

The term "Dark Net" is loosely defined, but most frequently refers to an area of the Internet only accessible by using an encryption tool called The Onion Router (Tor). Tor is a tool aimed at those desiring privacy online, although frequently attracts those with criminal intentions. An innovative feature of Tor is the ability to host websites anonymously and with a degree of impunity — designed to be used by those in repressive regimes who wish to host whistle-blowing or political content.

The study described in this paper collected data on the Tor Dark Net over a period of six months to analyze the type and popularity of the content. Perhaps unsurprisingly, the majority of sites were criminally oriented, with drug marketplaces featuring prominently. Notably, however, it was found that sites hosting child abuse imagery were the most frequently requested.

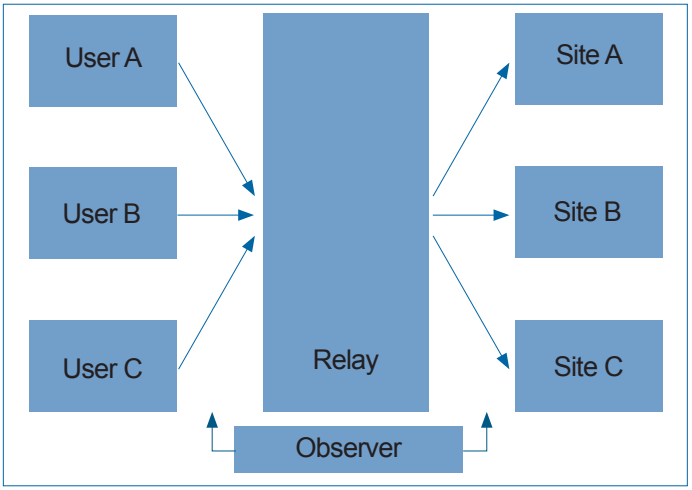
INTRODUCTION

Tor is an open-source tool that aims to provide anonymity and privacy to those using the Internet. It prevents someone who is observing the user from identifying which sites they are visiting and it prevents the sites from identifying the user. Some users value Tor's anonymity because it makes it difficult for governments to censor sites or content that may be hosted elsewhere in the world.

Tor has a critical mass of users, averaging two million per day as of June 2015 (Tor Project 2015), and is thus frequently cited as one of the key tools against government surveillance. Somewhat paradoxically, the Tor Project (the non-profit organization that manages Tor) receives the majority of its funding from the US government.

Tor volunteers run thousands of "relays," a server that any other user can ask to route traffic through. Figure 1 illustrates the simple case of a single relay, with three users asking it to route traffic to three sites. An observer can see traffic entering and leaving the relay, but they cannot determine which user is visiting which site (save for correlation attacks, which will be discussed later) because the traffic is encrypted; however, if the relay operator is malicious, they can trivially (with ease, from a technical standpoint) link the two.

Figure 1: Illustration of Relay-mixing Traffic

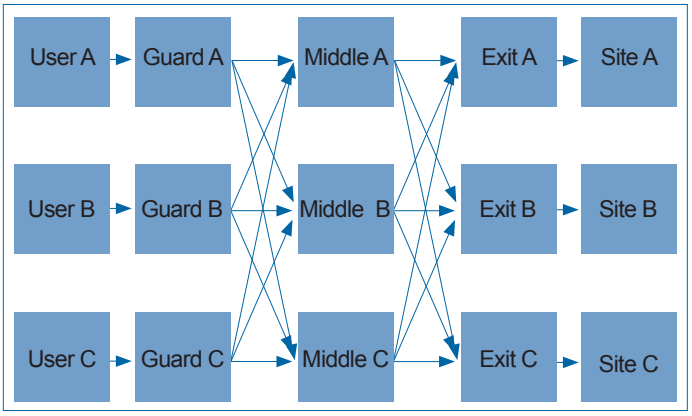


Source: Authors.

When a user visits any sites through a relay, his traffic appears to come from the relay rather than the user's computer. Thus, the user remains anonymous to the site itself.

To defend against a malicious relay operator, the user chooses three relays and chains them together, labelling them as the Guard, Middle and Exit, known as a three-hop circuit (see Figure 2).This raises the bar significantly for an attacker who would then need to control all three relays to be able to link users with certainty to the sites they are visiting. It is the user who chooses the three relays; the attacker is unable to influence his choice/decision. An attacker's only option would thus be to control a significant number of relays in the hope that a user chooses three within that controlled pool — this is thought to be impractical.

Figure 2: Typical Three-hop Tor Circuits



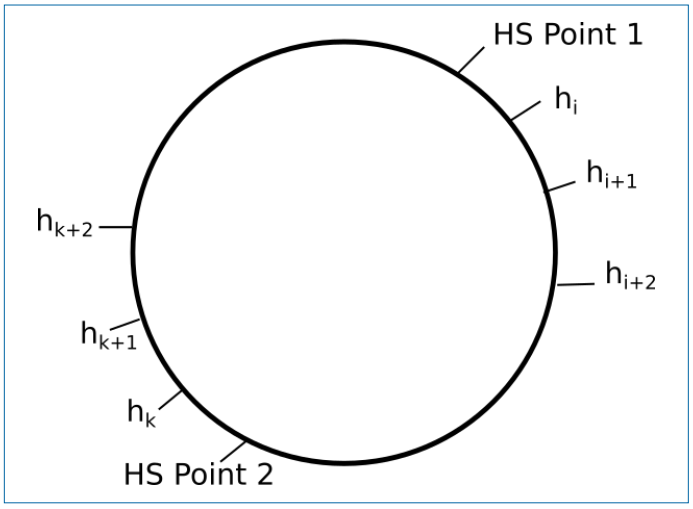
Source: Authors.

HIDDEN SERVICES

While the ability to access the Internet anonymously is valuable in countries where personal freedoms are restricted, it is only one feature of Tor (Jardine, n.d.).The other major feature is "hidden services" (HSes), the ability to host a website (or Internet service) anonymously. In this case, both the visitor and the site are anonymous to each other. Using this feature, political blogs or forums can be hosted in repressive regimes without fear of penalty. As with any technology such as this, it also allows the possibility of criminally oriented material to be hosted with a degree of impunity. The collection of Tor HSes is often referred to as the Dark Net, although there exist other, less popular tools that might also be considered under this umbrella (for example, the Invisible Internet Project, known as I2P).

