RAPID7

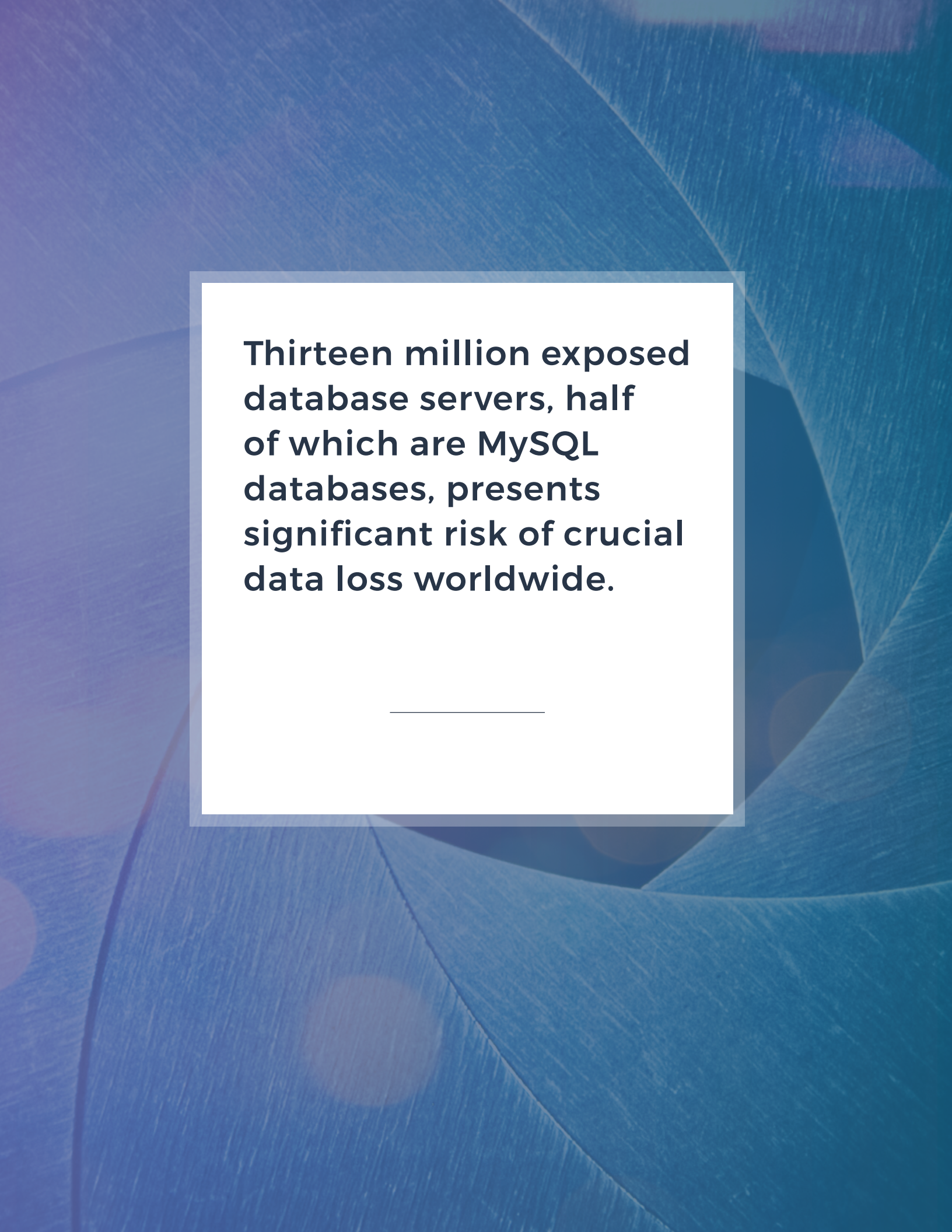
RAPID7 RESEARCH

NATIONAL EXPOSURE INDEX 2018

Inferring Internet Security Posture by Country Through Port Scanning

Rapid7 Labs | June 7, 2018

CONTENTS


Thirteen million exposed database servers, half of which are MySQL databases, presents significant risk of crucial data loss worldwide.

EXECUTIVE SUMMARY

In 2016, Rapid7 Labs launched the National Exposure Index in order to get a measurable, quantitative answer to a fairly fundamental question: What is the nature of internet exposure—services that either do not offer modern cryptographic protection, or are otherwise unsuitable to offer on the increasingly hostile internet—and where, physically, are these exposed services located?

Now in our third year, we continue this ongoing investigation into the risk of passive eavesdropping and active attack on the internet, and offer insight into the continuing changes involving these exposed services. We've also added a third dimension for exposure, "amplification potential," in the wake of the disastrous memcached[1] exposure uncovered in 2018.

Finally, we've modified our ranking algorithm in this edition. First, we're measuring and scoring amplification abuse potential. Second, we've added more studies targeting exposed databases, and weighted groups of protocols as "more risky" than others, such as SMB, memcached, and database ports. In addition, we're treating the especially responsive 2% of IPv4 nodes (0.08% of routable IPv4 addresses) as mere noise absorbers/generators in their networks and have removed those nodes from scoring entirely.

