Shining Light on the Dark Web

George Hurlburt, STEMCorp

Traditional endpoint protection will not address the looming cybersecurity crisis because it ignores the source of the problem—the vast online black market buried deep within the Internet.

On 6 August 1991, almost a year and a half after proposing the World Wide Web (WWW), physicist and computer scientist Tim Berners-Lee published the first website from his computer at CERN. Today, more than a billion websites continue to fuel this highly disruptive and transformative technology (www.internetlivestats.com/total-number-of-websites).

To many, the WWW is a wondrous place. We follow the news and weather online, often in near real time. Streaming videos entertain us or show us how to fix that pesky leaking faucet. We gather data from Wikipedia and other knowledge banks to answer questions. We keep in touch with distant loved ones and close friends via social media. We increasingly rely on the WWW to purchase goods and services—to the point that the global ecommerce market, estimated to reach $2.3 trillion this year,[1] is putting traditional retailers at risk of obsolescence.

The WWW resides on the Internet, whose transformative effect is far more immense. In January 1983, the Internet became a reality when the Arpanet switched from the Network Control Protocol (NCP) to the Transmission Control Protocol/Internet Protocol (TCP/IP). Just as the US Department of Defense spun off the operational Milnet from the research-oriented Arpanet, many other purposeful networks arose. Ultimately, the proliferation of networks led to the classification of networks as national (Class A), regional (Class B), or local (Class C) in scope. As the number of Internet hosts grew, the Domain Name Service (DNS) permitted scalable resolution of hierarchically organized host names to workable Internet addresses. Today's Internet has 3.5 billion users, which accounts for almost 45 percent of the world's 7 billion people (www.statista.com/topics/1145/internet-usage-worldwide). It subsumes many networks, only a fraction of which can be seen by the average WWW user.

BEYOND THE WWW: THE DEEP WEB AND DARK WEB

The openly searchable Internet, including the entire WWW, comprises only 6–10 percent of the whole Internet.[2] The remaining 90–94 percent holds content that is neither indexed nor cataloged. Much of this private data includes holdings on corporate Class B and C internal networks (intranets), email and/or databases, academic journals, or individually held information. This region,

0018-9162/17/$33.00 © 2017 IEEE

TABLE 1. Deep Web access control techniques.

Deep Web content	Access control techniques
Private Web	Private websites, such as intranets with public-facing webpages, typically require registration and login using password or other authentication mechanisms to gain access to the private, more protected, side of the website.
Contextual Web	Elements of value can be discovered through a history of navigation across websites with common contextual threads. For example, a unique semantic identity can be linked to an individual based on that person's online activity.
Dynamic content	A dynamic page appears as a hidden response to a specific query or through submission of a specific element or set of elements on a form. In either case, the resulting text fields are hard links to discover, much less navigate, without direct reference to a domain.
Limited access content	Many websites use a mechanism to limit access to or prevent duplication of their content. » The Robots Exclusion Standard or the Robot Standard serves to discourage searching. » A Completely Automated Public Turing Test to tell Computers and Humans Apart (CAPTCHA) requires users to mimic a random code to assure they aren't robots. » A no-store directive serves to prohibit creating cached copies.
Non-HTML content	The World Wide Web depends on Hypertext Markup Language (HTML) to perpetuate linkages. Non-HTML pages are hard to discover and access, as the accepted links are absent. » Textual content encoded as images, video files, or similar visual file formats, not discoverable via search engines, are also difficult to isolate by other means. » Scripted content involves links produced by JavaScript that are only accessible through links produced by JavaScript or content that is dynamically downloaded from webservers via Flash or Ajax solutions. » Specialized software, such as The Onion Router (TOR) or the Invisible Internet Project (I2P) anonymous peer-to-peer distributed communication network, are required to access content not otherwise discoverable on the WWW or Internet.
Unlinked content	Not all HTML pages contain external links. » Websites typically contain links from other websites, known as *backlinks* or in(bound) links. The number and quality of backlinks are a major factor in determining a website's ranking in search engine results. However, search engines don't usually detect all backlinks. In addition, not including backlinks can prevent search engines and web crawling programs from discovering and indexing website content. » Web archival services can reveal now-defunct webpages across their history. Websites such as the Wayback Machine (archive.org/web) allow viewing of old and often obsolete webpages that are often no longer accessible by today's search engines.

known as the Deep, Hidden, or Invisible Web, tends toward security by obscurity. As such, search and direct access aren't as straightforward as is the case in the WWW. Table 1 describes the wide range of techniques used to control access to Deep Web holdings.

