How Do Tor Users Interact With Onion Services?

Philipp Winter, Anne Edmundson, and Laura M. Roberts, ***Princeton University;***
Agnieszka Dutkowska-Żuk, ***Independent;***
Marshini Chetty and Nick Feamster, ***Princeton University***

https://www.usenix.org/conference/usenixsecurity18/presentation/winter

This paper is included in the Proceedings of the 27th USENIX Security Symposium.

August 15–17, 2018 • Baltimore, MD, USA

ISBN 978-1-931971-46-1

Open access to the Proceedings of the 27th USENIX Security Symposium is sponsored by USENIX.

How Do Tor Users Interact With Onion Services?

Philipp Winter
Princeton University

Anne Edmundson
Princeton University

Laura M. Roberts
Princeton University

Agnieszka Dutkowska-Żuk
Independent

Marshini Chetty
Princeton University

Nick Feamster
Princeton University

Abstract

Onion services are anonymous network services that are exposed over the Tor network. In contrast to conventional Internet services, onion services are private, generally not indexed by search engines, and use self-certifying domain names that are long and difficult for humans to read. In this paper, we study how people perceive, understand, and use onion services based on data from 17 semi-structured interviews and an online survey of 517 users. We find that users have an incomplete mental model of onion services, use these services for anonymity and have varying trust in onion services in general. Users also have difficulty discovering and tracking onion sites and authenticating them. Finally, users want technical improvements to onion services and better information on how to use them. Our findings suggest various improvements for the security and usability of Tor onion services, including ways to automatically detect phishing of onion services, more clear security indicators, and ways to manage onion domain names that are difficult to remember.

1 Introduction

The Tor Project's onion services provide a popular way of running an anonymous network service. In contrast to anonymity for clients (*e.g.*, obfuscating a client IP address using a virtual private network), Tor onion services provide anonymity for servers, allowing a web server to obfuscate its network location (specifically, its IP address). An operator of a web service may need to anonymize the location of a web service to escape harassment, speak out against power, or voice dissenting opinions.

Onion services were originally developed in 2004 and have recently seen growing numbers of both servers and users. As of June 2018, The Tor Project's statistics count more than 100,000 onion services each day, collectively serving traffic at a rate of nearly 1 Gbps. In addition to web sites, onion services include metadata-free instant messaging [4] and file sharing [15]. The Tor Project currently does not have data on the number of onion service users, but Facebook reported in 2016 that more than one million users logged into its onion service in one month [20].

Onion services differ from conventional web services in four ways; First, they can only be accessed over the Tor network. Second, onion domains are hashes over their public key, which make them difficult to remember. Third, the network path between client and the onion service is typically longer, increasing latency and thus reducing the performance of the service. Finally, onion services are private by default, meaning that users must discover these sites organically, rather than with a search engine.

In this paper, we study how users cope with these idiosyncrasies, by exploring the following questions:

- What are users' mental models of onion services?
- How do users use and manage onion services?
- What are the challenges of using onion services?

Because onion services depend on the Tor Browser and the underlying Tor network to exchange traffic, some of our study also explored users' mental models of Tor itself, but this topic is not the focus of our paper.

To answer these questions, we employed a mixed-methods approach. First, we conducted exploratory interviews with Tor and onion service users to guide the design of an online survey. We then conducted a large-scale online survey that included questions on Tor Browser, onion service usage and operation, onion site phishing, and users' general expectations of privacy. Next, we conducted follow-up interviews to further explore the topics and themes that we discovered in the exploratory interviews and survey. We complemented this qualitative data with an analysis of "leaked" DNS lookups to onion domains, as seen from a DNS root server; this data gave us insights into actual usage patterns and allowed us to corroborate some of the findings from the interviews and surveys.

We find that many Tor users misunderstand technical aspects of onion services, such as the nature of the domain format, rendering these users more vulnerable to phishing attacks. Second, we find that users have many issues using and managing onion services, including having trouble discovering and tracking new onion domains. Our data also suggests that users may visit onion domains that are slight variations of popular onion domains, suggesting that typos or phishing attacks may occur on onion domains. Third, users want improvements to onion services such as improved performance and easier ways to keep track of and verify onion domains as authentic. Many of the shortcomings that we discover could be addressed with straightforward and immediate improvements to the Tor Browser, including improved security indicators and mechanisms to automatically detect domains that may be typos or phishing attacks.

Tor is currently testing the next generation of onion services, which will address various security issues and upgrade to faster, future-proof cryptography. The findings from our work can inform the design of privacy and security enhancements to onion services and Tor Browser at a critical time as these improvements are being deployed. This paper makes the following contributions:

- We provide new, large-scale empirical evidence from Tor users that sheds light on how these users perceive, use, and manage onion services. Our work confirms and extends previous findings on Tor Browser users' mental models [9].
- We provide empirical evidence that characterizes onion domain name lookups based on a dataset from the `.onion` requests from DNS B root, both extending previous work on onion domain usage [18, 33] and corroborating our findings about usability and security problems that we identified in the survey and interview data.
- Based on our findings, we identify usability obstacles to the adoption of onion services and suggest possible design enhancements, including publishing mechanism for onion services and a Tor Browser extension that allows its users to securely and privately bookmark onion domains.

All code, data, and auxiliary resources are available at `https://nymity.ch/onion-services/`.

The rest of this paper is structured as follows. Section 2 provides background on onion services, and Section 3 presents related work. Section 4 presents the methods for our interviews, online survey, and DNS data analysis. Section 5 presents results, Section 6 discusses the implications of these findings, and Section 7 concludes.

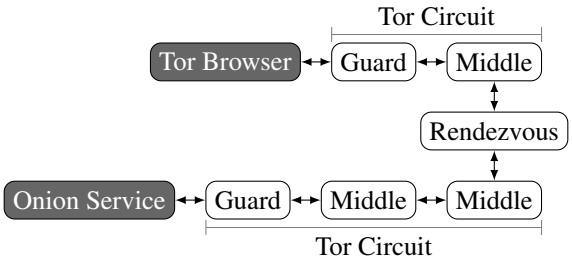


Figure 1: *A path to an onion service typically has six Tor relays. Both the client and the onion service create a Tor circuit (comprising two and three relays, respectively) to a rendezvous.*

2 Background: What Are Onion Services?

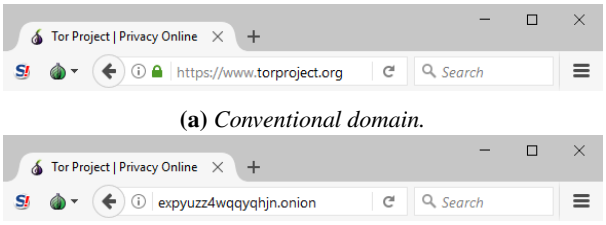
Originally called "hidden services", onion services were renamed in 2015 to reflect the fact that they provide more than just the "hiding" of a service [11]—more importantly, they provide end-to-end security and self-certifying domain names. Beyond The Tor Project's nomenclature, the "web" of onion services is occasionally referred to as the "Dark Web". In this paper, we use only the term onion services.

Onion services are TCP-based network services that are accessible only over the Tor network and provide mutual anonymity: the Tor client is anonymous to the server, and the server is anonymous to the client. Clients access onion services via onion domains that are meaningful only inside the Tor network. A path between a client and onion service has six Tor relays by default, as shown in Figure 1; the client builds a circuit to a "rendezvous" Tor relay, and the onion service builds a circuit to that same relay. Neither party learns the other's IP address.

To create an onion domain, a Tor daemon generates an RSA key pair, computes the SHA-1 hash over the RSA public key, truncates it to 80 bits, and encodes the result in a 16-character base32 string (*e.g.*, `expyuzz4wqqyqhjn`). Because an onion domain is derived directly from its public key, onion domains are self-certifying: if a client knows a domain, it automatically knows the corresponding public key. Unfortunately, this property makes the onion domain difficult to read, write, or remember.

As of February 2018, The Tor Project is deploying the next generation of onion services, whose domains have 56 characters [16, § 6] that include a base32 encoding of the onion service's public key, a checksum, and a version number. New onion services will also use elliptic curve cryptography, allowing the entire public key to be embedded in the domain, as opposed to only the hash of the public key. These changes will naturally improve the security of onion services but have important implications for usability, particularly as unreadable onion domain names get longer.

One way to make onion domains more readable is to repeatedly generate RSA keys until the result-



(a) *Conventional domain.*

(b) *Onion service.*

Figure 2: *Tor Browser 7.0.10's user interface on Windows 10 when accessing the Tor Project website via a conventional domain and the corresponding onion service. The onion service lacks a padlock; Tor developers are addressing this issue [1].*

ing domain contains some desired string (*e.g.*, "facebook"). These so-called *vanity onion domains* include Facebook (`facebookcorewwwi.onion`), ProPublica (`propub3r6espa33w.onion`), and the New York Times (`nytimes3xbfgragh.onion`). Vanity onion domains still typically have strings of characters that are not meaningful words, but they may be easier to memorize. These domains are relatively expensive to create: given base32's alphabet size of 32 characters, a vanity prefix of length n takes an average of $0.5 \cdot 32^n$ key creations, Given a set of domains that contain a vanity prefix, one can search this set for a domain that is the easiest to remember, for example by using a Markov model to filter domains that resemble English words. The popular `scallion` tool [30] parallelizes the search for vanity domains.

Even if the onion domain is more readable, the user still needs to have a way of discovering the onion service in the first place. In contrast to conventional network services, onion services are designed to be difficult to discover. The operator of an onion service must manually advertise the domain, for example by manually adding it to onion site search engines such as Ahmia [22]. The lack of a go-to service such as a "Google for onion services" prompted the community to devise various ways to disseminate onion services through a variety of search engines and curated lists.

Tor Browser aims to make user access to onion domains seamless. Figure 2a shows the interface when accessing The Tor Project's web site; Figure 2b shows a connection to the corresponding onion site. Additionally, because the unreadability of onion domains can make clients more susceptible to phishing attacks, website operators who want to provide their website as an onion service and do not care about their own anonymity can get an extended validation (EV) digital certificate for their .onion domain so that clients can be assured that they are connecting to the correct site. For example, Facebook's onion service has a certificate associated with it, and this added layer of security is reflected in the Tor Browser.