It is crucial to know how hidden services work to be able to understand the methodology used in measuring activity on the Dark Net. Let's assume Bob is hosting an HS and Alice wishes to visit his site. When Bob first creates his site, he constructs a document detailing "introduction points," or relays within the network that will be able to relay messages to him. He publishes this document in a distributed hash table (DHT), which can be thought of as a database or phone directory distributed across all the relays in the network — that is, no single relay controls or possesses all of the DHT at any one point in time. To create such a database, all the relays in the network are placed onto a circle and ordered according to a unique identifier (see Figure 3, relays are labelled h_i).The HS is then mapped onto the circle at two points. Bob publishes information on his introduction points to the three relays to the right of each of these two locations, so that copies exist on exactly six relays. When publishing the information, he uses a three-hop circuit to remain anonymous to the directory relay.

Figure 3: Tor HS Directory (DHT)



Source: Authors.

The location that Bob publishes to in this directory appears random, and changes every day, but it is possible for Alice to figure out his information. The location changes daily to make it more difficult for one person to control the relays that hold Bob's information.

Alice then calculates which relays on the circle contain Bob's information and builds three-hop circuits to the relays, requesting a copy of his information. By using a three-hop circuit, she remains anonymous to the directory relays. She now has information on Bob's introduction points and relays a message to one of them asking Bob to build a connection to a rendezvous relay that she chooses. The rendezvous relay proceeds to relay messages between Bob and Alice, and since both have connected to the rendezvous relay through three-hop circuits, the relay does not know the identity of either party. The rendezvous relay cannot inspect the traffic because it is encrypted; its service is wholly altruistic.

RELATED WORK

HSes were described in the original Tor Paper (Dingledine, Mathewson and Syverson 2004) and have since undergone several revisions. They are difficult to locate geographically, but they use a DHT (similar to those used in many other distributed Internet applications (Stoica et al. 2001)) to publish descriptors with information on how to connect to them. The Tor DHT is not resistant to Sybil attacks (Douceur 2002), in which one can run many nodes and gain control of a large proportion of the DHT. With that control, one can collect HS descriptors (as described in this paper) and deny service to legitimate users. There exist a number of Sybil-resistant DHT implementations (Lesniewski-Laas and Kaashoek 2010), but as of yet, Tor has not focused significantly on this aspect.

The Tor DHT consists of approximately 3,000 participating relays, and each must have been operating persistently for several days before it can participate in the DHT. Furthermore, one can operate only two relays per Internet Protocol (IP) address to increase the cost of launching a Sybil attack. Researchers at Luxembourg University (see Biryukov, Pustogarov and Weinmann 2013) describe a bug in the Tor core program that allows someone to launch a large number of relays on a single computer and selectively phase any into the network. Tor logs relays' uptime, even if they were not in the network, thus making it possible to launch a number of relays on a single computer. Biryukov, Pusatogarov and Weinmann took advantage of this bug and were able to collect the list of HS addresses in fewer than two days; however, the Tor Project has now fixed this bug. The authors used an automated classification algorithm to classify hidden sites into categories by content type. Their data shows they encountered popular abuse sites but chose to label them as "Adult"— an unfortunate side effect of the classification technique used.

Additionally, they only examined HSes present during a single 24-hour period. Therefore, the general question of the size, content and popularity of the Dark Net remains open. This paper addresses the question by collecting data over a significantly longer period of time and manually classifying sites to achieve greater precision.

STUDY OF HSes

To collect information on the Dark Net, a list of HSes must first be enumerated. By controlling all the relays in the Tor DHT, it is possible to collect a complete list of HSes by recording the descriptors as they have been published. It is then possible to count the number of requests for each descriptor and estimate their relative popularity.

Unfortunately, there are approximately 3,000 relays in the DHT, and even though one can create relays that participate in the DHT, it is impossible to control all of it. There is also a non-negligible cost associated with each participating relay one wishes to run. If one runs a handful of DHT participants, then one can observe a fraction of it at any one point in turn. Bearing in mind that HSes publish to two essentially random points every day, over time one would observe every HS that remained online during the collection period.

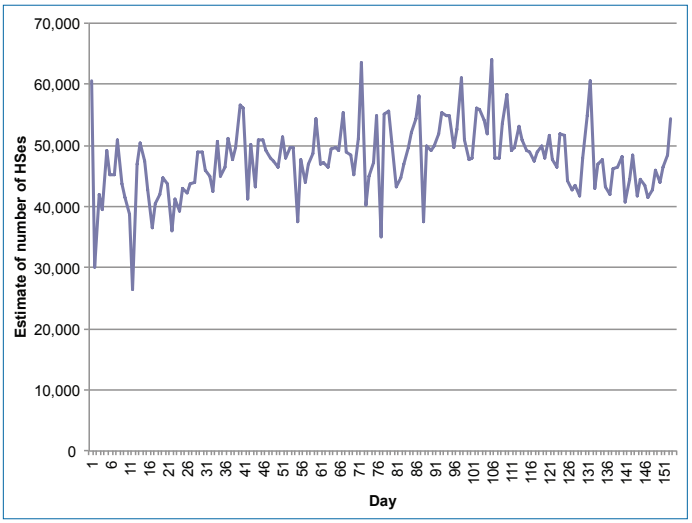
In this study, 40 relays were operated for a period of six months. Each relay recorded a list of published HS descriptors and the number of requests for each. Although only a small proportion of the DHT was observed each day, cumulatively all of the DHT was observed many times throughout the study.

SIZE AND TURNOVER ESTIMATION

Little has been known about the Dark Net to date, although a 2013 study estimated that there were 60,000 HSes at any one time (ibid.); however, this study was based on a single day and it was not known whether this was an outlier. Extrapolating from this paper's data for each day, it is possible to give an estimate for the number of HSes existing on any one day (see Figure 4).