Key Findings

- The United States leads all other countries in the 2018 exposure rankings, scoring the highest in nearly every exposure metric we measure. Following the U.S. is China, Canada, South Korea, and the United Kingdom, which together control over 61 million servers listening on at least one of the surveyed ports.
- There are 13 million exposed endpoints associated with direct database access, half of which are associated with MySQL. Along with millions of exposed PostgreSQL, Oracle DB, Microsoft SQL Server, Redis, DB2, and MongoDB endpoints, this exposure presents significant risk of crucial data loss in a coordinated attack.
- While the number of exposed Microsoft SMB Servers dropped considerably after the WannaCry attack of 2017, there remain about a half a million targets today, primarily in the U.S., Taiwan, Japan, Russia, and Germany.
- Amplification-based distributed denial of service (DDoS-A) remains a powerful technique for harming enterprises and providing cover for more sophisticated attacks. While the number of exposed UDP-based memcached servers is less than 4,000, there are about 40,000 unpatched, out-of-date memcached servers, which are at risk of being drafted into the next record-breaking DDoS attack.

These key findings tell us that the most risk to the internet originates in countries that have significant investment in, and reliance on, a safe and stable internet. This indicates to us that national internet service providers in these countries can use these findings to understand the risks of internet exposure, and that they, along with policymakers and other technical leaders, are in an excellent position to make significant progress in securing the global internet.

[1] https://blog.cloudflare.com/memcrashed-major-amplification-attacks-from-port-11211/

The year 2018 has already seen the largest distributed denial of service (DDoS) attack on record, using unsecured ‘memcached’ UDP servers.

MEASURING NATIONAL EXPOSURE

It's important to note that it's not just mature, traditionally "rich" or "large" countries that rely on a healthy and functioning internet. As of the start of 2018, more than half of all humans now maintain an active internet presence[2], after significant growth in both client-side and server-side infrastructure in Asia and Africa. We are in a crucial period of human history, and we need to actively measure the patterns of internet usage that impact the security and stability of this incredible, planet-wide resource. By comparing regions both globally and with their immediate neighbors, we believe it's possible to deliberately apply some "network husbandry" to the internet to ensure it remains supportive of technical innovation, cultural value, and economic prosperity.

For 2018's National Exposure Index, we once again took on the task of surveying the nature of the internet in order to determine (a) what is actually running on today's internet, versus what we believe should be present there, and (b) which geopolitical regions are most at risk for deliberate, wide-scale attacks on core internet services. Regional and global outages are still occuring with some frequency. In our first National Exposure Index in mid-2016, we warned of an impending disaster involving the millions of unsecured telnet servers, which turned out to be ripe hunting ground for the world's largest botnets, Mirai and its variants. In 2017, we were planning on shifting focus to Windows SMB, but WannaCry and its EternalBlue-powered variants beat our publish date to the punch.

The year 2018 has already seen the largest distributed denial of service (DDoS) attack on record, using unsecured 'memcached' UDP servers. Due to this event, we're paying much closer attention to memcached and other connectionless UDP services that can be abused in amplification attacks, and we have added this metric to the national exposure ranking system.

We also continue to worry about the exposure level of popular database servers, such as MySQL, PostgreSQL, Microsoft SQL Server, Oracle DB, and IBM DB2, as well as the "NoSQL" databases like MongoDB and Redis. It's our hope that by highlighting the prevalence of these services, and the specific geographic regions in which they reside, we can get ahead of a coming DB disaster.

[2] https://www.internetworldstats.com/stats.htm

Putting all this together, we believe that by measuring the most commonly deployed services on the internet and then breaking these statistics out by country, we can produce a ranked list of "most exposed" countries. Armed with this information, we have the opportunity to identify which nations can improve their local infrastructure's "natural" exposure to hostile actors. National borders are quite weak on the internet, as everyone is usually only a couple hundred milliseconds "away" from everyone else. Recent events suggest that nation-state actors are keenly interested in taking advantage of national internet exposure to pursue their own interests, so defenders can use the information presented in this paper to make informed decisions about how to best manage their own geopolitical region of the internet.

Changes to Methodology

We've again updated our regional ranking system algorithm in order to more accurately measure and report on the nature of the internet, both globally and regionally. First, we've added a small handful of new services to survey—namely, the aforementioned database services—and have altered the way we score those networks that are suspiciously responsive when probed on non-existent services, dubbed "canary ports." [3]

We're also able to delve a little deeper in both SMB (which is important for EternalBlue-powered malware families, such as WannaCry and its derivatives) and SSH (which we suspect will be important for future iterations of Mirai-like bots). While this second-level protocol analysis does not directly impact exposure ranking, it does offer deeper insight into the current deployment of these technologies.

The inclusion of more protocol-based scans, the identification of nodes responsive to canary port probes, and the statistical filtering of errant and anomalous endpoints have resulted in a higher efficacy data set with more representative and accurate results. These enhancements have made it difficult—if not impossible—to provide a reasonable or sensible year-over-year comparison with prior datasets. That said, we're confident that future studies using these techniques will have significant longitudinal value.