3 Related Work

Usage and mental models of Tor Browser. Forte *et al.* studied the privacy practices of contributors to open collaboration projects such as the Tor Project and Wikipedia to learn about how privacy concerns affect their contribution practices [9]. The study, based on 23 interviews, found that contributors worry about an array of threats, including surveillance, violence, harassment, and loss of opportunity. This study was not focused on hidden services at all. Additionally, Gallagher *et al.* conducted semi-structured interviews to understand both why people use Tor Browser and how they understand the technology [10]. The study found that experts tend to have a network-centric view of the Tor network and use it frequently, whereas non-experts have a goal-oriented view and see Tor Browser as a black-box service. Our work corroborates these findings but is focused on onion services, rather than generally on Tor Browser.

Usability of Tor Browser installation. Tor Browser has seen many usability improvements since its creation in 2003 [31], from a Tor "button" to Tor Browser Bundle (now called the Tor Browser). Ten years ago, Clark *et al.* used cognitive walkthroughs to study how users install, configure, and run Tor Browser [5]. The work revealed hurdles such as jargon-laden documentation, confusing menus, and insufficient visual feedback. Norcie *et al.* identified "stop-points" in the installation and use of the Tor Browser Bundle [21]; these stop-points require user action but instead cause confusion. the study recommended various changes to the installation process and evaluated them in a follow-up study. Lee *et al.* [14] studied the usability of Tor Launcher, the graphical configuration tool that allows users to configure Tor Browser, and found that 79% of users' connection attempts in a simulated censored environment failed, but that various design improvements could reduce these difficulties.

Usability of onion domain names. Previous work aimed to improve the usability of onion domain names. Sai and Fink proposed a mnemonic system that maps 80-bit onion domains to sentences [26]. Their work is inspired by mnemonicode, which maps binary data to words [36]. Victors *et al.* designed the Onion Name System [35], which allows users to reference an onion service by a readable, globally unique identifier. Kadianakis *et al.* designed an API that allows Tor clients to configure name systems (*e.g.*, GNS [28] or OnioNS [35]) on a per-domain basis [12].

Onion domain usage patterns. If a conventional DNS resolver attempts to resolve an `.onion` domain (as might happen when a user enters such a domain name into a normal browser), the resulting DNS lookup for the domain will "leak"to the DNS root servers. Previous studies have

taken advantage of this leaked information to characterize the popularity of various onion domains [18, 33]. We build on previous work, applying similar analysis with a focus on whether the lookups suggest usability problems with onion services or the presence of phishing attacks.

4 Method

We used a mixed-methods approach involving interview and survey data, as well as analysis of DNS query data. This section details our interviews (Section 4.1), large-scale online survey (Section 4.2), and the DNS dataset that we use for our analysis (Section 4.3).[1]

4.1 Interviews

To help us understand users' mental models of onion services, onion service usage, and the challenges and benefits of onion services, we conducted qualitative interviews, which allowed us to design the survey.

4.1.1 Procedure

Interview Guide. We developed a question set that served as the basis for each interview,[2] basing our design on prior work [9] but focusing particularly on onion services. The semi-structured nature of our interviews allowed us to deviate from this question set by asking follow-up questions as appropriate.

We followed standard consent procedures for all participants. We began by asking demographic information (gender, age range, occupation, country of residence, and level of education), followed by questions about users' general online behavior. We concluded with questions about Tor Browser and onion services (*e.g.*, when users started to use these services, how they track onion links as well as the drawbacks and strengths of these services based on their own experiences). To gather data about users' mental models of Tor browser and onion services, we designed a brief sketching exercise similar to those used in other work [25]. We asked participants to draw sketches of how they believed Tor and onion services worked and followed up on these drawings in interviews.

Recruitment. To select eligible interview subjects, we created a short pre-interview survey[3] asking users if they were over 18 years of age, if they had used Tor Browser and onion services, and how they would rate their general privacy and security knowledge. To the extent possible, we targeted lay-people and aimed to maximize cultural, gender, geographic location, education, and age diversity. The Tor Project advertised this survey both in a blog post [37] and via Twitter. We also advertised the study on Princeton's Center for Information Technology (CITP) blog and recruited participants in person at an Internet freedom event.

Recruiting a representative sample of Tor users is difficult, and our recruiting techniques likely resulted in a biased population for several reasons. First, we believe that The Tor Project's blog and Twitter account are followed by disproportionately more technical users, whereas non-technical users may not generally follow news and updates related to Tor via the project's blog and Twitter feed. Second, Tor users value their privacy more than the average Internet user, so the users we recruited may not be as honest and candid about their browsing habits as we would like.

Interviews. We conducted 13 interviews in person and four interviews remotely—over Skype, Signal, WhatsApp, and Jitsi—depending on the medium that our participants preferred. Two participants declined to have their interviews recorded; we recorded the rest of the interviews with the permission of the participant. All participants answered the interview questions and completed the sketching exercise. Each interview ended with a debriefing phase to ask if our participants had any remaining questions. We compensated participants with a $20 gift card. We conducted our first interview on July 13, 2017 and the last on October 20, 2017. The median interview time was 34 minutes, with interviews ranging from 20–50 minutes.

Transcription and Analysis. We transcribed our interview recordings and employed qualitative data coding to analyze the transcripts [29]. In the two cases where we did not have interview recordings, we relied on our field notes. We developed a codebook based on our research questions and used a combination of deductive coding to identify themes of interest we agreed upon and inductive coding to discover emergent phenomena and to expand the initial codebook. We had ten parent codes in total, with examples such as "Mental model of onion services", "Search habits", and "Reasons for using onion services"; and 168 child codes, including "Definition- anonymous", "Word of mouth", and "Curiosity". After we reached consensus on the phenomena of interest, at least two members of our team (sometimes up to four) read and coded each transcript. We also held regular research meetings with the entire team of authors to discuss the coded transcripts and reach consensus on the final themes.

4.1.2 Participants

We interviewed 17 subjects, as summarized in Table 1. We only present aggregate demographic information to

[1] Princeton University's institutional review board (IRB) approved this study (Protocol #8251).

[2] The question set is available at https://nymity.ch/onion-services/pdf/interview-checklist.pdf.

[3] The pre-interview survey is available at https://nymity.ch/onion-services/pdf/pre-interview-survey.pdf.

protect the identity of our interview participants. We believe that our sample is biased towards educated and technical users—almost 60% of our participants have a postgraduate degree—but our sample also shows the diversity among Tor's user base: our participants comprised human rights activists, legal professionals, writers, artists, and journalists, among others. In remainder of the paper, we use the denotation 'P' to refer to interview participants.

4.2 Online Survey

Shortly after we conducted our first batch of interviews, we designed, refined, and launched an online survey to complement our interview data.[4]

4.2.1 Procedure

Survey Design. We created our survey in Qualtrics because an unmodified Tor Browser could display it correctly. Unfortunately, Qualtrics requires JavaScript, and Tor Browser deactivates if it is set to its highest security setting. Several users complained about our reliance on JavaScript in the recruitment blog post comments [37]. All respondents consented to the survey and confirmed that they were at least 18 years old. Our survey was only available in English, but we targeted an international audience because Sawaya *et al.* showed that cultural differences yield different security behavior [27], and paying attention to these differences is central to The Tor Project's global mission.

Most of our survey focused on onion services, but we also included usage questions about Tor in general because Tor Browser is used to access onion services. Our survey had of 49 questions, most of which were closed-ended questions. The first set of questions asked for basic demographic information such as age, gender, privacy and security knowledge rating, and education level. Next, the survey asked about Tor usage, such as how frequently the Tor Browser was used. We also asked about onion services usage in detail, including questions concerning the usability of onion links, how users track and manage onion domain links, whether (and why) users had ever set up or operated an onion site, and whether users were aware of onion site phishing and impersonation. The last set of questions focused on users' general expectations of privacy and security when using onion services. We incorporated four attention checks to measure a respondent's *degree* of attention [3]. To ensure that participants felt comfortable answering questions, we did not make questions mandatory. The survey took about 15 minutes to complete.

Survey Testing. We used cognitive pretesting (sometimes also called cognitive interviewing) to improve the wording of our survey questions [6]. Pretesting reveals if respondents understand questions consistently and the way we intended them to be interpreted. Five pre-testers helped us iteratively improve the survey; after pre-testing and revisions, we launched the survey.

Recruitment. As with our interviews, we advertised our survey in a blog post on The Tor Project's blog [37], on its corresponding Twitter account, the CITP blog at Princeton, and on three Reddit subforums.[5] Unlike our interview participants, our survey respondents were self-selected. As with interview recruitment, we expect this recruitment strategy biased our sample towards engaged users because casual Tor users are unlikely to follow The Tor Project's social media accounts.

We did not offer incentives for participation because we wanted respondents to be able to participate anonymously without providing email addresses. Despite the lack of incentives, we collected enough responses. Our survey ran from August 16–September 11, 2017 (27 days).

Filtering and Analysis. Some of the survey responses were low-quality; people may have rushed their answers, aborted our survey prematurely, or given deliberately wrong answers. To mitigate these effects, we excluded participants who either did not finish the survey or who failed more than two out of four attention checks. We conducted a descriptive analysis on the survey data. We also computed correlation coefficients between every question pair in the survey, which did not yield significant results. We thus focus on results from the descriptive analysis. Each percentage is reported out of the total sample; we denote cases when survey participants chose not to respond as 'No Response'. Two researchers performed a deductive coding pass on the open-ended survey questions based on our interview codebook and held meetings to reach consensus on the final themes discussed. In rest of the paper, we denote survey participants with 'S'.

4.2.2 Participants

We collected 828 responses, but only 604 (73%) completed the survey, and 517 (62%) passed at least two attention checks. The rest of the paper focuses on these 517 responses. Table 2 shows the demographics of our survey. As we expected, respondents were young and educated: more than 71% were younger than 36, and 61% had at least a graduate or post-graduate degree. 44% percent also considered themselves at least highly knowledgeable in matters of Internet privacy and security.