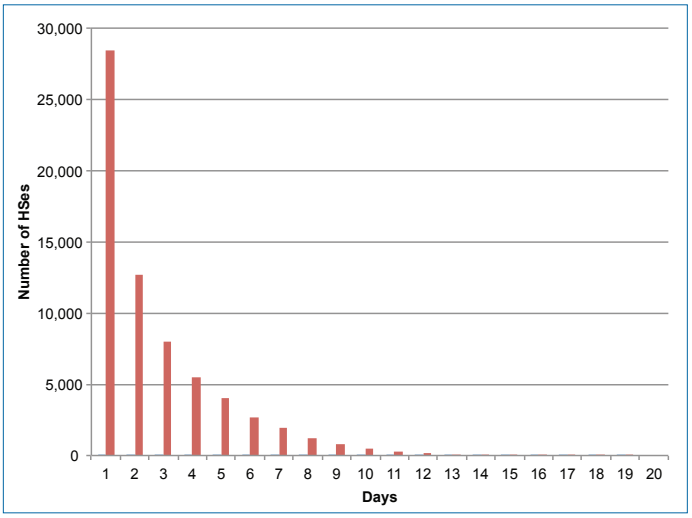
While on first observation it looks as if the number of HSes has high variance from day to day, one must bear in mind that only a small proportion of the DHT is being observed and then extrapolated. This means that errors will be amplified and this accounts for the variance. The long-term average throughout the study was 45,000 active sites and this is likely to be more indicative of the total number of HSes. In total, 80,000 unique HSes were observed during the study, but some only existed for a short period of time.

Figure 4: Estimate of the Number of HSes on Each Day of the Study



Source: Authors.

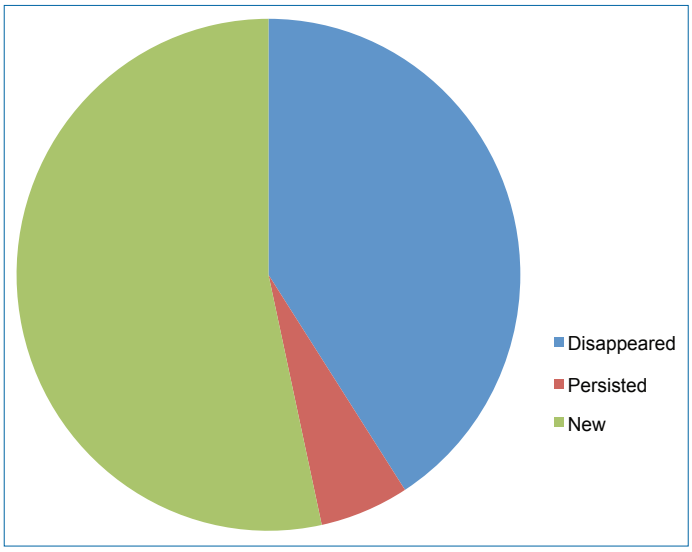
Figure 5: Number of Days an HS was Observed During the Study



Source: Authors.

While observing a fraction of the circle throughout the study, and bearing in mind that an HS publishes to two random points on the DHT circle each day, one would expect to see a long-lived HS publish again and again to the relays. Figure 5 shows the number of days an HS was observed during the study. The largest number of HSes were only seen once, which suggests that they existed for a short period of time and were never seen again. Longer-lived HSes accounted for only 15 percent of all HSes. Therefore, one can conclude that while there are many HSes, most only exist for a short period of time and are not long-lived services. The reason for this is unknown.

Figure 6: HSes that Persisted More than 18 Months



Source: Authors.

To further confirm the hypothesis that HSes have a high turnover, the authors of an earlier study (Biryukov, Pustogarov and Weinmann 2013) were approached to provide their collected list. Figure 6 shows the number of HSes that disappeared, persisted or been created between the two studies. As one might expect, given the previous results, most HSes did not persist for long and many newer ones had replaced them.

AUTHENTICATED HSes

Tor allows HSes to be authenticated in such a way that one cannot locate the position on the DHT circle without knowing a secret, such as a password; therefore, unless one knows the secret, the service cannot be accessed. On the directory relays, it is trivial to identify the descriptors belonging to authenticated HSes, because they are encrypted with the secret and so one can simply count encrypted descriptors (although not decrypt them). During this study, only 0.6 percent of HSes were authenticated. The content of these HSes is unknown, as without the secret it is impossible to access them.

CONTENT AND POPULARITY ANALYSIS

CLASSIFICATION

An HS does not have to be a website but could be, for example, a chat room, a file server or any other form of Internet service. There is no mechanism in Tor to find out which services are available for use by visitors, and the only way to discover them is to try each in turn (see Biryukov, Pustogarov and Weinmann 2013 for further analysis). Web pages were nearly universally offered by the most frequently requested HSes. To identify the type of content available, a custom crawler was developed that would connect to each HS, download web content and extract key data points. These data points were then used for classification of content type.

Classification of web pages is a difficult task, and while there exist automatic classifiers based on machine learning, the dataset in this case was small enough that manual classification was not unduly onerous. Additionally, the authors felt that given the range and complex, technical nature of some of the content, automatic classifiers would be insufficient due to difficulty in interpreting context and meaning (Samarawickrama and Jayaratne 2011).

Deciding when to crawl is not straightforward. At first glance, one may assume that crawling throughout the study is the logical approach; however, in doing so, one will overrepresent short-lived HSes, most of which were not online concurrently. Instead, the preferred approach is to take a snapshot of the content at a particular period in time; having observed the turnover and short longevity of services, the crawling took place over a one-month period toward the end of the study. It is acknowledged that there is room for imprecision, but there is presently no better approach available. As there is a high turnover of HSes, this is not significant.

It was considered at the outset that some HSes may contain content that was obscene or otherwise illegal to download, and it was more than likely that if this content existed it would be in the form of multimedia or images; hence, the crawler only fetched textual content from each HS, parsed key data points and stored them in a database. The crawled data were inspected to produce a list of categories that covered the majority of the content. Afterward, each site was manually examined and classified into distinct categories. Where a site spanned two categories, the authors chose the category that more precisely described the overwhelming or primary purpose.