Embedded within the Deep Web is the Dark Web or Dark Net. It's here that bad actors of all stripes—script kiddies out to deface websites; professional hackers who break into corporate and government networks to steal data, wreak havoc, and commit extortion; pedophiles circulating child pornography; drug, arms, and human traffickers; terrorists spreading propaganda, recruiting fighters, and planning attacks; digital media pirates; and cybermercenaries for rogue-state intelligence services—communicate with one another and trade in hacking tools, malware, ransomware, and various illegal goods and services. This underground market is vast enough to contain its own search engines, community forums, and rating systems just like the WWW. While highly organized, the Dark Web lacks an ethos except, perhaps, that of "honor among thieves." Trade is anonymous: unknown buyers and sellers often transact business using bitcoin and other crypto currencies. Ironically, however, the Dark Web is readily accessible via The Onion Router (TOR)—freely downloadable, open source software that assures user anonymity by disguising actual IP addresses (see Figure 1).[3]

The Dark Web is a hub of botnet activity. According to the most recent

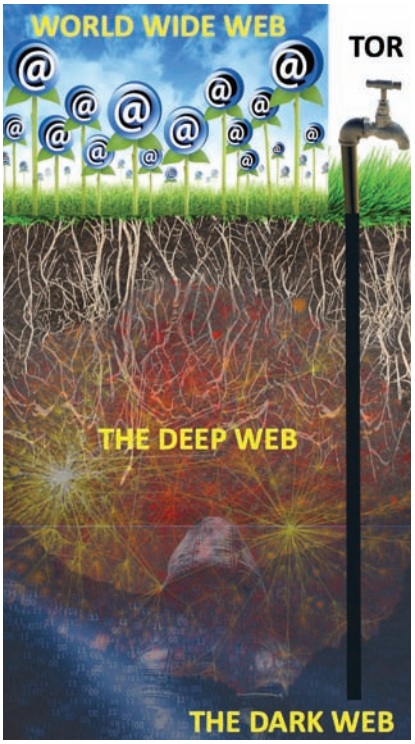

Figure 1. A metaphorical view of the Internet. The Onion Router (TOR) provides easy access to the Dark Web while also protecting users' anonymity.

Imperva Incapsula Bot Traffic Report (www.incapsula.com/blog/bot-traffic-report-2016.html), bots comprised 52 percent of Internet traffic in 2016, with 29 percent having malicious aims such as distributed denial of service (DDoS). Most alarmingly, 94.2 percent of domains surveyed experienced at least one bot attack within a 90-day period. Meanwhile, bots with latent payloads, such as the more than 350,000 automated Twitter accounts making up the "Star Wars" botnet,[4] are also surfacing from the depths of the Dark Web.

While the Dark Web, as part of the Internet, can't be shut down, it can be disrupted. Law enforcement agencies have expended considerable resources to penetrate the Dark Web and expose its anonymous users, occasionally with success.