---

[3] For a much more in-depth discussion of these canary ports, see the section of the same name on page 19, below.

MEASURING INTERNET ADOPTION

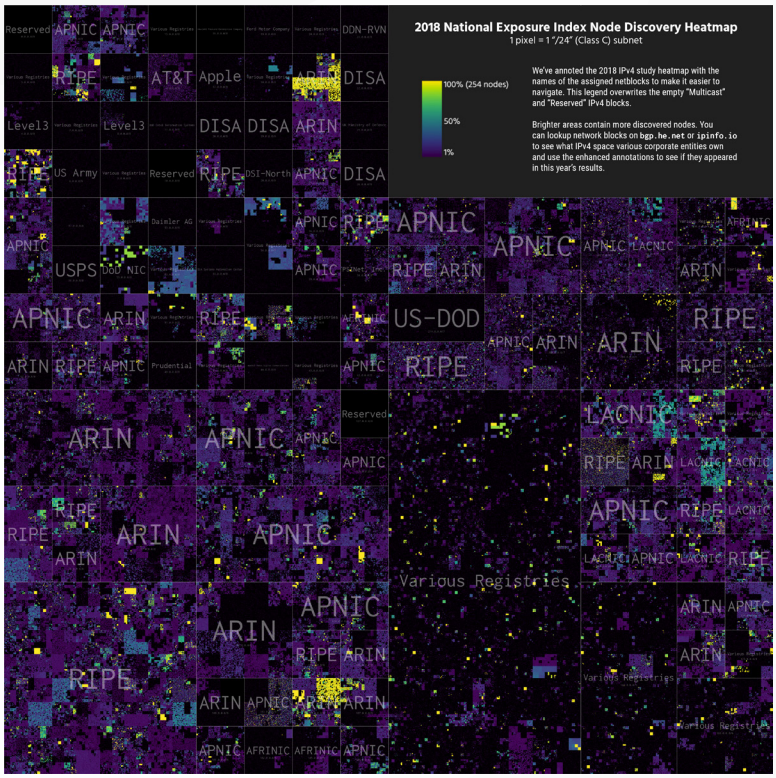
In order to be reachable on the internet, any service (such as a website, a mail server, or a database) must run on a **server**, which is reachable by a unique **IP address** and a standardized, well-known **port** associated with that address[4]. A client computer, such as a desktop or a smartphone, then makes a TCP/IP connection to that service, and the magic of packet exchange occurs. UDP services function in much the same way, but there is no initial "handshake" to establish the connection; for UDP services, communication begins without any guarantee the server is actually listening, and responses are sent based only on the apparent source of the initial request.

Given this standard model of client/server communication, we can measure the overall internet population of services offered by launching broad, shallow port scans across all of IPv4 address space, testing for responses from 38 selected ports that are most commonly found running TCP/IP services, and geolocating each server found by country. We regularly perform these actions through Project Sonar.[5]

Now, this is a very broad generalization of TCP/IP networking, and we will be the first to admit it does not capture the absolute universe of "the internet." After all, we are not counting the ongoing deployment of IPv6, we cannot count the population of client computers (including smartphones) through port scanning, and we are not able to reach through NATs and firewalls. For more details on these factors that necessarily limit Project Sonar's telemetry capabilities, please see Appendix B.

Keeping in mind these caveats, Figure 6[6], is just about the most accurate map you will find of "the internet."

Figure 1: Heat map of the internet



As you can see, 0.0.0.0 is in the upper left corner, 255.255.255.255 is in the upper right, 80.0.0.0/4 is in the bottom left and 168.0.0.0/6 is in the bottom right. Each pixel represents one block of 255 addresses. The black areas are addresses that are either unresponsive, unroutable (private), or otherwise unreachable by our Sonar scans. The colored areas have a higher density of responsive ports, while darker (but not black) areas are lower density regions of the internet.

The addition of selected UDP scanning, in particular, helped to fill in some gaps when compared to previous maps in past iterations of the National Exposure Index. We've also laid in some helpful gridlines to indicate which authorities control which netblocks.

[4] Complete service notations are expressed as a "tuple" of an IP address and port, such as 127.0.0.1:445 (where "445" is the port and "127.0.0.1" is the IP address).

[5] https://sonar.labs.rapid7.com/

[6] A high resolution version of this map, along with all the data behind its generation, is available at https://github.com/rapid7/data/blob/master/national-exposure/2018

Today's Internet

While visually interesting for the more mathy reader, sorting simply by IP address does not give us the view we're after; ultimately, it's more useful for this study to see the utilization of IP address space by country. For that, we've generated a view of the internet, colored by size and arranged by region in (Figure 2)[7]. Each square represents one country, and by giving them a uniform size the physical land mass each country occupies does not trick the eye into believing there are more addresses present than there really are (a common problem with world choropleth maps). Each country is placed in relative proximity to their bordering counterparts, and you can quickly spot the countries with the most IPv4 resources.