[4] The full survey is available at https://nymity.ch/onion-services/pdf/survey-questions.pdf.

[5] https://reddit.com/r/tor/, https://reddit.com/r/onions/ https://reddit.com/r/samplesize/.

Age	#	%	Gender	#	%	Continent of residence	#	%	Education	#	%
18–25	2	11.8	Female	5	29.4	Asia	3	17.6	No degree	1	5.9
26–35	10	58.8	Male	12	70.6	Australia	1	5.9	High school	3	17.7
36–45	4	23.5				Europe	4	23.5	Graduate	3	17.7
46–55	1	5.9				North America	8	47.1	Postgraduate	10	58.8
						South America	1	5.9			

Table 1: *The distribution over gender, age, country of residence, and education for our 17 interview subjects. We do not show per-person demographic information to protect the identity of our interview subjects.*

Gender	#	%	Age	#	%	Education	#	%	Domain knowledge	#	%
Male	438	84.7	18–25	186	35.9	No degree	25	4.8	None	1	0.2
Female	49	9.4	26–35	180	34.8	High school	172	33.2	Mild	35	6.8
Other	25	4.8	36–45	87	16.8	Graduate	214	41.4	Moderate	178	34.4
No Response	5	1.0	46–55	43	8.3	Post graduate	102	19.7	High	227	43.9
			56–65	16	3.1	No Response	4	0.4	Expert	75	14.5
			> 65	3	0.6				No Response	1	0.2
			No Response	2	0.4						

Table 2: *The distribution over gender, age, education, and domain knowledge of the survey respondents. Providing demographic information was optional, so we lack data for some respondents.*

4.3 Domain Name Service (DNS) Queries

We analyzed `.onion` domains leaked via the Domain Name System (DNS) to better understand onion service usage and look for specific evidence of usability issues (*e.g.*, onion domains with typographical errors, phishing attacks). Although onion domains are only resolvable inside the Tor network, Internet users may attempt to access an onion site using a browser that is not configured to use Tor, resulting in the DNS query for onion domain "leaking" to conventional DNS resolvers—and ultimately to a DNS root server. Because all onion lookups to a conventional DNS server will result in a cache miss, all leaked onion lookups will ultimately go to a DNS root server. Thus, DNS root servers see a good sample of leaked onion domains. Our work builds on a previous analysis of a similar data set that was conducted several years ago and which was not focused on onion services specifically like our work [18, 33].

We obtained about several days of DNS data from the B root server through the IMPACT Cyber Trust program [34]. This data has several hundred `pcap` files, which contain full packet captures with pseudonymized IP addresses of all DNS traffic to the B root from September 19, 2017 10:00 UTC to September 21, 2017 23:59 UTC. We analyzed the DNS queries dataset and present our results alongside our findings from the survey and interview results. We extracted the QNAME of each DNS query, which yielded 15,471 correctly formatted onion domains that were 16 characters long (representing an 80-bit hash of the owner's public key) had has any letters of the alphabet and numbers between 2 and 7. These lookups, of course, may not always correspond to a real onion site, but they do reflect that some machine issued a DNS query for that onion domain for some reason.

4.4 Limitations

As we previously mentioned, we asked The Tor Project to disseminate our survey on its blog and Twitter account, which likely yielded the following biases.

Non-response bias. People who noticed our call for volunteers but decided against participating may have valued their privacy too much, falsely believed that their perspective is irrelevant, lacked time, or had other reasons not to participate. Nevertheless, non-respondents may exhibit traits that are fundamentally different from those who did participate.

Survivor bias. Our participants generally were able to tolerate Tor Browser's usability issues, which is why they are still around to tell their tale. We likely did not hear from people who decided that Tor Browser was not for them and were thus unable to tell us what drove them away. The danger of survivor bias lies in optimizing the user experience for the subset of people whose tolerance for inconvenience is higher than the rest.

Self-selection bias. Due to the nature of our online survey, participants could voluntarily select themselves into our set of respondents. These respondents may be unusually engaged, technical, and opinionated. Indeed, the demographic for our online survey in Section 4.2 was young and educated; perhaps Tor Browser's population is young and educated, as well, but we have no way of knowing.

5 Results

We organize the presentation of our findings by topic, including how users *perceive and use* (Section 5.1), *manage* (Section 5.2), and *wish to improve* (Section 5.3) onion

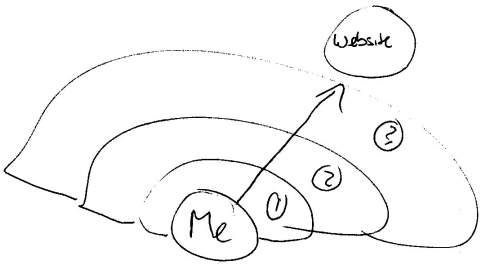


Figure 3: *A sketch of interviewee P03's mental model of onion services. The participant referred to several layers of protection.*

services. We interleave the results from our online survey with our interviews and domain name system data as appropriate.

5.1 Perception and Use

We first explore how users perceive onion site technology and why they use onion sites.

5.1.1 Incomplete mental models of onion services

We asked only our interviewees (not our survey participants) about their mental models of onion services because it is difficult to collect this type of information from a survey. This section thus presents results from the interviews only.

Perceptions of what an onion service is. We asked our interview participants how they defined an onion service, how they work, and what types of content and services they tend to host. Terminology was inconsistent and sometimes confusing: some interviewees referred to onion services as the dark web and others as hidden services. (Recall that The Tor Project only uses the term onion services). About half of our interviewees (9/17) knew that onion services enabled a user to access Web content anonymously. Six interviewees stated that onion services provide extra layers of protection, an idea that is well-illustrated in Figure 3,[6] and further elaborated on by participant P03: *"I think it's to do with the different hops that you build - different layers of making it difficult to find out who this person is."* Four interviewees stated that onion services work in a similar manner to Tor but with different encryption methods, which we can see on Figure 4. A minority of participants had sophisticated understanding: they referred to the encryption of data on the end points of a connection; three interviewees referred to the fact that last hop along the encrypted path corresponds to an onion link.

Perception of anonymity. Five interview participants drew the connection between Tor and onion services, stating that onion services have to be accessed through Tor

[6] All sketches are available online at https://nymity.ch/onion-services/mental-models/.

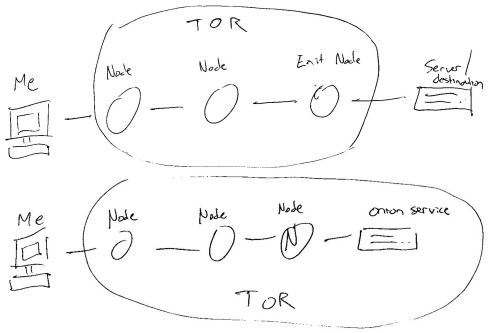


Figure 4: *Comparison of two sketches from interviewee P13. The first sketch shows the P13's mental model of Tor and the second one P13's mental model of onion services.*

browser but at least one did not see any connection between Tor and onion services. Only three interview participants knew that onion services do not only provide anonymity to the visitors to a website but also to the onion website provider themselves. In contrast to these interviewees who had some sense of what an onion service was, nearly half of our interviewees (8/17) were confused about how to define onion services, were unsure how onion services function or how to describe them, and did not understand how onion services protect them. Some of our interviewees did not distinguish disguising their IP address from disguising their real-world identity and instead used the umbrella term "anonymity" to refer to both concepts. This conflation of concepts paints an incomplete picture of the security and privacy guarantees that the Tor network provides, with only a few interviewees recognizing that anonymity is not completely achievable with Tor onion services: *"What's the point of going to Facebook using onion services when their business model is still about collecting your data?"* (P7). Other participants simply thought of onion services as P08 characterized them: *"[the] Internet without hyperlinks."* Some of our participants were not aware that onion services provide end-to-end security and self-certifying names. Syverson and Boyce explored how onion services can improve website authentication [32], but these benefits are difficult to convey to non-technical users, and even some experts advocated an "all or nothing" approach to online anonymity, overlooking important nuances.

The presence of a large quantity onion domains in the root DNS data corroborates prior studies that suggest either Internet users are attempting to visit an onion domain in a non-Tor browser indicating a misunderstanding of onion links, that browsers are loading content with onion links using pre-fetching, or that some web pages or malware are attempting to load resources from onion sites [18, 33].

Perceptions of what an onion service is used for. Interviewees had various perceptions of what onion services were used for or why they existed in the first place. In-

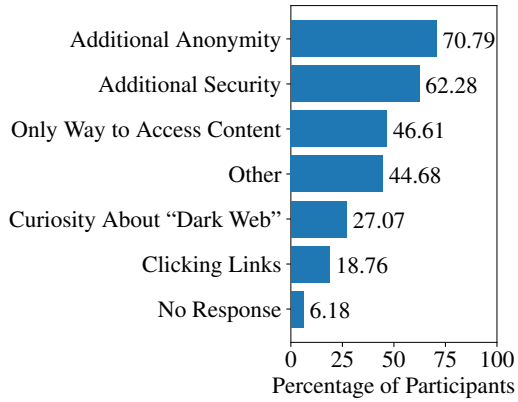


Figure 5: *Reasons for using onion services.*

terviewees sometimes associated onion services with illicit content such as the drug trade or credit card data sales (2/17) or felt that onion services may be the technology behind anonymous purchases. Similarly, as reported later in the paper, many survey respondents also voiced concern about illegal and questionable content on onion services, described by some as a "Wild West". Phishing sites, honeypots, and compromised onion sites further contribute to this perception.

5.1.2 Onion services used mostly for more anonymity

Usage. Our survey asked how often our respondents browse onion services. The usage frequency was almost uniformly distributed among our survey respondents; 24% use onion sites less than once a month, 22% use them about monthly, 25% weekly, and 23% daily. The remaining 6% had never used an onion service. We also asked our interviewees if they had used onion in the last three months; seven had and seven had not, with four of the latter group explaining that they had used onion services before, just not in the last three months. Only two interviewees had never used onion services before at all.