While there is often debate about the division of illegal and legal content on the Dark Net, it is difficult to classify sites into either legal or illegal due to discrepancies and intricacies between legal jurisdictions: for example, whistle-blowing sites are often considered legal, but may not be if used to disclose classified documents or by persons in repressive regimes. Therefore, classification into legal versus illegal has not been undertaken. That said, the majority of sites on the Tor Dark Net are likely to be illegal (or considered immoral) in many Western countries.

The classification categories are as follows, with notable examples where appropriate:

- **Abuse:** sites where the title indicates some form of sexual abuse (typically minors), likely to be illegal in most Western jurisdictions. Sadly, these pages were easily identifiable from the metadata, suggesting

webmasters had confidence that Tor would provide robust anonymity. For some sites, it was difficult to discern whether they were facilitating abuse or providing adult pornographic services, and due to legal restrictions we were unwilling to download images to confirm. Where this was the case, the site was put into the porn category.

- **Anonymity:** sites aimed at promoting (or teaching) the use of anonymity tools or anonymous culture.
- **Bitcoin:** currency exchange from a mainstream currency to bitcoin, but more often money-laundering services.
- **Blog:** personal or topical blog, often covering topics such as hacktivism.
- **Books:** ebook service typically offering copyrighted material for free.
- **Chat:** web-based chat service, excluding services such as Jabber and Internet Relay Chat.
- **Counterfeit:** sites offering counterfeit items; notable fake currency, such as notes, or fake passports/identity documents.
- **Directory:** site offering links to other sites within the Dark Net, often used for discovering other sites.
- **Drugs:** the sale or purchase of narcotics; typically, marketplaces connecting buyers and sellers.
- **Forum:** web-based forum whose primary purpose does not fit into another category; for example, generalist forum.
- **Fraud:** sites attempting to obtain a pecuniary advantage by deception.
- **Gambling:** any site that promotes/supports gambling. Bitcoin gambling services were most prevalent here, whereby users would first convert their fiat currency to bitcoin.
- **Guns:** sites exclusively aimed at selling guns.
- **Hacking:** site providing instructional information on illegal computer hacking.
- **Hosting:** Dark Net hosting service allowing users to host another Dark Net site.
- **Mail:** Dark Net web-based email or messaging service; examples include Mail2Tor and the now defunct TorMail.
- **Market:** a marketplace selling items other than drugs or services covered in other categories.
- **News:** news service such as current affairs or news specific to the Dark Net.
- **Porn:** Pornography sites that carry material that would be legal in most Western jurisdictions.
- **Search:** site providing a search engine-type service; one example is Ahmia.
- **Whistleblower:** sites typically operated by journalists for whistleblowers to submit documents. The GlobaLeaks platform (Hermes Center for Transparency and Digital Human Rights 2014) and SecureDrop platform (Freedom of the Press Foundation 2014) were prominently featured in this category.
- **Wiki:** user-editable content, such as the Hidden Wiki.

POPULARITY

The popularity data here shows the number of requests made for the descriptor for a particular HS. Bearing in mind the earlier point about the Tor program caching descriptors, one can interpret the number of requests as between the number of visits and the number of visitors. Due to the anonymity offered by Tor, it is not possible to link two separate requests to the same person, but since their computer will remember descriptors until the Tor software is restarted, it often will not make multiple requests within a 24-hour period.

Table 1 shows the popularity of HSes for which the authors received a descriptor request, but did not received a publication during the study. These are addresses that no longer exist, but are still being requested by Tor clients. In many cases, it was possible to identify the purpose of these now extinct HSes by examining online malware reports or by word prefixes present in the .onion address.

Almost all the top 40 HSes requested but no longer operating were botnet command and control (C&C) servers (Stone-Gross et al. 2009). Botnet C&C servers are used to control computers infected with malware (called bots) remotely; the bot will connect to the server regularly for new instructions or to upload data (such as stolen passwords). Malware authors and researchers have been involved in a cat-and-mouse game in the last 10 years, whereby authors have attempted to produce C&C servers that are difficult to takedown. Tor has become a popular tool for C&C infrastructure, due to the difficulty in taking down and locating servers. Interestingly, most botnets represented in the dataset had many (as opposed to a single) HS addresses, which paradoxically may make them more vulnerable to deanonymization attacks if these services are distributed across several Tor processes (Murdoch and Zielinski 2007; Johnson et al. 2013).

Table 1: Popularity of No-longer Existent HSes

HS Address	Requests/day	Days observed	Description
l77ukkijtdca2tsy	679,470	9	Botnet Sefnit
7sc6xyn3rrxtknu6	525,930	11	Botnet Sefnit
pomyeasfnmtn544p	514,766	10	Botnet Sefnit
ceif2rmdoput3wjh	247,296	6	Botnet Sefnit
censored	6,603	10	Child abuse

Source: Authors.

Table 2 shows a cross-section of the widely known onion addresses by the number of visitors they received each day. Abuse sites were by far the most popular and these sites were easily identified by words in the page title or by prefixes used on the .onion address. The Hidden Wiki also featured and is often used as a starting point for many visitors to the Dark Net. It is perhaps surprising, given the amount of media attention that Silk Road receives, that the number of its requests is fewer than 10,000.

Table 2: Non-sequential Snapshot of Popular HSes

HS Address	Requests/ day	Days observed	Description
censored	168,152	12	Child abuse
silkroad6ownowfk	8,067	11	Silk Road
agorabasakxmewww	3,035	8	Agora
k5zq47j6wd3wdvjq	2,589	5	Evolution
xmh57jrzrnw6insl	1,341	7	Torch
3g2upl4pq6kufc4m	1,223	4	DuckDuckGo
wikitjerrta4ggz4	555	12	HiddenWiki
mail2tor2zyjdctd	266	8	Mail

Source: Authors.