On 2 October 2013, at a branch of the San Francisco library, federal agents apprehended Ross William Ulbricht, administrator of the online criminal marketplace Silk Road, while he was logged into the site on a laptop through a temporarily encrypted TOR connection over the library's public Wi-Fi system. Since the site's launch in January 2011, more than 100,000 buyers had used it to conduct $1.2 billion worth of bitcoin transactions for illegal drugs and other unlawful goods and services ranging from pirated media to forged documents to murder for hire.[5] Ulbricht later received a life sentence and was ordered to pay more than $180 million in restitution for an array of felonies.[6]

There have been other high-profile incidents. In November 2014, EU and US police agencies seized hundreds of TOR ".onion" domains used by multiple illegal drug markets, including Silk Road 2.0—which emerged on the Dark Web shortly after Ulbricht's arrest[7]—and other black-market operations and arrested 17 people in Operation Onymous.[8] In February 2015, the FBI seized a server for a TOR-hidden child porn site called Playpen and deployed malware to reveal the true IP addresses of the site's 215,000 users, resulting in more than 50 prosecutions to date.[9] In another recent case, an international task force took down the Avalanche syndicate, a global trafficker in Botnet malware, in a series of raids on 1 December 2016 after 4 years of painstaking surveillance.[10] Agents from 30 countries arrested 5 kingpins; confiscated 37 servers; took more than 200 others offline; and seized, blocked, or disrupted more than 800,000 domains. Avalanche had used more than 20 families of malware to infect victims in 180 countries, causing hundreds of millions of dollars in damage.

Red hat hackers have also scored hits. In February 2017, a member of the vigilante group Anonymous took 10,000 child porn sites offline—an estimated 20 percent of the Dark Web—and leaked users' account information.[11]

A CYBERSECURITY CRISIS

The Ulbricht, Avalanche, and other cases that have garnered international attention in recent years highlight the Dark Net's role as a breeding ground of cybercrime, which is becoming ever more destructive and costly. A 2014 report by the Center for Strategic and International Studies estimated the annual losses to the global economy from cybercrime to be $445 billion,[12] but a 2015 study by Jupiter Research projected that, with the rapidly increasing digitization of consumer and business data, those costs will more than quadruple by 2019.[13]

In response to this growing threat, both governments and companies are investing considerably more money in cybersecurity products, services, and research. Worth $3.5 billion in 2004, the global cybersecurity market is expected to reach $120 billion this year,[14] with cumulative spending to eclipse $1 trillion over the next five years.[15]

Traditionally, practical security solutions have focused on protecting endpoint devices such as desktops, laptops, tablets, point-of-sale terminals, bar-code readers, and smartphones from attack when connected to the Internet and other networks. Such solutions include antivirus/antimalware protection, intrusion detection and prevention, patch distribution, a firewall, application and device management, network access control, URL blocking, email server protection, directory integration, automated backup, password management, file-level encryption, vulnerability scanning, and incident reporting. Available tools vary according to an organization's size and resources and can be managed and operated by in-house IT staff or third-party vendors. The global endpoint security market, valued at $11.6 billion in 2015, is estimated to grow to $17.4 billion by 2020.[16]

However, dark clouds are gathering. The sheer number of devices in the emerging Internet of Things—projected to soar from over 17 billion in 2016 to about 30 billion in 2020[17]—suggests that trying to secure every device would be prohibitively expensive and ultimately futile. While

endpoint protection will still have value for some time, it's unlikely to be a long-term solution.

In addition, some researchers are beginning to question the efficacy of the most straightforward endpoint protection: antivirus software. They note that sometimes innately insecure antivirus packages aggressively attach themselves to other software such as browsers and word processors, significantly adding to system overhead.[18] Others claim that endpoint solutions leave data residing within networks, including clouds, at risk, yet organizations continue to spend 10 times as much on endpoint security than on fundamental data encryption.[19] Still others advocate for redesigning operating systems from the ground up to protect networks by default.[20]

Efforts are underway to consider whether the Internet itself, which has evolved incrementally, should be completely re-architected. The National Science Foundation, for example, is exploring such clean-state approaches in its Future Internet Design (FIND) and Global Environment for Networking Innovations (GENI) initiatives.[21] However, such ambitious projects are in their infancy and, if ever realized, won't be implemented anytime soon. There are also proposals to better secure parts of the Internet—such as eHealth networks[22]—through targeted improvements in technology, policy, and digital literacy, but such schemes have limited effectiveness.

PENETRATING THE DARK WEB

The root of many, if not most, cybersecurity threats lies not at the edge of the Internet but deep within it, in the Dark Web. Endpoint protection is thus doomed to fail because it only addresses the symptoms of cybercrime, not the disease itself.