Anonymity and onion service content. The majority of our survey participants who used onion services did so because of the additional anonymity (71%) and the additional security (62%) (see Figure 5). For instance, six survey respondents commented on the onion domain format, indicating that they believed the seemingly-random characters in onion domains are the reason why onion services are anonymous: *"Onion services stay anonymous through changing their domain, and I feel that there is a possibility of decreased anonymity with a constant domain name." (S436).* These participants also believed that vanity domains are "less anonymous" because part of their domains is clearly not random. One survey participant (S454) further wrote: *"I understand vanity onion domains are a sign of the weakness of the hash algorithm used by the Tor network."*

Anonymity was also the main reason why our interviewees used onion services (6/17). Another reassuring factor for two of our interviewees was the feeling of security and safety that onion services provide. Furthermore, two interview participants thought of onion services as *"harm reduction technique."* P10 preferred to use Facebook's onion domain because it impedes tracking efforts. Additionally, 47% of survey respondents and three interviewees viewed onion services as the only way to access content they enjoy, making the use of onion services a necessity.

Non-browsing activities. Of our survey respondents who used onion services (485/517), 64% had these services for purposes other than web browsing. Several protocols such as the chat application Ricochet [4] and the file sharing application OnionShare [15] were purpose-built on top of onion services while existing TCP-based tools such as `ssh` can transparently use onion addresses instead of traditional IP addresses. Less than a quarter (21%) of our survey participants used onion services for non-browsing activities at least once a month such as remote login (ssh) or chat (IRC or XMPP). Our interviewees similarly mentioned using onion services to access Pirate Bay (1/17), Ricochet (1/17), TorChat (1/17), and OnionShare (1/17).

Work or personal reasons. Survey respondents who selected "Other" (45%) for onion service usage provided many reasons, including personal (18/517), with the most predominant personal reason being that an onion service gives a machine behind a network address translation (NAT) device a stable identifier and can be reached from any other user on the Tor network (there are other ways to achieve this goal, but for these users, setting up an onion service was the easiest way). Several interviewees used onion services to accomplish specific tasks. Five interviewees reported that they use onion services simply for their work, while four stated personal reasons, such as for a personal blog, or giving someone access to their home network. Two interview participants used onion services for educational purposes. P3 used onion services to help teach students about the dark web: *"I was teaching a class on Internet technology and regulations. We were basically showing students how Tor works and part of what I have to do as a teaching assistant was make students go and basically get to the moment where they either hire a hitman, buy drugs, or buy weapons. Just to show that it's possible. And then obviously we didn't buy it."*

Other survey respondents reported using onion services to reduce the load on exit relays, to do technical research, and to access sites that are otherwise unavailable. For instance, 7/517 used onion services for hosting a service, one survey respondent admitted using onion services for e-book piracy, two used onion services as an alternative

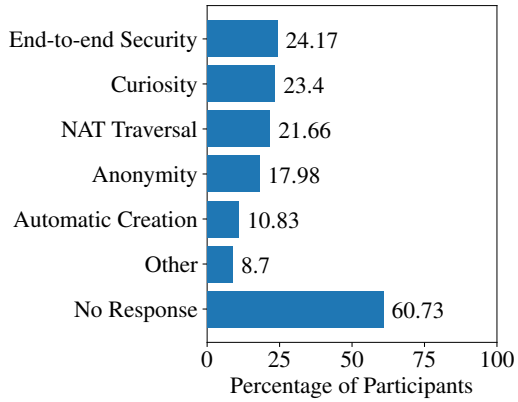


Figure 6: *Reasons for running onion services.*

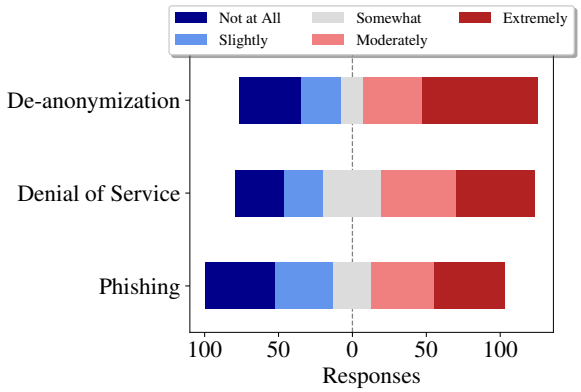


Figure 7: *Concerns of onion service operators about attacks.*

to a virtual private network and two used them to make their website as private and personal as they could.

Exploring the dark web. 27% of our survey respondents and two interviewees wanted to find out more about the dark web and onion domain content (3/517) as reasons to use onion services. Two interviewees used onion services for fun and social reasons—to *"toy around"* (P7) and also, as a way of spending time with friends, as well as to *"show off"* around them by using a technology unfamiliar to most users. Interestingly, 19% of survey respondents said that they use onion services for no particular reason but have clicked on onion links occasionally.

5.1.3 Onion sites operated for various reasons

Setting up an onion service. 39% of survey respondents had set up an onion service at some point. Of the respondents who had set up onion services of their own (266/517), 31% had run their onion service for private use while 21% had run them for the public. Figure 6 gives an overview of the reasons our respondents have for running onion services. For instance, the majority of those with onion services used them for end-to-end security, curiosity, or NAT traversal. Only 18% survey respondents had set up onion services for anonymity, such as to protect their visitors and provide security on their sites. In the open-ended responses, eleven survey respondents set up onion services because then their websites could be accessed from anywhere in the world, and seven survey respondents set up an onion service simply to test and learn how they work. Another two survey participants ran onion mirror sites to their personal websites, and at least one had an onion service as a backup website in case he lost control over his personal domain. Finally, at least two survey respondents set up onion for business purposes, work requirements, or to add valuable content to the onion community. In a similar vein, at least two interviewees spoke about setting up onion services or using onion services for work, such as to help Internet users upload leaked documents to their whistleblower website anonymously. In another example, P5 used onion services in the academic peer review process to allow authors to submit source code or supplementary material anonymously: *"If one of the other reviewers connects to our university site, and we have some sort of tracking information on there, we would be deanonymizing the reviewer. We put it on a Tor hidden service to make sure that the reviewer remains blind in academic review process."*

Phishing concerns. We inquired how concerned the survey respondents were about three potential attacks on their own onion services: (*i*) somebody setting up a phishing site for the operator's site, (*ii*) a denial-of-service attack, and (*iii*) a deanonymization attack. According to the results, shown in Figure 7, less than 8% of our survey respondents who operated an onion service were at least somewhat concerned about all of these attacks. Only a small percentage, 15%, claimed to be extremely concerned about somebody deanonymizing their onion service, 10% were extremely concerned about an onion site being taken offline, and only 9% were concerned about an onion site being impersonated for phishing purposes. Indeed, in the open-ended responses, we noted that several respondents lamented the difficulty of protecting onion services from application-layer deanonymization attacks. Matic *et al.* demonstrated some of these attacks in 2015 [17].

5.1.4 Varying trust in Tor and onion services

Our survey asked how safe our respondents feel when using Tor Browser and onion services, respectively. Figure 8 shows that onion services were actually perceived as less safe than Tor browser. 85% of survey respondents feel at least somewhat safe or very safe using Tor Browser as compared to only 66% of onion service users.

Reasons for trust. Survey responses indicated that par-

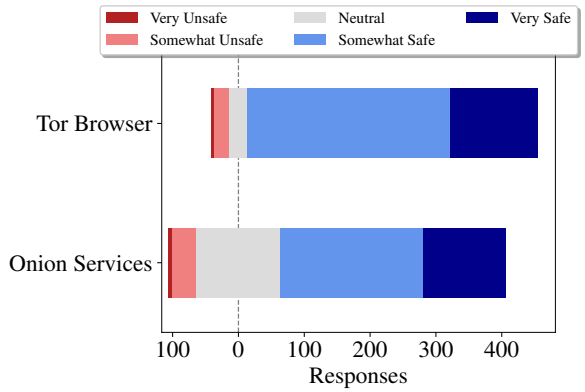


Figure 8: *Safety that respondents perceive when using Tor Browser and onion services.*

ticipants, most of whom (85%) rated themselves as non-experts (versus 15% self-rated experts) in knowledge about Internet privacy and security, lacked the ability to evaluate (or even understand) the Tor network's design which is why they deferred to expert opinion, their gut feeling, or the trust they place in Tor developers to gauge how much to trust these services. As S450 put it: *'There's a safety tradeoff. My connection to onion sites is more secure from outside eyes, but onion sites are more likely to be scams.'* With respect to onion services, the majority of survey respondents expressed that the added security and anonymity made them feel safe (117/517). Another factor contributing to the perceived security of onion services is that advertising companies are nowhere near as present on onion services as they are on the Web. 80/517 respondents trusted Tor and themselves to be safe on onion services while only a minority of interviewees were content and believed in the future of onion services (4/17) or placed their trust in them (2/17). Additionally, 30/517 participants said they would also choose onion services over regular websites because they trust them.

Reasons for distrust. 90/517 of survey respondents were skeptical of trusting onion services because of the possibility of phishing, the fact that onion services are hard to verify as authentic, and a concern that tracking can still occur even with onion services (59/517). Furthermore, at least 20/517 respondents said their trust of onion services would depend on the content of the services themselves. Some survey respondents did not have a clear understanding of onion services or thought they were the same as regular websites and reported as much (34/517).

Although our interviewees tended to see onion services as safer than corresponding websites (eight versus four participants), six participants felt that users should be careful when using onion services. Not all participants trusted onion services (5/17) and one expressed frustration such as P06: *"I'm pretty distrusting with most of the content I access over onion services. When I want content from a service, I tend to distrust it from the beginning."* Two interviewees mentioned that websites cannot identify you as the general advantage of onion services but at least three participants pointed out that websites actually can determine your identity if you write down your personal details as well as if you log in into any private accounts while using onion services. Similarly, 20 survey respondents also raised concerned and mentioned not wanting to log in to onion sites because they believe it defeats the purpose by revealing private data.