There are some "surprises" when it comes to the National Exposure Index view of IPv4 country IP-space utilization. While the overall order is generally expected—with the United States, China, Germany, South Korea, and Japan home to the lion's share of discovered IP addresses—countries such as Mexico, Iran, Ireland, Saudi Arabia, and the United Arab Emirates have more of their IPv4 allocations used by services reached by the National Exposure scans (Figure 3). Even if a country hits 100% utilization of IPv4 space, all is not lost since IPv6 address space—with practically infinite capacity—usage is increasing[8] and available at-will for expansion.

Figure 2: Country-level IPv4 totals

Internet services by region and country

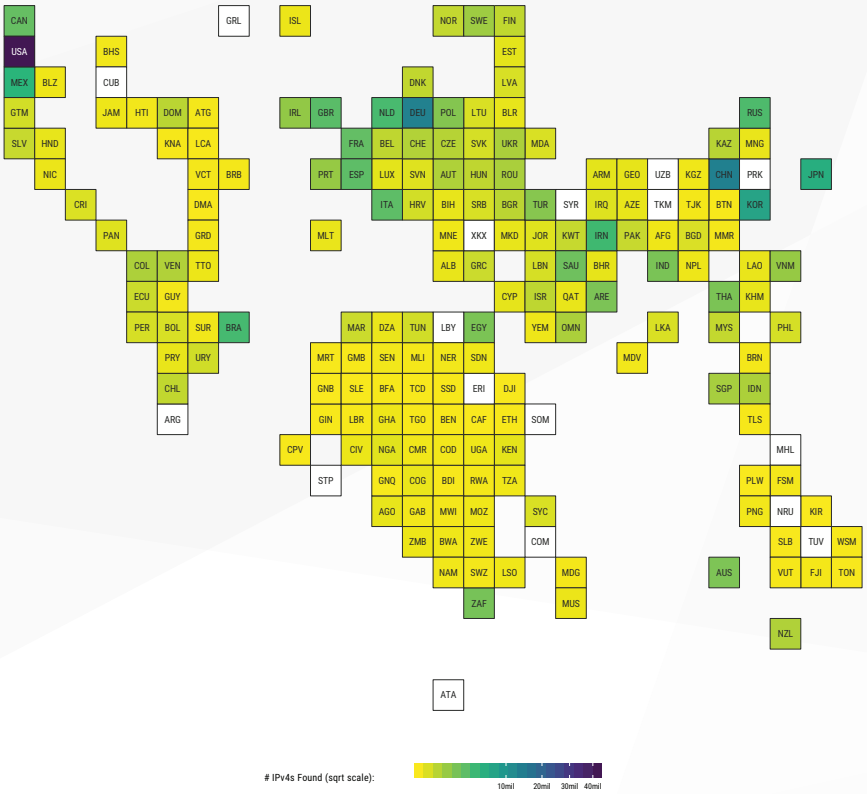
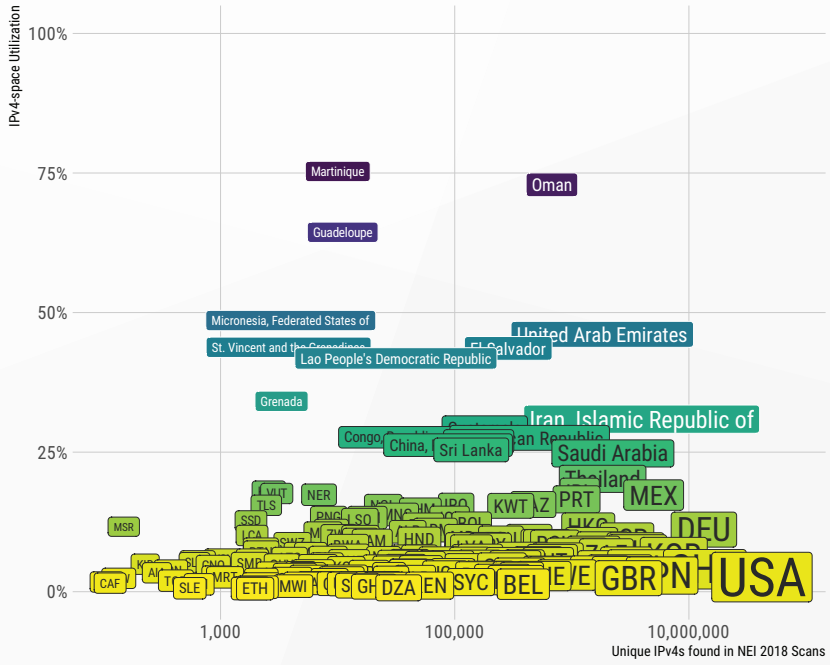


Figure 3: Identified exposed addresses and corresponding IPv4 assigned allocation utilization

Countries like the United States and China have enormous IPv4 allocations while others do not. A handful of those with smaller allocations are exposing services on over 50% of their allocated space..



Country label size proportionally (sqrt) scaled to total IPv4 allocations

[7] Note that this color scale moves from light-to-dark to indicate less-to-more volume, but for graphs on a dark background (like the internet heat map, above), the color scale moves from dark-to-light instead.