However, the Dark Web is becoming harder to crack as privacy and encryption techniques become more sophisticated. TOR is reportedly adding a layer of privacy later this year that will make discovery of who is hosting and visiting a given site next to impossible. Sites will be far less discoverable and will be accessible by invitation only.[23] In addition, bitcoin, once the cryptocurrency of choice on the Dark Web, is being rapidly replaced by Monero, which offers stealth mechanisms that prevent the indirect tracing of those conducting transactions—a vulnerability that has dogged bitcoin.[24] The same open source tools touted by privacy advocates to protect personal data and to elude government censorship and surveillance are also fueling widespread criminal activity.

Given the Dark Web's sophisticated infrastructure and the superior technical capabilities of many of its denizens, traditional forensic techniques are unlikely to have a substantial or lasting effect. However, emerging machine learning, data mining, and analytics tools are poised to become formidable offensive weapons in the fight against cybercrime.

Emerging machine learning, data mining, and analytics tools are poised to become formidable offensive weapons in the fight against cybercrime.

Because the Internet is a large network of networks with trillions of interconnected nodes, patterns indicative of potentially harmful or illegal activity—for example, botnets, malware distribution, and peer-to-peer file sharing—can be discovered through advanced algorithms and visualization software.[25] These tools can not only help target and disable Dark Net sites but also provide legal evidence against identified offenders. Law enforcement agencies often employ secret and controversial techniques to take down illegal sites and arrest their operators—in some cases, foregoing prosecution to avoid revealing the technology they used[26]—but AI and big data analytics offer an alternative method to discover perpetrators without running afoul of civil liberties advocates.

For this approach to work, however, there must be a reliable baseline of information on cybercriminal behavior. Over the years, many have called for data on DDoS attacks, breaches, and other cybercrimes to be shared openly for the benefit of all. Corporations, however, have been reluctant to admit they've been attacked, or to acknowledge the full extent of damage, for fear of loss of market share or reputation or the imposition of further regulation.

Yet, making such data readily available is essential to establishing cybercrime patterns in different economic sectors. Toward this end, IBM has opened its extensive cybersecurity event dataset to public scrutiny.[27] In addition, various companies, security firms, government agencies, and nonprofits issue regular reports on data breaches. Verizon, for example, publishes an annual *Data Breach Investigations Report* with lessons learned from a dataset that, as of 2016, contained more than 100,000 incidents and 2,200 confirmed data breaches in 82 countries.[28]

What's needed is a well-secured, publicly accessible repository of cybercrime data, perhaps run by a public–private intermediary, to which companies and other entities could contribute without fear of negative consequences. This differs significantly from a security practices rating system, previously proposed in this column, although the two could work synergistically. Promising initiatives in this direction include the VERIS (Vocabulary for Event Recording and Incident Sharing) Community Database (vcdb.org), which collects

and disseminates data from publicly disclosed breaches in an open format from various government agencies, media reports, and press releases, and Gemalto's Breach Level Index (breachlevelindex.com).

Beyond the goal of detecting and gaining information about cybercriminals hiding in the Dark Web, there's the question of what to do about them. It's impractical to root out every perpetrator—investigations are time-consuming and expensive and often require simultaneous raids by multiple law enforcement agencies at locations around the world as well as costly, lengthy legal proceedings. In many cases, the criminal enterprise itself is small but, because of its huge online reach, does far more extensive economic damage than comparable syndicates in the pre-Internet days.

Consequently, we might have no alternative than to fight fire with fire and employ the same tactics used by Dark Web sites against them, including DDoS attacks and malware infiltration. This approach, however, would require carefully crafted guidelines to avoid ensnaring legitimate users or causing other unintended negative consequences. Although most Dark Web content is illicit,[29] not all TOR users are cybercriminals; journalists, political activists, and citizens in countries with extreme censorship also use the software to provide anonymity.