Moreover, one interview participant (P10) claimed that using onion links may influence the usability of their "normal" corresponding websites—the person shared a story in which they postulated that their Facebook account had been flagged for suspicious activity and then was deactivated because they had logged in through Tor Browser. These interview participants did not realize that while the company indeed knows who is logging in, it does not know Tor users' IP address or operating system.

5.2 Discovery and Management

We now explore how users discover and keep track of onion sites.

5.2.1 Discovering onion links is not straightforward

Recall that a freshly set up onion service is private by default, leaving it up to its operator to disseminate the domain. Established search engines such as Google are therefore generally inadequate to find content on onion services. Therefore discovering onion services is not as straightforward as with regular domains Figure 9 illustrates the results from our survey.

Social networking site and search engines. The three most popular ways that almost half of our survey participants discovered onion sites by were via (*i*) social networking sites such as Twitter and Reddit (48%), (*ii*) search engines such as Ahmia,[7] (46%) and (*iii*) randomly encountering links when browsing the Web (46%). Survey respondents who selected "Other" (16%) for how they discover onion links predominantly brought up independently-maintained onion domain aggregators. A noteworthy example is the Hidden Wiki used by 13 survey respondents, a community-curated and frequently-forked wiki that contains categorized links to onion services. At least 34 survey respondents searched for onion links on regular browsers and 18 of these respondents looked specifically at regular websites to see if they had

[7] Ahmia.fi is an onion site search engine that crawls user-submitted onion domains. It publishes the list of all indexed onion services at https://ahmia.fi/onions/.

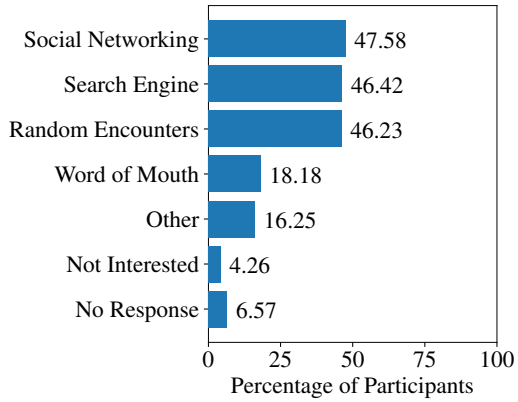


Figure 9: *Methods of discovering onion services.*

a corresponding onion link. In our interviews, two participants mentioned these techniques too. Between one to three survey respondents mentioned each of the following: using onion link lists generated by onion spiders, onion.torproject.org, ddg.onion, Imageboard, Google, and even Wikipedia.

We observed similar patterns in our interview respondents. Interviewees told us that they find onion links by word of mouth (6/17), using a search engine tool (5/17) including tools like DuckDuckGo (1/17), The Pirate Bay (1/17), Reddit (1/17), ahmia.fi (1/17), and the search widget in the Tor browser (1/17). More of our interviewees discovered onion services passively (6/17) by just happening to hear about or know about specific onion services while five interviewees told us that they looked actively for onion links, browsing for the content they needed.

Random encounters or word of mouth. A significantly less popular discovery mechanism was discovering links through word of mouth, which has the advantage that domains come from a trusted source (18% of survey respondents). 19/517 were frustrated that it was difficult to find out if a regular website had an onion service version even if they visited their website. Only 4% of our survey respondents—indicated that they were not interested in learning about new onion services because they only use their own sites (7/517). Similarly, two interviewees claimed that they never searched for new onion links.

Link discovery challenges. The majority of our survey respondents (55%) reported that they were satisfied with how they discover onion services but a significant proportion of our participants (38%) were not and 7% did not respond to this question. Those satisfied reported that they had no interest in learning about new onion services, in part because they only use a small set of onion services. Among the survey respondents who were not satisfied with how they discover onion services (38%), many (28/517) complained in the open-ended responses about link rot on aggregators where onion links were broken, unusable, or outdated. There is significant churn among onion sites, and our respondents were frustrated that aggregators are typically not curated and therefore link to numerous dead domains. The lack of curation also leads to these aggregators' containing the occasional scam and phishing site. The difficulty of telling apart two given onion domain names exacerbates this issue for users. 15/517 did not trust onion link lists because it is hard to validate if they are legitimate or not. 28/517 complained about filtering onion sites related to their interests with several wanting to avoid illegal and pornographic content, which is often difficult if the description is vague and the onion domain reveals nothing about its content. For this reason, 5/517 wished aggregators were more verbose in their description of onion sites.

Lack of good search engines. Many survey respondents complained about the lack of good search engines (33/517) and were not aware of search engines such as Ahmia. Among survey respondents who were aware of such engines, many were dissatisfied with both the search results and the number of indexed onion sites. Unsurprisingly, a *"Google for onion sites"* was a frequent wish. Similarly, one of the biggest issues for our interview participants was that onion sites are hard to find (5/17), or as P13 put it: *"How do you find stuff if you don't know what you're looking for or only have a vague idea?"* 10 survey respondents desired a better searching solution for onion services even with recognizing that this would be a tradeoff for security so services should have opt-in and opt-out options for discovery. As summarized by one survey respondent: *"Tor is still like the early 1990s Internet where websites were spread by word of mouth and by lists of links. In Tor, people publish lists of onion sites and I pick the ones I'm interested in. Every Tor search engine is poor and unreliable. Lists of links like Fresh Onions, while useful, often get out of date quickly, since many onion sites are unreliably hosted. Tor desperately needs a good search engine to find onion sites and ideally some way of identifying what those sites are about before clicking on them, since we lack that info in the URL."* (S339)

5.2.2 Saving and tracking onion links is difficult

Bookmarking links. Conventional domains are often easy to remember and recognize; most onion domains are random strings. We explored how users coped with this challenge. Most survey respondents (52%) use Tor Browser's bookmarks or a web-based bookmarking tool (3%) to save onion domains as seen in Figure 10. At least two interview participants reported bookmarking links as well. While convenient, this method of saving onion links

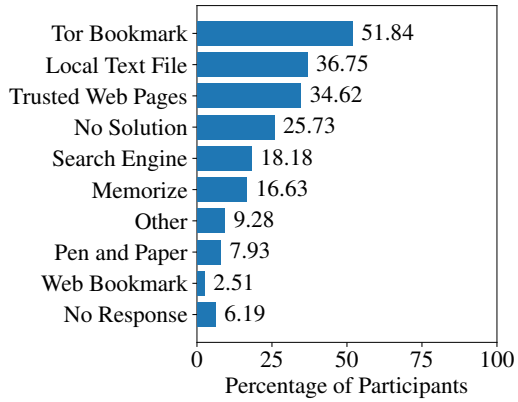


Figure 10: *Strategies to manage onion domains.*

leaves a trace of (presumably) visited sites on somebody's computer. One of Tor Browser's security requirements is "disk avoidance"—the browser must not write anything to disk that would reveal the user's browsing history [24, § 2.1]. Bookmarking links is a violation of this security requirement, albeit one that users seem to want.

Ad-hoc tracking methods. Somewhat less popular amongst our survey participants was saving onion domains in local text files (37%), getting them from trusted websites (35%), using search engines (18%), memorizing domains (17%), using some other techniques (9%), or employing pen and paper (8%). Of the 9% of our survey respondents who selected "Other", 15/517 stated that they store onion domains in an encrypted manner—either in a text file or in their password manager. Other techniques mentioned by only one or two survey respondents each included using auto-complete, storing them on a personal blog or using Twitter to find links, emailing the links to oneself, using redirect rules to automatically go to the .onion domain, storing the links in a virtual machine, or using Hidden Wiki. Four of our interviewees reported that they store onion services in a list and three remember (some) onion services. Other techniques for saving onion links mentioned by interviewees mirrored those of the survey and included using a Twitter feed to track onion links (1/17) and using TorChat as storage places for onion links (1/17). Moreover, one interviewee believed that Tor Browser remembers onion links and another interview participant (P1) explained: *"The onion services we run professionally we keep track of because we operate the server, so that's easy."* Notably, just over one-quarter of our survey respondents (26%) did not have a good solution to the problem of tracking onion links and similarly two interviewees pointed out that they lacked an onion link management mechanism.

Reaching onion domains quickly. We also asked our interviewees how they typically reach onion services. The most often mentioned technique was copy and pasting domains, done by four interviewees, followed by three interviewees who simply click on links they encounter. Two interviewees would go to onion sites using bookmarks while another two use Google to get to onion services. Only one interview participant told us that they typed the domains from their notes. Given the high number of (possibly insecure) home-baked solutions, a Tor Browser extension that solves the problem of saving and tracking onion links seems warranted.

5.2.3 Onion domains are hard to remember

Memorization reasons. Our participants often memorized onion domains to make it easier to visit onion sites and to minimize traces of their browsing habits. Of the survey respondents who memorize onion domains, we found that most respondents do no memorize any onion domains (60%) and less than a third (30%) memorize one to four onion domains. Only 3% can memorize more than four domains. Survey respondents who memorized domains (65% of all respondents) did so (*i*) automatically because of typing a domain many times (20%) (*ii*) to allow them to open an onion site more quickly (17%), and (*iii*) to ensure that they are visiting the correct site and not a phishing site (15%). Only 9% were privacy conscious and did so because bookmarking onion domains leaves a trace. 5% of the respondents gave other reasons for memorizing onion links. In these open-ended responses, 18 survey participants said that memorizing was simply easy for them, even unintentional. Among these participants, there were only 8/517 that specifically mentioned the Facebook onion site as very easy to remember. Only a few survey respondents (3/517) did not memorize onion sites at all.

Memorization challenges. Our interview participants generally found onion domains problematic in terms of having to remember random strings of letters and numbers. Four interviewees perceived onion domains as too long. Among these participant was one who further complained about random characters in onion domains. At least two interviewees criticized onion links for being hard to remember. This viewpoint was echoed in our survey, where participants rated URLs such as expyuzz4wqqyqhjn.onion and torproz4wqqyqhjn.onion as harder to remember because the *"numbers make the names harder to remember."* Other survey respondents stated that vanity domains are easier to remember when they can be pronounced as described in the example quote by survey respondent (S46): *"phonetic pronunciation plays a large part in how I remember onions."* Many other survey respondents stated that onion domains that are supported by a mnemonic are also easier to remember; we elaborate on this result in Section 5.2.4.