[8] https://www.google.com/intl/en/ipv6/statistics.html#tab=per-country-ipv6-adoption&tab=per-country-ipv6-adoption

MEASURING EXPOSURE

What do we mean when we say "exposure"? For our purposes, we would consider a system to be "exposed" if it's (a) offering a natively unencrypted service on the public internet, (b) offering a service on the internet that is unsuitable for public access, or (c) subject to amplification abuse through connectionless communication. If any of these conditions are met for a given IP-addressable server, it counts against that IP address's geolocated country's exposure. While exposure is a useful shorthand for security professionals, we should take a moment to unpack all three of these conditions.

Cleartext Services

The internet was originally designed to allow for any computer to communicate with any other computer—this is a core feature of TCP/IP networking. This was revolutionary in the computing environment of the late 20th Century, which was dominated by terminals physically wired to mainframes, as it essentially democratizes and decentralizes data, storage capability, and computing power. Anyone with a computer on the internetwork could connect to any server and interact with it. However, this decentralization also means that anyone with a view into the underlying network—the hubs, routers, and switches that actually handle the packets flowing between endpoints—could eavesdrop, impersonate, and alter any communications in transit, both actively and passively.

Modern, certificate-based encryption can prevent these man-in-the-middle shenanigans[9]. Even if an adversary controls one of the routers between you and yourbank.com, you have assurances built in to your web browser that https://yourbank.com is both **authenticated** as truly yourbank.com (and not an imposter), and that your transactions between you and yourbank.com are **confidential**.

Without encryption, no service on the internet can reasonably guarantee that computers at either end of a connection are who they say they are, nor can they guarantee that the data passed between them is both authentic and private. Unencrypted data is commonly referred to as **cleartext**.

Today, we know that some national security organizations in some countries have the capability to conduct large scale, passive monitoring of internet activity, and that the Internet Engineering Task Force proposed in 2014 that "Pervasive Monitoring Is An Attack" in RFC 7258, an official memorandum with that title[10].

While we acknowledge there is tension between the need for strong security controls and the need for reasonable and lawful surveillance capabilities for national security, we contend that **cleartext services are necessarily insecure** from eavesdropping, data alteration, or data breach[11]. After all, an adversary need not have the formidable capabilities of a three-letter agency to snoop on cleartext communications; they need only to compromise one hop, or network segment, between the target (or target population) and the intended service. This is well within the capability of even amatuer cyber criminals camped out on local WiFi access points.

[9] Encryption is hard and there are ways to subvert it, but these technical nuances are beyond the scope of this paper.

[10] https://tools.ietf.org/html/rfc7258

[11] For more on the virtues of encryption see the National Exposure Index of 2016 at https://information.rapid7.com/national-exposure-index.html

Scanned Cleartext and Encrypted Services

For purposes of this investigation, the cleartext services listed below were chosen as scan targets. A complete list of all scan targets, sorted by port number, can be found in Appendix A.

Table 1: Scanned cleartext services

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
80	62,656,633	HTTP	HyperText Transfer Protocol, used to serve web pages and web applications
25	15,664,213	SMTP	Simple Mail Transfer Protocol, used to send email
21	13,359,961	FTP	File Transfer Protocol, used to send and receive data and text files; FTPS, SSH, and HTTPS are all encrypted alternatives
8080	9,243,677	http-alt0	A common alternative port for HTTP, usually used for websites and web proxy services
53	8,832,463	DNS (TCP)	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
110	7,114,795	POP3	Post Office Protocol version 3, used to receive email
143	6,668,963	IMAP	Internet Message Access Protocol, used to receive email
3306	6,087,830	MySQL	MySQL, used to communicate with the (usually) open source MySQL Server published by Oracle
23	5,814,024	telnet	Telnet, a remote command shell interface, one of the oldest protocols on the internet; SSH is an encrypted alternative
8081	5,452,401	http-alt1	A common alternative port for HTTP, usually used for websites and web proxy services
587	5,262,246	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
3389	4,934,495	RDP	Remote Desktop Protocol, a graphical user interface to remotely administer (usually) Microsoft Windows servers and desktops
445	3,507,183	SMB	Server Message Block, a file transfer and remote administration protocol for (usually) Microsoft operating systems
111	3,375,227	rpcbind	Remote Procedure Call port mapping service, usually used on Unix-like operating systems, usually for NFS file sharing
81	2,464,657	http-alt	A common alternative port for HTTP, usually used for websites and web proxy services
135	2,437,524	MS-RPC	Microsoft Remote Procedure Call, usually used on Microsoft OSes for distributed computing
5000	2,300,261	uPNP	Universal Plug-and-Play, a protocol for machine-to-machine discovery and configuration
139	1,934,357	NBSS	NetBIOS Session Service, used in NetBIOS over TCP/IP, usually on Microsoft OSes for file and print sharing
8888	1,855,002	http-alt8	A common alternative port for HTTP, usually used for websites and web proxy services