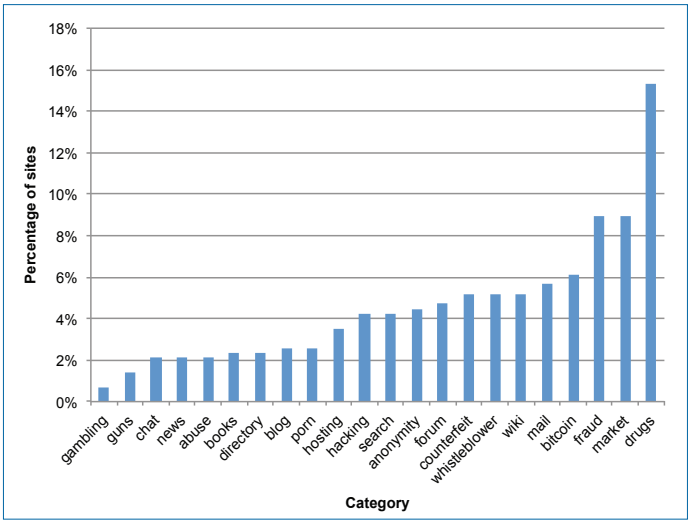
CLASSIFICATION

There are two representations of the classifications of data, the first being the number of sites in each category (see Figure 7). Figure 7 shows that the Dark Net's content is diverse, with the largest number of sites beingrepresented in the drugs category, but only by a small margin.

When each category is plotted against the percentage of HS directory requests it received (using the previous hits data), an entirely different picture emerges (see Figure 8). Requests to abuse sites represented more than 80 percent of total requests observed, although they accounted for only two percent of the total HSes available (see Figure 7).

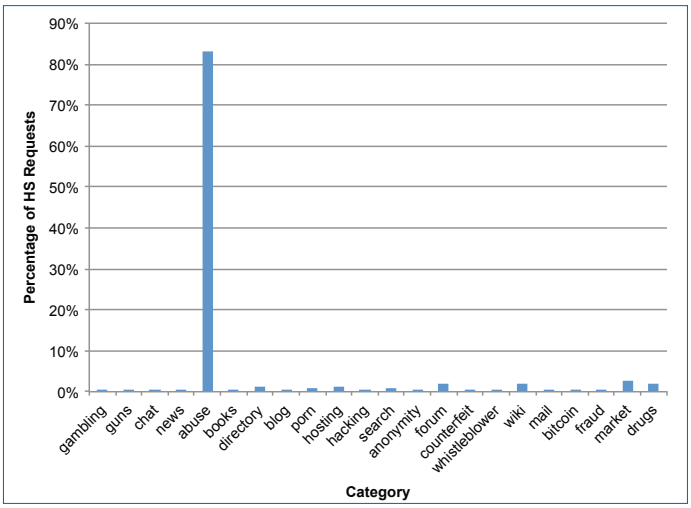
It is important to emphasize what is being measured. The popularity data is a measure of the number of HS directory requests, and when grouped into content-type categories the picture may become somewhat misleading. First, law enforcement frequently patrol abuse sites and this may inflate the figures; however, crawlers are likely

Figure 7: Percentage of Sites in each Classification Category



Source: Authors.

Figure 8: Percentage of Requests by Classification Category



Source: Authors.

to account for a single request in a 24-hour period and we are seeing a large number of requests to these sites. Even if it's assumed that all national forces crawl these sites daily, they would still only account for a small proportion of the total requests. The second possibility is denial of service attacks, where one could flood the HS directory with requests for descriptors in an effort to take the directory offline. This is likely to be ineffective because the attacker would need to take all six directories offline and then these relays would be dropped from the consensus and the responsibility would shift to other relays. It is worth noting that most of these sites were observed on several random days during the study, so an attack of this nature would have to persist for most of the duration of the study. While denial of service attacks are impossible to rule out,

due to the anonymity offered by Tor, it seems unlikely and in any case none of our servers were taken offline or received requests far exceeding expectations.

Tor offers a tool called Tor2Web, which allows non-Tor users to visit HSes through a web gateway. These web gateways will operate one or a small number of Tor clients, so although there might be several visitors to a site, only one request will be seen because the gateway has cached the descriptor; hence, it is possible for some sites to be underrepresented in the data if they are largely accessed through Tor2Web. The popularity data is the proportion of HS directory requests observed on HSes offering a website that the authors were able to crawl and classify. One should be extremely cautious before trying to link this data to a number of users, as the data approximates somewhere between visits and visitors. Interpreting the figure as visits will underestimate the number of users.

CONNECTIVITY

For each HS website that was crawled, information was extracted on the hyperlinks listed on their site. Each link was categorized into one of three categories: Dark Net, clearnet or own-site. Dark Net links were links to other Tor HSes, clearnet links were links to websites that were hosted on the Internet (for example, regular non-Tor domains) and own-site were links within the site being crawled.

Of the HSes that were crawled, 59 percent did not link to any site other than itself, seven percent linked only to other Dark Net sites, 23 percent linked only to clearnet sites and 11 percent linked to both. Aggregating the first two figures, one can say that two-thirds of sites were not connected by links to any sites outside of the Dark Net.

DEANONYMIZATION OF TOR USERS AND HSES

A common misconception is that Tor is resistant to state-level surveillance and that its users can therefore act with impunity. In reality, any suitably resourced entity can launch attacks with high success rates while maintaining a minimal risk of detection.

While an observer cannot see where traffic is routed in the Tor network, he can treat the network as a black box and observe traffic entering and leaving it. An interesting analogy would be the postal service, whereby one cannot see what happens in the sorting office, but can see how many letters/parcels every address posts and receives on each day. Assume the intelligence services think that two people are pen pals: they can observe letters leaving one person and arriving at the other and vice versa. Observing the mail of both parties over a period of time can give a degree of confidence about whether they are communicating with each other without opening their mail or tracking it through the postal system.