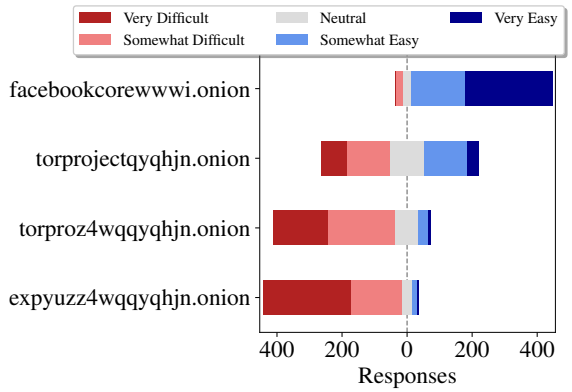


Figure 11: *Expected difficulty memorizing four onion domains.*

5.2.4 Vanity domains: more memorable, less trusted

Memorizability. The majority of our survey respondents appreciated vanity domains because they were easy to remember (64%) and easy to recognize (64%), and they provided a unique *"branding"* (34%). Some survey respondents indicated that a vanity prefix—like a traditional domain—informs about an onion service's content, letting visitors know what to expect and thus preventing unpleasant surprises but at least 3/517 wanted more clues to let visitors know more about what the domain content is or for some content to be harder to find. As S423 wrote: *"For less important, high traffic sites (social media like Facebook), it's okay. For sites handling much more sensitive/potentially illicit content, its a good idea to make it difficult to find."*

Only 15% did not have an opinion about vanity domains, 8% reported that they disliked vanity onion domains, and 7% did not see a benefit of vanity domains. We asked survey respondents about whether or not they memorize vanity domains—specifically facebookcorewwwi.onion—and how difficult they find it to memorize onion domains of differing levels of vanity. Only 20% of respondents replied that facebookcorewwwi.onion is among the sites that they have memorized. This is because it is "easy to memorize" (S391) and "after seeing [it] many times, I automatically start to memorize it."(S94) Depending on the format of the vanity domain, our survey respondents expressed differing levels of ease for memorizing them; these results are shown in Figure 11. Most participants found it easier to memorize vanity domains with a longer recognizable prefix such as Facebook's. Interestingly, only 4/517 survey respondents considered vanity domains economically unfair because wealthy entities can afford to generate longer prefixes such as Facebook.

Usable links. Ten out of seventeen interviewees saw vanity domains as a significant usability improvement to the regular onion domains: *"In terms of mnemonics and easier recollection if you can chunk words that are associated with daily life and not just a random. If there's entropy in the stream, there's no way I'm going to remember more than a few characters"* (P18). P10 had a different perspective that suggested these vanity domains make onion services more usable: *"I think that for people who don't spend a lot of time using those types of services, it definitely gives you a more familiar framework for thinking about where you are on the Internet. If people think . . . people have a pretty strange geographic metaphors for navigating the Internet, but I think this idea of where are you? Well, I'm at this place I can't even name, I can't say it out loud, I think that can be a barrier for people."*

Phishing and security. If users focus on the vanity part of a domain only, attackers can create an similar domain that features the original's prefix but differs in subsequent characters. Nurmi [23] and Monteiro [19] have both documented such an attack, but its effectiveness is not known.

Indeed, in several cases, both survey (29/517) and interview participants found that vanity domains were not practical and seemed to distrust them because they felt they made phishing easier: *"I don't think it's useful because . . . it's followed by another random word . . . and phishing can still copy that . . . I don't think what I can remember is safe now."* (P17). Similarly, as S94 explained: *"We also get false expectations of security from such domains. Somebody can generate another onion key with same facebookcorewwwi address. It's hard but may be possible. People who believe in uniqueness of generated characters, will be caught and impersonated."*. Among our survey respondents, there was also concern that the short and recognizable prefixes tempt users to verify only the prefix and ignore the non-vanity part of the onion domain, as epitomized by one survey respondent: *"I only memorize the first part of the domain."* (S96) while another wrote: *"If there isn't some cognizable word at the start, it'll be more difficult for me to determine if I'm going to the correct domain or a scam. I may end up going to less onion sites as a result."* (S355)

This viewpoint was echoed by our interview participants, who noticed that vanity domains can negatively affect security. P13 explained: *"I think in theory, on the one [hand], it makes it easier for you to recognize where you are, it makes it easier for you to perhaps, share the URL or type it out. On the other hand, I've seen concerns that, by having a vanity URL where perhaps people only look for the Facebook portion and they don't pay attention to what comes after it could potentially make it easier to exploit unsuspecting users. Send them a link that also says Facebook but the numbers after it are different, but you just see the Facebook part and go, 'It's fine, it's Facebook.' That can be a risk to them."* P5 also shared

their view on vanity domains: *"It seems like it would encourage more trust on behalf of the user, but then again, maybe make phishing easier too, if phishers are making vanity domains themselves. Yeah, that seems like it could go both ways actually."*

5.2.5 Onion sites are hard to verify as authentic

Verification techniques. We asked our participants about verifying the authenticity of an onion site. The majority of our survey respondents (79%) did want to verify an onion service as authentic. Figure 12 gives an overview of the strategies that our respondents employ. Most of the respondents (64%) copied and pasted onion links from trusted sources (*e.g.*, friends or another, trusted website) or used bookmarks when revisiting onion services (52%). Many survey respondents also verified the domain in the browser's address bar (45%), checked if the corresponding website had a link to its onion site (40%), or checked that the onion service has a valid HTTPS certificate (36%).[8] Survey respondents reporting checking the corresponding regular website for verification, verifying if familiar images were recognized, or checking for HTTPS (9/517). 8/517 only used links if received form a trusted resource or trusted member of a community or check with their notes (4/517). 5/517 trusted their perception of a website as verification of authenticity or Tor or the fact that onion sites are self-certified by design (3/517) or use the fact that they could log into a site as verification (5/517). Only a few mentioned using multiple sources to verify authenticity (3/517) and at least 9 survey respondents said that they did not use onion links at all.

When asked how many characters our survey respondents verify in onion domains, 19% verified thirteen to sixteen digits, *i.e.*, (almost) the full domain, while 20% verified up to nine digits, which is within the realm of brute force attacks, and 5% verified between nine to twelve digits. More than half of respondents provided no response at all (54%).

For those interviewees (7/17) who did attempt to ensure they were visiting an authentic onion site, we observed two strategies: relying on someone else to ensure a link was authentic and trying to work out authenticity using various techniques on their own. Most interviewees in the first group stated that they rely on word of mouth for verification (5/17), followed by assistance from someone else (4/17). P3 explained *"[I] let people show me them. I don't go there myself."* Two interview participants relied on resources they already trusted for onion links, like friends and other communities and two accessed onion services by first visiting their corresponding

[8] DigiCert is issuing EV certificates for onion sites [7], but adoption has been slow—presumably in part because EV certificates require the CA to verify the applicant's identity and they are not free.

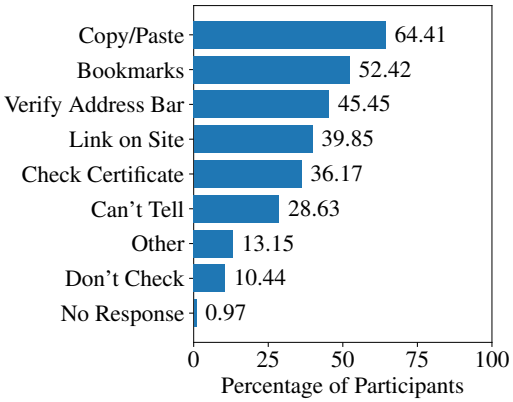


Figure 12: *Determining an onion service's legitimacy.*

publicly available websites if they could to verify authenticity. One of the most common approaches in the second group (3/17) was to check and compare URLs to see whether they matched to a *"clearnet site"* (P14), its unencrypted version on the regular Internet. Furthermore, two interview participants rely on their own experience, one on HTTPS certificates, and another one would lower the security settings in Tor Browser using the security slider to check the website more thoroughly: *"Sometimes, it worries me, but before that I access, in Tor, I turn off, I always. First, I always turn off the Java service and etcetera, to check the website. I think it's good, then I will lower the security level in Tor browser, but mostly, I will ask anything, maybe, in the Reddit or in the forum—in my country forum—of what the service [may be]."* (P17). One interviewee believed that just using Tor is verification in itself and another participant avoided onion sites altogether.

Verification challenges. Indicative of potential security issues, 29% of survey respondents stated that they sometimes could not tell the difference between an authentic service and an impersonation, and 10% never checked a service's legitimacy in the first place. Survey participants who selected "Other" (13%) provided a wide variety of ad-hoc verification strategies, further highlighting the importance of being able to verify a site as being the one that they were trying to reach. For instance, 13 survey respondents said there is no good way of verifying onion services or they do not know how to.

We also asked our interview participants how they knew that the site they went to was the one that they wanted to visit. Similar to the survey respondents, six interviewees reported that they did not know how to verify the authenticity on an onion site and they were concerned about being on an impersonating website because it is easy to mistype onion domains and onion domains change frequently if an onion service is short-lived or moves. P1 summarized the issue as being inherent to the nature of

onion services *"I wouldn't know how to do that, no. Isn't that the whole point of onion services? That people can run anonymous things without being to find out who owns and operates them?"* Two interviewees even believed onion site authentication to be impossible. For this reason, some interviewees also proposed that onion domain formats without numbers or with a stable patterns of letters and numbers could potentially make sites easier to reach and verify for authenticity.

5.2.6 Onion lookups suggest typos or phishing

Phishing remains an issue despite onion services' extra anonymity and security properties. Past work has documented phishing onion sites that transparently rewrote Bitcoin addresses to hijack Bitcoin transactions [19, 23, 38]. Key to this attack is the difficulty of telling apart an authentic onion domain from an impersonation. For conventional domains we rely on EV certificates, browser protections, search results, and long-lived reputation, but none of these methods have matured for onion services. Does the nature of onion services facilitate phishing attacks? If so, what can we do to mitigate the issue?