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
5432	1,594,877	PostgreSQL	PostgreSQL listening service, used to communicate with the T-SQL server of the same name
9100	1,536,469	jetdirect	HP JetDirect, a printer control service used to manage print jobs
1521	1,489,749	oracle	Oracle Database listening service, used to communicate with the T-SQL server of the same name
1433	1,369,495	MSSQL	Microsoft SQL Server service, used to communicate with Microsoft database servers of the same name
6379	1,258,944	Redis	RESP, the Redis Serialization Protocol, used to communicate with Redis, a popular open source database and caching service
5900	1,142,393	RFB	Remote Frame Buffer, a remote GUI for desktop administration, usually used by VNC (Virtual Network Computing)
50000	1,066,201	DB2	IBM DB2 service, used to communicate with DB2 database servers
389	862,686	LDAP	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
27017	561,471	Mongo	Mongo Wire Protocol, used to communicate with MongoDB, a popular open source document database
11211	39,799	Memcached	Memcached, a distributed memory object caching system

While many of these services do offer **opportunistic encryption**[12], such protocols are still susceptible to active attacks where an adversary can rewrite requests and responses to subvert the initial negotiated encryption request. Opportunistic encryption over cleartext protocols is a useful defense against pervasive, passive monitoring, but it is not designed to be sufficient against active attacks.

For comparison to encrypted counterparts, these ports are associated with fully encrypted protocols. Barring implementation errors and software vulnerabilities, these services provide reasonable encryption by default[13].

[12] Services utilizing opportunistic encryption attempt to establish encrypted connections for transmitting data, but resort to cleartext communications if an encrypted connection cannot be established. SMTP's STARTTLS implementation is one example of opportunistic encryption. See the STARTTLS RFC, at https://tools.ietf.org/html/rfc3207.

[13] PPTP relies on older encryption algorithms to guarantee confidentiality and authentication, and has been shown to be quite crackable in practice. While it is "encrypted," it is no longer considered

Table 2: Scanned encrypted services

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
443	44,849,191	HTTPS	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	19,061,180	SSH	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
1723	5,334,237	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers.
993	5,155,630	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	4,997,893	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
8443	4,515,905	https-alt	A common alternative port for HTTPS, usually used for test web sites
465	4,416,327	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
990	1,046,579	FTPS	Secure FTP, an encrypted-by-default alternative to FTP

Inappropriate Services

In addition to the cleartext problems introduced in the early design of the internet, we now see that when you have a network where literally anyone on the planet can establish a connection to anyone else, some less-than-neighborly behavior emerges. In addition to the eavesdropping on and altering of cleartext data as described above, it is common to see more targeted attacks against specific services that have no reason to be accessible to absolutely anyone. Even if a service is otherwise inherently secure against tampering through encryption, unrestricted open access to that service creates an exposure simply by being available.

For example, we surveyed the internet for a number of database services, as noted in Table 1: Microsoft SQL Server, MySQL, MongoDB, PostgreSQL, DB2, Oracle DB, and Redis. Many of these database systems offer perfectly adequate authentication protocols and encryption guarantees (notably Microsoft SQL Server, MySQL, and PostgreSQL), but these services also offer direct access to random strangers when, in practice, there is no earthly reason to do so. There is no case when a database administrator (DBA) would recommend that anonymous users should be able to run any custom search query, if only for the performance disasters that poorly constructed statements provided by amateur DBAs would cause. Databases should always be mediated by a simplified, restricted front end, like a web application[14].

In addition to their sensitivity to denial-of-service conditions, accidental or intentional, services that are inappropriate to deploy on the public internet tend to be among the most complex software applications ever invented. One example is Server Message Block, or SMB. SMB is an all-in-one file sharing and remote administration protocol, usually associated with Windows, that has been an attractive target for attackers and researchers alike for decades, from MS03-049[15] in 2003, to MS08-067[16] in 2008 and "EternalBlue" in MS17-010[17] in 2017, precisely due to the likelihood of vulnerabilities in its complex implementation. Exposing an SMB service on the internet is simply asking for trouble.

[14] Web applications that fail to restrict SQL queries, of course, have SQL injection (SQLi) vulnerabilities. If an entire web app vulnerability class is dedicated to the unplanned reach into a database server, then surely a population of open database servers should be even more troubling.

[15] https://www.rapid7.com/db/modules/exploit/windows/smb/ms03_049_netapi

[16] https://www.rapid7.com/db/modules/exploit/windows/smb/ms08_067_netapi

[17] https://www.rapid7.com/db/modules/exploit/windows/smb/ms17_010_eternalblue

Finally, there are some TCP services that are inappropriate for today's opportunistically hostile internet. These services, such as telnet and FTP, nearly always require exposing credentials and sensitive data to an eavesdropper. Others, such as Universal Plug-and-Play, rcpbind and MS-RPC, and HP JetDirect[18], offer far more control over networked resources than is appropriate for the casual internet stranger.

With these considerations in mind, IPv4 servers that expose these inappropriately open services are **necessarily exposed to increased risk** more than machines that offer only appropriately curated, internet-ready services. This increase in attack surface is irrespective of any controls that would limit access, access attempts, or allowed usage of these services.