The postal system analogy may seem like there is a lot of room for error, but with Tor, a typical user may send millions of letters and an observer can see the precise time they were sent and received. It is therefore easy to confirm with high probability that two parties are communicating. A slightly harder version of this problem occurs when one can observe traffic exiting the Tor network to a jihadist website: can he identify the original user while still treating Tor as a black box? The answer is often yes, provided he has enough visibility of traffic entering the Tor network to correlate the number of messages, the rate and time at which they are sent. One does not need to control guard relays to be able to launch traffic correlation attacks; one needs to be able to observe traffic between a user and his guard even though the traffic cannot be read at that point. Recent leaks from Edward Snowden indicate that UK and US intelligence services can observe traffic from entire countries, enabling them to observe all guards within those countries. Guards are presently changed every 30 to 90 days, so a targeted user may fall within the net at some point in the future when global observation is not possible.

While HSes are believed to be the Holy Grail in anonymity protection, in reality these correlation attacks are much more successful compared with attacks against general web browsing through Tor. Typically with HSes, one wishes to deanonymize the visitors and the service itself. In the last example, with general Tor usage one, was observing traffic entering and leaving the Tor network, while with HSes the attacker can control one end of the connection and inject patterns of traffic to spot. In the case of a user, the attacker can control the relay in the DHT and send a specific pattern back to the user and try to identify it leaving the network. The attacker will be able to identify the service visited and the user but not what the user does on the site, because the content is encrypted end-to-end between the two parties. In some cases, the mere fact a user has visited an HS may be enough to gain a conviction (for example, in particular where the site contains illegal content on the front page). That said, once a user has been identified, his home and equipment can be searched, where there may be stored evidence of wrongdoing. Thus far, the authors are aware of no cases whereby a deanonymization attack alone has been used to seek a conviction.

In the case of the service, the attacker can simply connect to the HS and send a pattern, and again attempt to identify it leaving the Tor network. Deanonymization attacks against HSes can be highly successful with very low (even absent) false positives.

BLOCKING OF TOR

Tor is often described as being censorship resistant and impossible to block — this is not the case. There is a misunderstanding of how Tor works and some nations have attempted naive approaches that have, predictably,

failed. There are many effective approaches to blocking Tor and the problem of building a truly censorship-resistant network is presently an open one.

When a user wishes to connect to the Tor network, he needs a list of all the relays, which he obtains from the directory authorities — special relays operated by the Tor developers. He then selects a relay from the list that meets certain characteristics (principally uptime and bandwidth) and chooses this as the first node in any circuit. Since the list of relays (known as the consensus) is public, anyone is able to download the list and block access to all of them. The user would then be unable to connect to the first hop and into the network.

An attempt to mitigate these blocking attempts was made through the introduction of "bridges." Bridges are not listed in the consensus and one has to visit the Tor Project website and enter a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) to obtain a small number of them. The Tor website will only release a small fraction of bridges to any one user on any one day, which makes it difficult for an attacker to obtain the full list and block them.

Notably, China operates a country-wide firewall used to censor material that citizens can access. China has attempted to block Tor by stopping access to all of the relays listed in the consensus as described above (Winter and Lindskog 2012). They have also attempted to block bridges by looking for connections traversing the firewall to see if they met characteristics typical of Tor. This worked for some time until the Tor developers modified the program so that Tor connections were indistinguishable from ordinary web traffic. In response to this, China then monitored any encrypted connections and would try to connect to the remote server and talk to the Tor protocol: if it responded, they concluded it was a Tor relay; if not, it was an ordinary website.

There exist, however, many more successful techniques for bridge enumeration (Ling, Luo and Yang 2012) — that is, detecting potential bridges without asking the Tor Project for a list or scanning suspect servers. A simple approach is to run a Tor relay and monitor all of the circuits built through your relay. It is easy to identify whether you are the middle relay, so you can simply identify the previous hop and if it is not in the network it is probably a bridge. This technique is not foolproof because not all bridges will connect through your relay all of the time; hence, you must run many relays offering a significant proportion of the bandwidth to detect most of them, and even then you will only detect most, but not all.

That said, it is believed that these techniques would be effective, as only the most determined user would continue to persist with a tool that failed most of the time.

HS BLOCKING

While Tor is designed to be resistant to censorship, at present HSes are not particularly robust against technical attacks (they will resist physical attacks if the operator is unknown). At present, groups of individuals or the Tor Project itself could choose to block these sites by the following methods:

- An individual can block a single site by launching several relays and ensuring they occupy the positions in the DHT of the responsible relays for that service. If someone comes to the relay asking for the descriptor, the individual can simply deny it.

- Operators of Tor relays could themselves choose to block the content by patching their relays to deny requests to these sites. This would require the cooperation of a large percentage of relay operators to be effective, but it would be a decentralized blocking mechanism requiring some consensus.

The Tor Project itself can choose to trivially block the content by modifying the Tor program to block requests for such sites at the relay and client level. This might seem to place a large amount of power into the developer's hands, but it is worth remembering they already control the authorities and the consensus, and can abuse this to deanonymize users or block sites anyway. At present, the Tor Project has stated that it is not willing to censor HSes, because it fears it will be a slippery slope with future requests widening the categories blocked. This is unfortunate because child abuse sites do cause real harm and may encourage offenders. The number of requests for whistleblowing sites is minuscule in comparison to those aimed at child abusers.

CONCLUSION

This paper does suggest that child abuse content is the most popular type of content on the Tor Dark Net. While law enforcement may crawl such sites, the number of requests that would be seen would be only a tiny fraction, and hence not skew the outline ratios. Similarly, denial of service attacks were not observed and so are also unlikely to account for the high requests. The usage of Tor2Web may underrepresent some categories, but it is not currently clear whether, or why, such groups would exclusively use this tool.

An explicit categorization of sites into illegal and legal was not undertaken, but it was abundantly clear to the authors that the majority of sites were of questionable legality. While anonymity and privacy tools such as Tor might fight online surveillance, they also give an easy and accessible route for those with criminal motivations. There are alternatives, such as botnets, available for criminal activity, but these do not negate the comparative ease with which Tor can be used.

It is technically possible to block Tor, although it is likely that the Tor Project will deploy countermeasures resulting in the endeavour descending into a cat-and-mouse game of "circumvent-and-censor." In any case, Tor does not provide the absolute impunity that is often attributed to it.