Most interview participants (9/17) agreed that phishing constitutes a serious risk, one of them explained the phenomenon this way: *"the two approaches I know from the normal Web still apply here, which is typo-squatting, registering an onion [domain] that's only a slight variation away, or bit-squatting, which is slightly different, but it involves a single or a few bit flips within an onion address, so that it looks relatively similar"* (P6), while another interview participant presented their solution to this problem: *"If you're manually typing it in I suppose they could be a problem, but I primarily cut and paste"* (P16).

We evaluated how often lookups to two different onion domains are extremely similar to one another, which can shed light on how often an onion domain may be phished, since it is unlikely for distinct onion services to have extremely similar strings for onion domains.

To do so, we computed the Jaro-Winkler similarity metric between each unique pair of correctly formatted onion domains, which is the edit distance between two strings that gives more weight to strings with common prefixes. We used this metric because people tend to check the first part of the domain. Values range between $[0, 1]$, where 0 represents completely different strings and 1 represents matching strings, to each unique domain pair. We find that 0.007% (8,672) of all unique domain pairs (119,668,185) have an extremely high similarity ($> .90$); for example, `bitfog2jzic5tnh7.onion` and `bitfog2y7y2pfv75.onion` have a Jaro-Winkler similarity of 0.917.

We first analyzed the results of the similarity metric for any well-known vanity domains. We found that Facebook's onion site (`facebookcorewwwi.onion`) has a similarity score of 0.953 with another onion domain that was looked up `facebookizqekmhz.onion`, which only appeared in our dataset twice (in comparison to the 101 instances of `facebookcorewwwi.onion`). Another frequently looked up onion domain is `blockchainbdgpzk.onion`, which is a popular Bitcoin wallet; it was extremely similar to blockchatvqztbll.onion (similarity score 0.949). These cases of similar domains could be a potential indicator of phishing sites for popular domains.

Onion 1	#	Onion 2	#	J-W
57g7spgrzlojinas	1,621	57g7spgrziojinas	14	0.989
xxlvbrloxvriy2c5	1,593	xxlvbrioxvriy2c5	4	0.949
gx7ekbenv2riucmf	1,476	gm7ekbenv2riucmf	4	0.973
mischapuk6hyrn72	1,062	mischa5xyir2mrhd	8	0.902
petya3jxfp2f7g3i	1,061	petya3jxfb2f7g3i	8	0.997
petya3jxfp2f7g3i	1,061	petya37h5tbhyvki	58	0.907
mischa5xyix2mrhd	786	mischa5xyir2mrhd	8	0.999
hydraruzxpnew4af	529	hydraruzxpnew1af	2	0.999
hydraruzxpnew4af	529	hydraruehfq5poj5	2	0.927
hydraruzxpnew4af	529	hydraruzxpnew3af	2	0.999
3g2upl4pq6kufc4m	472	tg2upl4pq6kufc4m	2	0.971
3g2upl4pq6kufc4m	472	3g2upl4t5houfo4y	2	0.924
3g2upl4pq6kufc4m	472	3g2upl4oq6kuc4mm	2	0.954
3g2upl4pq6kufc4m	472	3g2upl4pe3kcf24d	2	0.973
zqktlwi4fecvo6ri	410	zqktlwipcfe3siu2	2	0.931
zqktlwi4fecvo6ri	410	zqktlwi4i34kbat3	12	0.946

Table 3: *The Jaro-Winkler similarity score for frequently visited onion domains in the DNS root dataset.*

We next explored the top 20 most frequently requested onion domains dataset by checking: whether they are extremely similar to another onion domain in our dataset, and whether there is a large difference in frequency of the two similar domains. Of the top 20 onion domains, 16 had a Jaro-Winkler similarity score > 0.90 with at least one other onion domain in the data. Table 3 shows the characteristics of these domains. Many of the domains in the table under "Onion 1" are associated with either the WannaCry Ransomware, the Mischa Ransomware, or the Petya Ransomware. The remaining domains in that column are real onion domains that returned search results when used as input to `https://ahmia.fi`; these include a Russian Market (`hydraruzxpnew4af.onion`), DuckDuckGo (`3g2upl4pq6kufc4m.onion`), and The Hidden Wiki (`zqktlwi4fecvo6ri.onion`).

5.3 Areas for Improvement

When we asked about areas for improvement in the survey and interviews, participants told us that onion services could be enhanced technically and performance-wise, and that privacy and security, educational resources on, and methods for discovering onion content could be improved.

Technical Improvements. In our open ended question on improvements to onion services, 43/517 did not provide

an answer and 36/517 expressed their gratitude for Tor and Torproject and were satisfied with the service overall. However, many respondents spoke of possible enhancements. The majority of survey respondents (59/517) mentioned technical improvements they would like to see for onion services such as improving support for Javascript, making onion services available in other browsers, and having more support for mobile devices. 17/517 wanted a better user interface and user experience with onion services in general. Our interviewees also mentioned various technical improvements they would like to see in onion services. Two wanted a secure bookmarking tool and another interviewee wanted CAPTCHAs to be gone (these are triggered more often with onion services). Only four talked about wanting to see influential websites or even all websites set up corresponding onion sites.

Performance Concerns. At least 48 survey respondents had performance concerns about onion services. For example, one survey user stated, *"I would always prefer the onion site but for video sites like YouTube I would likely often use the normal site to be able to get a higher quality stream due to higher bandwidth."* (S435) Three interview participants similarly raised the "slowness" of onion services.

Privacy and Security. 34 survey participants expressed concern about anonymity and security issues and would like to feel and be safer over the Tor network more generally. For instance, S70 wrote: *'I hear a lot of social media questions from casual or unsophisticated users, and the single biggest problem is that they don't have the slightest idea of exactly what's being protected and what isn't. Vague pronouncements that "doing X is safer" don't help. Tor needs to stop being muddy in explaining what it protects, and stop promoting itself to people who don't understand what it can and can't do for them.'* 11/517 complained about lack of anonymity protection specifically from government, big companies or even Federal Bureau of Investigation (FBI). 8/517 wanted to verify onion services as legitimate or live and only 2/517 spoke about not wanting the dark net to contain criminal content.

Education and Resources. 24 survey respondents believed that there was a "'knowledge" issue with not enough resources and documentation for newcomers to Tor and onion services. Many of our interviewees felt similarly (7/17). Interviewees lamented about a lack of documentation or resources that would allow newcomers to learn more about onion services. P8, for example, wanted to know how to use onion services correctly and stop being uncertain about its properties: *"Really clear user education in the installation process would be great for people like me . . . who are like 'Okay, this is a thing I can use, why am I using it again? What am I using it for? What does it do?"* Three of our interviewees also referred to the lack of proper education as *"cultural mysticism."* Uneducated users often misunderstand concepts, as P10 explained: *"The perception that these are hardcore security tools sometimes signals to ordinary users that they are also difficult or badly designed or complicated to use, and that's not really the case with Tor."* Even if knowledge was not an issue, fear of consequences may deter users otherwise, as P8 mentioned before: *"Because it's also super scary. You think you're playing with this spy thing . . . Sometimes it's actually a really simple technical thing that's not terrifying. And to demystify those things would be really nice."*

Improved Search. 15/517 survey respondents wanted onion services to be more accessible, such as via a good search engine or organized database. At least four interviewees also desired improved search engines. As an example of this sentiment, S116 wrote: *'Ask someone to develop a really good search engine so that sites may be found. I am sure that the dark net has to be more than a few illicit sites that are selling stolen credit cards, and running Bitcoin scams. I feel like when I browse the dark net, I am floating in space waiting for another planet to suddenly appear. Whatever content is out there needs to be discovered, lest people will make misinformed judgments about the dark net. The dark net should be understood to be preeminently about privacy, not criminality.'* In addition, many survey respondents expressed frustration about the difficulty of finding out if a particular public website has a corresponding onion service. A common wish was to have a website list its onion service prominently in a footer or on the corresponding Internet site (3/517). Ironically, some survey respondents were surprised that `torproject.org` has a corresponding onion site—they could not find it on the website.

6 Future Directions

Our work highlights several opportunities for improvements to current onion services.

Security indicators for onion services. First, many of our participants had an incomplete mental model of how onion services work and trusted them less than other Tor services, which suggests that a better indicator of the protections an onion service offers should be made visible to onion service users. Currently, The Tor Project is working on a security indicator for onion services [1]. Figure 2b illustrates that Tor Browser currently, in version 7.0.10, displays an onion service connection as an insecure HTTP connection, thus greatly "under-selling" the security and privacy that an onion service connection provides. The design process for such indicators should evaluate whether users understand the meaning of the indicator, as well as how it differs from an HTTPS indicator.

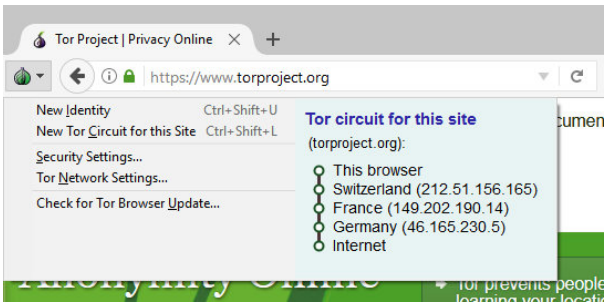


Figure 13: *A click on the onion icon reveals the Tor relays that constitute the circuit that was used to fetch the current page. As of February 2018, the user interface is subject to a redesign [2].*

(Felt *et al.* found the subtleties that one must consider when designing similar security indicators [8].)

The Tor Browser's circuit display interface is also being redesigned (see Figure 13) [2]. As with an onion service indicator, an evaluation of the circuit display could reveal user misunderstandings that may improve perceptions of and trust in onion services. For example, we found that some users are not familiar with the concept of guard relays and incorrectly expect each relay in their circuit to change, which suggests the need for an improved interface. Users also found it difficult to verify the authenticity of an onion site; while certificates do help, many sites still do not have them, and some may never have them.