Scanned Inappropriate and Appropriate Services

The inappropriate services listed below were chosen as scan targets. A complete list of all scan targets, sorted by port number, can be found in Appendix A.

Table 3: Scanned inappropriate services

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
21	13,359,961	FTP	File Transfer Protocol, used to send and receive data and text files; FTPS, SSH, and HTTPS are all encrypted alternatives
3306	6,087,830	MySQL	MySQL, used to communicate with the (usually) open source MySQL Server published by Oracle
23	5,814,024	telnet	Telnet, a remote command shell interface, one of the oldest protocols on the internet; SSH is an encrypted alternative
1723	5,334,237	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers.
3389	4,934,495	RDP	Remote Desktop Protocol, a graphical user interface to remotely administer (usually) Microsoft Windows servers and desktops
445	3,507,183	SMB	Server Message Block, a file transfer and remote administration protocol for (usually) Microsoft operating systems
111	3,375,227	rpcbind	Remote Procedure Call port mapping service, usually used on Unix-like operating systems, usually for NFS file sharing
135	2,437,524	MS-RPC	Microsoft Remote Procedure Call, usually used on Microsoft OSes for distributed computing
5000	2,300,261	uPNP	Universal Plug-and-Play, a protocol for machine-to-machine discovery and configuration
139	1,934,357	NBSS	NetBIOS Session Service, used in NetBIOS over TCP/IP, usually on Microsoft OSes for file and print sharing
5432	1,594,877	PostgreSQL	PostgreSQL listening service, used to communicate with the T-SQL server of the same name
9100	1,536,469	jetdirect	HP JetDirect, a printer control service used to manage print jobs

[18] https://www.nytimes.com/2016/03/29/nyregion/hacker-weev-says-he-printed-anti-semitic-and-racist-fliers-at-colleges-across-us.html

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
1521	1,489,749	oracle	Oracle Database listening service, used to communicate with the T-SQL server of the same name
1433	1,369,495	MSSQL	Microsoft SQL Server service, used to communicate with Microsoft database servers of the same name
6379	1,258,944	Redis	RESP, the Redis Serialization Protocol, used to communicate with Redis, a popular open source database and caching service
5900	1,142,393	RFB	Remote Frame Buffer, a remote GUI for desktop administration, usually used by VNC (Virtual Network Computing)
50000	1,066,201	DB2	IBM DB2 service, used to communicate with DB2 database servers
389	862,686	LDAP	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
27017	561,471	Mongo	Mongo Wire Protocol, used to communicate with MongoDB, a popular open source document database
11211	39,799	Memcached	Memcached, a distributed memory object caching system

For comparison, the below lists the services that are generally considered appropriate for public internet use.

Table 4: Scanned internet-appropriate services

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
80	62,656,633	HTTP	HyperText Transfer Protocol, used to serve web pages and web applications
443	44,849,191	HTTPS	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	19,061,180	SSH	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
25	15,664,213	SMTP	Simple Mail Transfer Protocol, used to send email
8080	9,243,677	http-alt0	A common alternative port for HTTP, usually used for web sites and web proxy services
53	8,832,463	DNS	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
110	7,114,795	POP3	Post Office Protocol version 3, used to receive email
143	6,668,963	IMAP	Internet Message Access Protocol, used to receive email
8081	5,452,401	http-alt1	A common alternative port for HTTP, usually used for web sites and web proxy services
587	5,262,246	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
993	5,155,630	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	4,997,893	POP3S	Secure POP3, an encrypted-by-default alternative to POP3

PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
8443	4,515,905	https-alt	A common alternative port for HTTPS, usually used for test web sites
465	4,416,327	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
81	2,464,657	http-alt	A common alternative port for HTTP, usually used for web sites and web proxy services
8888	1,855,002	http-alt8	A common alternative port for HTTP, usually used for web sites and web proxy services
990	1,046,579	FTPS	Secure FTP, an encrypted-by-default alternative to FTP

The astute reader will notice that a little more than half of these protocols, while designed to be exposed on the public internet, are also natively unencrypted. We will explore this apparent dichotomy in the "Characterizing TCP Protocols" section, but briefly: while all inappropriate services are themselves not natively encrypted by default, the reverse isn't true on today's internet.

That said, a truly ideal internet would see new protocols that always guarantee authenticity and integrity of data, and those existing protocols that already are both appropriate for the internet and natively encrypted, as described by the table on the following page.