With the Internet's explosive growth, vast, intertwined networks now underlie every facet of life, from commerce to entertainment to communication. Such networks are continually changing, making it difficult to predict their evolution. This reality can be unsettling—even scary—when applied to activity driven by humanity's darker side.

In dealing with cybercrime, the security community must rethink its approach. Endpoint protection is important but will never address the source of the growing cybersecurity crisis. Cybercrime is a cancer, spreading from the Dark Web into the rest of the Internet. Eradicating this disease will require understanding dynamic, nonlinear network mechanics to prevent sick cells from clustering into malignant growths that threaten to stifle the exchange of goods and services worldwide. It ultimately calls for exploiting advances in big data mining and analytics to create automated mechanisms that systematically identify and eliminate cybercriminals.

REFERENCES

1. "Worldwide Retail Ecommerce Sales Will Reach $1.915 Trillion This Year," *eMarketer*, 22 Aug. 2016; www.emarketer.com/Article/Worldwide-Retail-Ecommerce-Sales-Will-Reach-1915-Trillion-This-Year/1014369.
2. M.K. Bergman, "The Deep Web: Surfacing Hidden Value," *J. Electronic Publishing*, vol. 7, no. 1, 2001; quod.lib.umich.edu/j/jep/3336451.0007.104%20?view=text;rgn=main.
3. M. Chertoff and T. Simon, *The Impact of the Dark Web on Internet Governance and Cyber Security*, Centre for Int'l Governance Innovation and Chatham House, Feb. 2015; www.ourinternet.org/sites/default/files/publications/GCIG_Paper_No6.pdf.
4. "Cybersecurity Experts Uncover Dormant Botnet of 350,000 Twitter Accounts," *MIT Technology Rev.*, 20 Jan. 2017; www.technologyreview.com/s/603404/ cybersecurity-experts-uncover-dormant-botnet-of-350000-twitter-accounts.
5. D. Segal, "Eagle Scout. Idealist. Drug Trafficker?," *The New York Times*, 18 Jan. 2014; www.nytimes.com/2014/01/19/business/eagle-scout-idealist-drug-trafficker.html.
6. A. Greenberg, "Silk Road Creator Ross Ulbricht Sentenced to Life in Prison," *Wired*, 29 May 2015; www.wired.com/2015/05/silk-road-creator-ross-ulbricht-sentenced-life-prison.
7. A. Greenberg, "'Silk Road 2.0' Launches, Promising a Resurrected Black Market for the Dark Web," *Forbes*, 6 Nov. 2013; www.forbes.com/sites/andygreenberg/2013/11/06/silk-road-2-0-launches-promising-a-resurrected-black-market-for-the-dark-web/#7280363161c5.
8. A. Greenberg, "Global Web Crackdown Arrests 17, Seizes Hundreds of Dark Net Domains," *Wired*, 7 Nov. 2014; www.wired.com/2014/11/operation-onymous-dark-web-arrests.
9. D. Kravets, "Playpen Moderator Sentenced to 20 Years in Prison," *Ars Technica*, 7 Feb. 2017; arstechnica.com/tech-policy/2017/02/moderator-for-darknet-child-porn-site-playpen-gets-20-years.
10. J. Greuel, "It Took 4 Years to Take Down 'Avalanche,' a Huge Online Crime Ring," *Wired*, 2 Dec. 2016; www.wired.com/2016/12/took-4-years-take-avalanche-huge-online-crime-ring.
11. A. Verma, "20% of the Dark Web Taken Down by Hacker, Here's How He Did It," *Fossbytes*, 6 Feb. 2017; fossbytes.com/20-dark-web-taken-anonymous-hackers.
12. Center for Strategic and Int'l Studies, *Net Losses: Estimating the Global Cost of Cybercrime*; *Economic Impact of Cybercrime II*, June 2014; www.mcafee.com/us/resources/reports/rp-economic-impact-cybercrime2.pdf.
13. J. Moar, *The Future of Cybercrime & Security: Financial & Corporate Threats & Mitigation 2015–2020*, Juniper Research, 5 Dec. 2015; www.juniperresearch.com/researchstore/strategy-competition/cybercrime-security/financial-corporate-threats-mitigation.
14. A. Ross, "Want Job Security? Try Online Security," *Wired*, 25 Apr. 2016; www.wired.co.uk/article/job-security-cybersecurity-alec-ross.
15. S. Morgan, *Cybersecurity Market Report: Q 1 2017*, Cybersecurity Ventures, 17 Feb. 2017; cybersecurityventures.com/cybersecurity-market-report.
16. *Endpoint Security Market by Solution (Anti-Virus, Antispyware/Antimalware, Firewall, Endpoint Device Control, Intrusion Prevention, Endpoint Application Control), Service, Deployment*