Automatic detection of phishing onion domains. Our findings that some onion domains in the root DNS data have small edit distance to popular onion domains suggests that users may fall victim typos to phishing attacks; on the other hand, because the number of popular onion domains is still relatively small and (through our analysis and previous work [18, 33]) relatively well-known, the Tor Browser could raise an alert when the user attempts to access an onion domain that has a small edit distance to a popular onion domain.

Opt-in publishing of onion sites. Our participants often wanted more services to be available as onion services and did not often know if an onion service for a popular website existed. Participants found it difficult to discover new onion services, which suggests the need for better ways to find active onion services. While search engines and curated lists do exist, they do not generally allow users to locate an onion service of interest without also stumbling upon unwanted content. One possibility is an opt-in public log, whereby users can learn about new onion domains as they are added. Many participants also expressed interest in a browser feature that could automatically "upgrade" from a regular web site to its corresponding onion service. (The Tor Project is currently investigating this problem space [13].)

Privacy-preserving onion bookmarking. Participants found it difficult to track and save onion links; they often resorted to memorizing links to avoid security issues with storing onion links. This problem suggests the need for a privacy-preserving bookmarking tool that allows users to bookmark sites without leaving a trail in their browser storage or elsewhere on their system.

7 Conclusion

Onion services resemble the 1990s web: Pages load slowly, user interfaces are clumsy, and search engines are inadequate. Users appreciate the extra security, privacy, and NAT punching properties of onion services, which gives rise to a variety of use cases. Yet, users are confronted with a variety of privacy, security and usability concerns that should be addressed in future generations of onion services. For example, users are concerned about the susceptibility of onion domains to phishing attacks, and the onion domains that are leaked to the public Internet illustrate that this threat is real—and unaddressed. Users have limited ways of discovering the existence of onion services, let alone navigating to them.

A range of design improvements, from better discovery mechanisms to automatic "upgrading" to a corresponding onion service when it is available are initial steps to improve usability. Some of these desired features have clear analogs in the public Internet, such as the padlock icon as a security indicator for HTTPS, and HTTP Strict Transport Security (HSTS) to automatically upgrade an HTTP connection to HTTPS. We expect that many of the usability design lessons from the public Internet may in some cases also apply to onion services.

Acknowledgments

This research was supported by the National Science Foundation Awards CNS-1540066, CNS-1602399, and CNS-1664786. We thank George Kadianakis for helpful feedback on our survey questions, Katherine Haenschen for helping us improve our method, Mark Martinez for conducting interviews, Stephanie Whited for helping us disseminate our survey, and Antonela Debiasi for informing us about current user experience efforts around the Tor Browser. We thank Roya Ensafi, Will Scott, Jens Kubiziel, and Vasilis Ververis for pre-testing our survey, and USC's Information Sciences Institute for access to the DNS B root data. We also thank the Tor community for helpful feedback, for volunteering for our interviews, and for taking our survey.

References

[1] I. Bagueros. Communicating security expectations for .onion: what to say about different padlock states for .onion services. https://bugs.torproject.org/23247.

[2] I. Bagueros. Improve how circuits are displayed to the user. https://bugs.torproject.org/24309.

[3] A. J. Berinsky, M. F. Margolis, and M. W. Sances. Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American Journal of Political Science*, 58(3), 2014. http://web.mit.edu/berinsky/www/files/shirkers1.pdf.

[4] J. Brooks. Ricochet. https://ricochet.im.

[5] J. Clark, P. C. V. Oorschot, and C. Adams. Usability of anonymous web browsing: An examination of Tor interfaces and deployability. In *SOUPS*. ACM, 2007. https://www.freehaven.net/anonbib/cache/tor-soups07.pdf.

[6] D. Collins. Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12(3), 2003. https://link.springer.com/content/pdf/10.1023%2FA%3A1023254226592.pdf.

[7] DigiCert. Ordering a .onion certificate from DigiCert, Dec. 2015. https://www.digicert.com/blog/ordering-a-onion-certificate-from-digicert/.

[8] A. P. Felt, R. W. Reeder, A. Ainslie, H. Harris, M. Walker, C. Thompson, M. E. Acer, E. Morant, and S. Consolvo. Rethinking connection security indicators. In *SOUPS*. USENIX, 2016. https://www.usenix.org/system/files/conference/soups2016/soups2016-paper-porter-felt.pdf.

[9] A. Forte, N. Andalibi, and R. Greenstadt. Privacy, anonymity, and perceived risk in open collaboration: A study of Tor users and Wikipedians. In *CSCW*. ACM, 2017. http://andreaforte.net/ForteCSCW17-Anonymity.pdf.

[10] K. Gallagher, S. Patil, and N. Memon. New me: Understanding expert and non-expert perceptions and usage of the Tor anonymity network. In *SOUPS*. ACM, 2017. https://www.usenix.org/system/files/conference/soups2017/soups2017-gallagher.pdf.

[11] A. Johnson. A proposal to change hidden service terminology, Feb. 2015. https://lists.torproject.org/pipermail/tor-dev/2015-February/008256.html.

[12] G. Kadianakis, Y. Angel, and D. Goulet. A name system API for Tor onion services, 2016. https://gitweb.torproject.org/torspec.git/tree/proposals/279-naming-layer-api.txt.

[13] L. Lee. .onion everywhere?: increasing the use of onion services through automatic redirects and aliasing. https://bugs.torproject.org/21952.

[14] L. Lee, D. Fifield, N. Malkin, G. Iyer, S. Egelman, and D. Wagner. A usability evaluation of Tor launcher. *PoPETS*, 2017(3), 2017. https://petsymposium.org/2017/papers/issue3/paper2-2017-3-source.pdf.

[15] M. Lee. OnionShare. https://onionshare.org.

[16] N. Mathewson. Next-generation hidden services in Tor, 2013. https://gitweb.torproject.org/torspec.git/tree/proposals/224-rend-spec-ng.txt.

[17] S. Matic, P. Kotzias, and J. Caballero. Caronte: Detecting location leaks for deanonymizing Tor hidden services. In *CCS*. ACM, 2015. https://software.imdea.org/~juanca/papers/caronte_ccs15.pdf.

[18] A. Mohaisen and K. Ren. Leakage of .onion at the DNS Root: Measurements, Causes, and Countermeasures. *IEEE/ACM Transactions on Networking*, 25(5):3059–3072, 2017.

[19] C. Monteiro. Intercepting drug deals, charity, and onionland, Oct. 2016. https://pirate.london/intercepting-drug-deals-charity-and-onionland-a2f9bb306b04.

[20] A. Muffett. 1 million people use Facebook over Tor, Apr. 2016. https://www.facebook.com/notes/facebook-over-tor/1-million-people-use-facebook-over-tor/865624066877648/.

[21] G. Norcie, J. Blythe, K. Caine, and L. J. Camp. Why Johnny can't blow the whistle: Identifying and reducing usability issues in anonymity systems. In *USENIX*. Internet Society, 2014. https://www.freehaven.net/anonbib/cache/usableTor.pdf.

[22] J. Nurmi. Ahmia – search Tor hidden services. https://ahmia.fi.

[23] J. Nurmi. Warning: 255 fake and booby trapped onion sites, June 2015. https://lists.torproject.org/pipermail/tor-talk/2015-June/038295.html.

[24] M. Perry, E. Clark, S. Murdoch, and G. Koppen. The design and implementation of the Tor Browser, Mar. 2017. https://www.torproject.org/projects/torbrowser/design/.

[25] E. S. Poole, M. Chetty, R. E. Grinter, and W. K. Edwards. More than meets the eye: Transforming the user experience of home network management. In *Proceedings of the 7th ACM Conference on Designing Interactive Systems*, DIS '08, pages 455–464, New York, NY, USA, 2008. ACM. http://doi.acm.org.proxy-um.researchport.umd.edu/10.1145/1394445.1394494.

[26] Sai and A. Fink. Mnemonic .onion URLs, Feb. 2012. https://gitweb.torproject.org/torspec.git/tree/proposals/194-mnemonic-urls.txt.

[27] Y. Sawaya, M. Sharif, N. Christin, A. Kubota, A. Nakarai, and A. Yamada. Self-confidence trumps knowledge: A cross-cultural study of security behavior. In *CHI*. ACM, 2017. https://users.ece.cmu.edu/~mahmoods/publications/chi17-cross-cultural-study.pdf.

[28] M. Schanzenbach. The GNU name system, 2012. https://gnunet.org/gns.

[29] I. Seidman. *Interviewing As Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. Teachers college press, 2013.

[30] E. Swanson. Scallion: GPU-based onion hash generator. https://github.com/lachesis/scallion.

[31] P. Syverson. Onion routing: Brief selected history, 2005. https://www.onion-router.net/History.html.

[32] P. Syverson and G. Boyce. Genuine onion: Simple, fast, flexible, and cheap website authentication. In *Web 2.0 Security & Privacy*. IEEE, 2015. https://www.ieee-security.org/TC/SPW2015/W2SP/papers/W2SP_2015_submission_27.pdf.

[33] M. Thomas and A. Mohaisen. Measuring the leakage of onion at the root: A measurement of Tor's .onion pseudo-TLD in the global domain name system. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 173–180. ACM, 2014.

[34] University of Southern California—Information Sciences Institute. B root traffic for DITL, 2017. https://impactcybertrust.org/dataset_view?idDataset=814.

[35] J. Victors, M. Li, and X. Fu. The Onion Name System. *PoPETS*, 2017(1), 2017. https://www.degruyter.com/downloadpdf/j/popets.2017.2017.issue-1/popets-2017-0003/popets-2017-0003.pdf.

[36] S. P. Weber. mnemonicode, 2017. https://github.com/singpolyma/mnemonicode.

[37] P. Winter. Take part in a study to help improve onion services. https://blog.torproject.org/take-part-study-help-improve-onion-services.

[38] P. Winter, R. Ensafi, K. Loesing, and N. Feamster. Identifying and characterizing Sybils in the Tor network. In *USENIX Security*. USENIX, 2016. https://nymity.ch/sybilhunting/pdf/sybilhunting-sec16.pdf.