WORKS CITED

Biryukov, A., I. Pustogarov and R.-P. Weinmann. 2013. "Trawling for Tor Hidden Services: Detection, Measurement and Deanonymisation." *Proceedings of IEEE Symposium on Security and Privacy*: 80–94.

Dingledine, R., N. Mathewson and P. Syverson. 2004. "Tor: The Second-Generation Onion Router." Washington, DC: Naval Research Laboratory.

Douceur, J. R. 2002. "The Sybil Attack." *Revised Papers from the First International Workshop on Peer-to-Peer Systems*: 251–60.

Freedom of the Press Foundation. 2014. "SecureDrop Platform."

Hermes Center for Transparency and Digital Human Rights. 2014. "GlobaLeaks Platform."

Jardine, Eric. n.d. "Tor, What Is It Good For? Political Rights and Online Anonymity-Granting Technologies." Unpublished manuscript.

Johnson, A., C. Wacek, R. Jansen, M. Sherr and P. Syverson. 2013. "Users Get Routed: Traffic Correlation on Tor by Realistic Adversaries." *Proceedings of the 20th ACM conference on Computer and Communications Security.*

Lesniewski-Laas, C. and M. F. Kaashoek. 2010. "Whanau: A Sybil-proof Distributed Hash Table." *Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation.*

Ling, Z., J. Luo and M. Yang. 2012. "Extensive Analysis and Large-scale Empirical Evaluation of Tor Bridge Discovery." IEEE INFOCOM. Orlando, FL.

Murdoch, S. J. and P. Zielinski. 2007. "Sampled Traffic Analysis by Internet-Exchange-Level Adversaries." *Proceedings of the Seventh Workshop on Privacy Enhancing Technologies.*

Samarawickrama, S. and L. Jayaratne. 2011. "Automatic text classification and focused crawling." Sixth International Conference on Digital Information Management.

Stoica, I., R. Morris, D. Karger, M. FransKaashoek and H. Balakrishnan. 2001. "Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications." *Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*: 149–60.

Stone-Gross, B., M. Cova, L. Cavallaro, B. Gilbert, M. Szdlowski, R. Kemmerer, C. Kruegel and G. Vigna. 2009. "Your Botnet is My Botnet: Analysis of a Botnet Takeover." *Proceedings of the 16th ACM Conference on Computer and Communications Security*: 635–47.

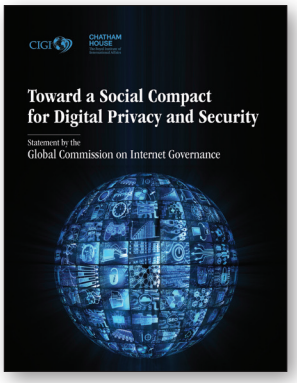
Tor Project. 2015. "Metrics Portal." June.

Winter, P. and S. Lindskog. 2012. "How the Great Firewall of China is blocking Tor." *Proceedings of USENIX Workshop on Free and Open Communications on the Internet.*

CIGI PUBLICATIONS
ADVANCING POLICY IDEAS AND DEBATE

Global Commission on Internet Governance

The Global Commission on Internet Governance (GCIG) was established in January 2014 to articulate and advance a strategic vision for the future of Internet governance. The two-year project conducts and supports independent research on Internet-related dimensions of global public policy, culminating in an official commission report that will articulate concrete policy recommendations for the future of Internet governance. These recommendations will address concerns about the stability, interoperability, security and resilience of the Internet ecosystem. Launched by two independent global think tanks, the Centre for International Governance Innovation and Chatham House, the GCIG will help educate the wider public on the most effective ways to promote Internet access, while simultaneously championing the principles of freedom of expression and the free flow of ideas over the Internet.

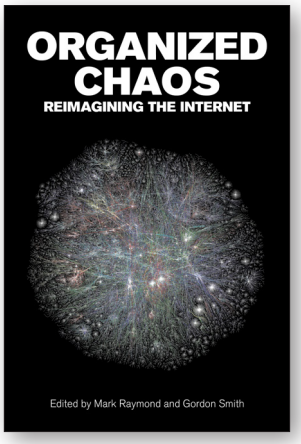


Toward a Social Compact for Digital Privacy and Security

Statement by the Global Commission on Internet Governance

On the occasion of the April 2015 Global Conference on Cyberspace meeting in The Hague, the Global Commission on Internet Governance calls on the global community to build a new social compact between citizens and their elected representatives, the judiciary, law enforcement and intelligence agencies, business, civil society and the Internet technical community, with the goal of restoring trust and enhancing confidence in the Internet. It is now essential that governments, collaborating with all other stakeholders, take steps to build confidence that the right to privacy of all people is respected on the Internet. This statement provides the Commission's view of the issues at stake and describes in greater detail the core elements that are essential to achieving a social compact for digital privacy and security.

Available for free download at www.cigionline.org/publications



Organized Chaos

CDN$25

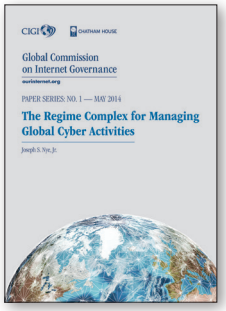
Edited by Mark Raymond and Gordon Smith

"Anonymous." Cybercrime. Hacktivist. Cyber security. Now part of the lexicon of our daily language, these words were unknown a decade ago. The evolution and expansion of the Internet has transformed communication, business and politics, and the Internet has become a powerful influence on everyday life globally. But the Internet is a medium that is not controlled by one centralized system, and the debate over who will govern the Internet has commanded attention from a wide range of actors, including states, policy makers and those beyond the traditional tech industries.

Organized Chaos: Reimagining the Internet examines the contemporary international politics of Internet governance problems, exploring issues such as cybercrime, activities of the global hacktivist network Anonymous and "swing states," and highlighting central trends that will play a role in shaping a universal policy to govern the Internet. In this book, some of the world's foremost Internet governance scholars consider the critical problems facing efforts to update and refine Internet governance at an international level and the appropriate framework for doing so. This volume provides the basis for developing a high-level strategic vision required to successfully navigate a multi-faceted, shifting and uncertain governance environment.