Type, Organization Size, Vertical, and Region—Global Forecast to 2020, TC 2287, MarketsandMarkets, Nov. 2015; www.marketsandmarkets.com/Market-Reports/endpoint-security-market-29081235.html?gclid=COCym_LGl9ICFQQvaQodI14Nfg.
17. A. Nordrum, "Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated," *IEEE Spectrum*, 18 Aug. 2016; spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated.
18. S. Anthony, "It Might Be Time to Stop Using Antivirus," *Ars Technica*, 27 Jan. 2017; arstechnica.com/information-technology/2017/01/antivirus-is-bad.
19. P. Kuper, "The State of Security," *IEEE Security & Privacy*, vol. 3, no. 5, 2005, pp. 51–53.
20. P A. Loscocco et al., "The Inevitability of Failure: The Flawed Assumption of Security in Modern Computing Environments," *Proc. 21st Nat'l Information Systems Security Conf.* (NISSC 98), 1998, pp. 303–314.
21. A. Feldmann, "Internet Clean-Slate Design: What and Why?," *ACM SIGCOMM Computer Communication Rev.*, vol. 37, no. 3, 2007, pp. 59–64.
22. *Safety and Security on the Internet: Challenges and Advances in Member States*, World Health Org., 2011; www.who.int/goe/publications/goe_security_web.pdf.
23. A. Greenberg, "It's About to Get Even Easier to Hide on the Dark Web," *Wired*, 20 Jan. 2017; www.wired.com/2017/01/get-even-easier-hide-dark-web.
24. A. Greenberg, "Monero, the Drug Dealer's Cryptocurrency of Choice, Is on Fire," *Wired*, 25 Jan. 2017; www.wired.com/2017/01/monero-drug-dealers-cryptocurrency-choice-fire.
25. E. Nunes et al., "Darknet and Deepnet Mining for Proactive Cybersecurity Threat Intelligence," arXiv:1607.08583, 2016; arxiv.org/pdf/1607.08583.pdf.
26. C. Farivar, "Feds May Let Playpen Child Porn Suspect Go to Keep Concealing Their Source Code," *Ars Technica*, 9 Jan. 2017; arstechnica.com/tech-policy/2017/01/feds-may-let-playpen-child-porn-suspect-go-to-keep-concealing-their-source-code.
27. C. Barlow, "Where Is Cybercrime Really Coming From?," TED Talk, Nov. 2016; www.ted.com/talks/caleb_barlow_where_is_cybercrime_really_coming_from.
28. *2016 Data Breach Investigations Report*, Verizon, 2016; www.verizonenterprise.com/verizon-insights-lab/dbir.
29. C. McGoogan, "Dark Web Browser Tor Is Overwhelmingly Used for Crime, Says Study," *The Telegraph*, 2 Feb. 2016; www.telegraph.co.uk/technology/2016/02/02/dark-web-browser-tor-is-overwhelmingly-used-for-crime-says-study.

GEORGE HURLBURT is chief scientist at STEMCorp, a nonprofit that works to further economic development via adoption of network science and to advance autonomous technologies as useful tools for human use. He is engaged in dynamic, graph-based Internet of Things architecture. Hurlburt is on the editorial board of *IT Professional* and is a member of the board of governors of the Southern Maryland Higher Education Center. Contact him at ghurlburt@change-index.com.