Available for purchase at www.cigionline.org/bookstore

GLOBAL COMMISSION ON INTERNET GOVERNANCE PAPER SERIES



The Regime Complex for Managing Global Cyber Activities
GCIG Paper Series No. 1
Joseph S. Nye, Jr.

Tipping the Scale: An Analysis of Global Swing States in the Internet Governance Debate
GCIG Paper Series No. 2
Tim Maurer and Robert Morgus

Legal Mechanisms for Governing the Transition of Key Domain Name Functions to the Global Multi-stakeholder Community
GCIG Paper Series No. 3
Aaron Shull, Paul Twomey and Christopher S. Yoo

Legal Interoperability as a Tool for Combatting Fragmentation
GCIG Paper Series No. 4
Rolf H. Weber

Innovations in Global Governance: Toward a Distributed Internet Governance Ecosystem
GCIG Paper Series No. 5
Stefaan G. Verhulst, Beth S. Noveck, Jillian Raines and Antony Declercq

The Impact of the Dark Web on Internet Governance and Cyber Security
GCIG Paper Series No. 6
Tobby Simon and Michael Chertoff

On the Nature of the Internet
GCIG Paper Series No. 7
Leslie Daigle

Understanding Digital Intelligence and the Norms That Might Govern It
GCIG Paper Series No. 8
David Omand

ICANN: Bridging the Trust Gap
GCIG Paper Series No. 9
Emily Taylor

A Primer on Globally Harmonizing Internet Jurisdiction and Regulations
GCIG Paper Series No. 10
Michael Chertoff and Paul Rosenzweig

Connected Choices: How the Internet is Challenging Sovereign Decisions
GCIG Paper Series No. 11
Melissa E. Hathaway

Solving the International Internet Policy Coordination Problem
GCIG Paper Series No. 12
Nick Ashton-Hart

Net Neutrality: Reflections on the Current Debate
GCIG Paper Series No. 13
Pablo Bello and Juan Jung

Addressing the Impact of Data Location Regulation in Financial Services
GCIG Paper Series No. 14
James M. Kaplan and Kayvaun Rowshankish

Cyber Security and Cyber Resilience in East Africa
GCIG Paper Series No. 15
Iginio Gagliardone and Nanjira Sambuli

Global Cyberspace Is Safer than You Think: Real Trends in Cybercrime
GCIG Paper Series No. 16
Eric Jardine

The Emergence of Contention in Global Internet Governance
GCIG Paper Series No. 17
Samantha Bradshaw, Laura DeNardis, Fen Osler Hampson, Eric Jardine and Mark Raymond

Landmark EU and US Net Neutrality Decisions: How Might Pending Decisions Impact Internet Fragmentation?
GCIG Paper Series No. 18
Ben Scott, Stefan Heumann and Jan-Peter Kleinhans

The Strengths and Weaknesses of the "Brazilian Internet Bill of Rights": Examining a Human Rights Framework for the Internet
GCIG Paper Series No. 19
Carolina Rossini, Francisco Brito Cruz, Danilo Doneda

Available for free download at www.cigionline.org/publications

CIGI
Centre for International Governance Innovation
www.cigionline.org

ABOUT CIGI

The Centre for International Governance Innovation is an independent, non-partisan think tank on international governance. Led by experienced practitioners and distinguished academics, CIGI supports research, forms networks, advances policy debate and generates ideas for multilateral governance improvements. Conducting an active agenda of research, events and publications, CIGI's interdisciplinary work includes collaboration with policy, business and academic communities around the world.

CIGI's current research programs focus on three themes: the global economy; global security & politics; and international law.

CIGI was founded in 2001 by Jim Balsillie, then co-CEO of Research In Motion (BlackBerry), and collaborates with and gratefully acknowledges support from a number of strategic partners, in particular the Government of Canada and the Government of Ontario.

Le CIGI a été fondé en 2001 par Jim Balsillie, qui était alors co-chef de la direction de Research In Motion (BlackBerry). Il collabore avec de nombreux partenaires stratégiques et exprime sa reconnaissance du soutien reçu de ceux-ci, notamment de l'appui reçu du gouvernement du Canada et de celui du gouvernement de l'Ontario.

For more information, please visit www.cigionline.org.

ABOUT CHATHAM HOUSE

Chatham House, the Royal Institute of International Affairs, is based in London. Chatham House's mission is to be a world-leading source of independent analysis, informed debate and influential ideas on how to build a prosperous and secure world for all. The institute: engages governments, the private sector, civil society and its members in open debates and confidential discussions about significant developments in international affairs; produces independent and rigorous analysis of critical global, regional and country-specific challenges and opportunities; and offers new ideas to decision-makers and -shapers on how these could best be tackled from the near- to the long-term. For more information, please visit: www.chathamhouse.org.

CIGI MASTHEAD

EXECUTIVE

President	Rohinton P. Medhora
Director of the International Law Research Program	Oonagh Fitzgerald
Director of the Global Security & Politics Program	Fen Osler Hampson
Director of Human Resources	Susan Hirst
Director of the Global Economy Program	Domenico Lombardi
Vice President of Finance	Mark Menard
Chief of Staff and General Counsel	Aaron Shull

PUBLICATIONS

Managing Editor, Publications	Carol Bonnett
Publications Editor	Jennifer Goyder
Publications Editor	Vivian Moser
Publications Editor	Patricia Holmes
Publications Editor	Nicole Langlois
Graphic Designer	Melodie Wakefield
Graphic Designer	Sara Moore

COMMUNICATIONS

Communications Manager	Tammy Bender	tbender@cigionline.org (1 519 885 2444 x 7356)

CIGI

67 Erb Street West
Waterloo, Ontario N2L 6C2
tel +1 519 885 2444 fax +1 519 885 5450
www.cigionline.org

CHATHAM HOUSE
The Royal Institute of International Affairs

10 St James's Square
London, England SW1Y 4LE, United Kingdom
tel +44 (0)20 7957 5700 fax +44 (0)20 7957 5710
www.chathamhouse.org