Table 5: Scanned "ideal" services which are both encrypted and appropriate

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
443	44,849,191	HTTPS	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	19,061,180	SSH	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
1723	5,334,237	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers.
993	5,155,630	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	4,997,893	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
465	4,416,327	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
8443	4,515,905	https-alt	A common alternative port for HTTPS, usually used for test web sites
990	1,046,579	FTPS	Secure FTP, an encrypted-by-default alternative to FTP
8081	5,452,401	http-alt1	A common alternative port for HTTP, usually used for web sites and web proxy services
587	5,262,246	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
993	5,155,630	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	4,997,893	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
8443	4,515,905	https-alt	A common alternative port for HTTPS, usually used for test web sites
465	4,416,327	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
81	2,464,657	http-alt	A common alternative port for HTTP, usually used for web sites and web proxy services

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
8888	1,855,002	http-alt8	A common alternative port for HTTP, usually used for web sites and web proxy services
990	1,046,579	FTPS	Secure FTP, an encrypted-by-default alternative to FTP

UDP Services

New to the 2018 edition of the National Exposure Index, we have included a number of surveys of UDP-based services. Because UDP is a "connectionless" protocol, there is no initial handshake that we test for responsiveness. Instead, each of these services requires a well-formed request in the data portion of the packet, which will elicit a data packet from the target with a response. As a result, the scans tend to run a little slower, but we also have much greater certainty that the nodes being tested are, in fact, hosting the requested service (or are at least medium-interaction honeypots). A selection of these services are discussed further on page 33, in the "Characterizing UDP Protocols" section of this paper.

Scanned Inappropriate UDP Services

The table below lists the UDP protocols that should not be available on the internet, either because they are subject to amplification attacks or offer unreasonable control over the target network's connected devices.

Table 6: Scanned inappropriate UDP services

UDP PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
5060	14,001,928	SIP	Session Initiation Protocol, usually used in Voice over IP applications
1900	1,289,184	SSDP	Simple Service Discovery Protocol, used with UPnP
389	810,656	LDAP	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
137	737,185	NBSN	NetBIOS Name Service, used in NetBIOS over TCP/IP, usually on Microsoft OSes for file and print sharing
5353	463,924	mDNS	Multicast DNS, useful in networks without dedicated name services
11211	3,777	Memcached	Memcached, a distributed caching service
19	3,756	Chargen	Chargen, a service that echos a list of characters
990	1,046,579	FTPS	Secure FTP, an encrypted-by-default alternative to FTP
8081	5,452,401	http-alt1	A common alternative port for HTTP, usually used for web sites and web proxy services
587	5,262,246	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
993	5,155,630	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	4,997,893	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
8443	4,515,905	https-alt	A common alternative port for HTTPS, usually used for test web sites
465	4,416,327	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP

UDP PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
81	2,464,657	http-alt	A common alternative port for HTTP, usually used for web sites and web proxy services
8888	1,855,002	http-alt8	A common alternative port for HTTP, usually used for web sites and web proxy services
990	1,046,579	FTPS	Secure FTP, an encrypted-by-default alternative to FTP

Two other UDP protocols were also scanned, DNS and NTP:

Table 7: Scanned appropriate UDP services

UDP PORT	OBSERVED COUNT	PROTOCOL / SERVICE	DESCRIPTION
53	7,352,839	DNS	Domain Name Service, used to resolve human-memorable names to IP addresses
123	2,738,152	NTP	NTP, the Network Time Protocol

As with TCP services, the savvy network engineer will know that DNS (UDP port 53) and NTP (UDP port 123) are currently considered reasonable and appropriate for the internet with sufficient configuration hardening, even though they appear to fail tests of reasonable encryption and the amplification potential.

Canary Ports

Along with the 37 TCP services and 9 UDP services chosen for scanning, we also scanned for two "canary" TCP ports: port 5 and port 61439. These are TCP ports that are unlikely to ever respond to any port scanning with an affirmative response[19], since there are no well-known services associated with them. Yet, we picked up responses from approximately 2.39 million devices that respond on both of these ports, and these "ghost" servers account for approximately 0.08% of routable IPv4 address space and 2% of all IPv4 addresses discovered through National Exposure scans. These suspiciously responsive IP addressable servers imply that "something funny" is going on with local firewall rules on that subnet, which is causing those machines to behave as if **any** service asked for is listening. Upon further investigation, it became clear that these IPv4 nodes are often associated with content distribution networks (CDNs), and occasionally associated with unusually behaving edge routers which may be honeypots or misconfigured devices. For context, the following heatmap shows the distribution of these nodes responsive to canary port scans. Clearly, they're not uniform across the total IPv4 space and tend to cluster around netblocks associated with pretty high density of normally responsive servers.

[19] In technical networking parlance, we would not expect "SYN/ACK" responses to any "SYN" packet sent to these canary ports, since no normal services are associated with them.

The heat map[20] in Figure 4 illustrates where, in IPv4 space, the most responsive canary ports reside.

Another way to look at responsive canary port distribution is the circlepack graph by country, in FIgure 5.

Canary Ports and Scoring Exposure

We've determined that the most fair way to score the protocols that appear to be offered by these "ghost nodes" is to treat those host nodes as if they offer no services at all. That said, there is **something** there to respond to our probes. So, while the individual services that appear to be offered (but aren't) are not counted in the exposure scoring algorithm, the presence of these nodes does count against the total IPv4 server utilization for that country. After all, no purely client node or truly absent endpoint would generate an affirmative response. This (admittedly, somewhat paradoxical) treatment of canary ports and ghost nodes seems to give us the best chance at accurately measuring a given network's overall exposure to attack.

[20] As with the overall heatmap of the internet, a high-res version of this map is available at our Github data repository.

Figure 4: Heatmap of canary ports

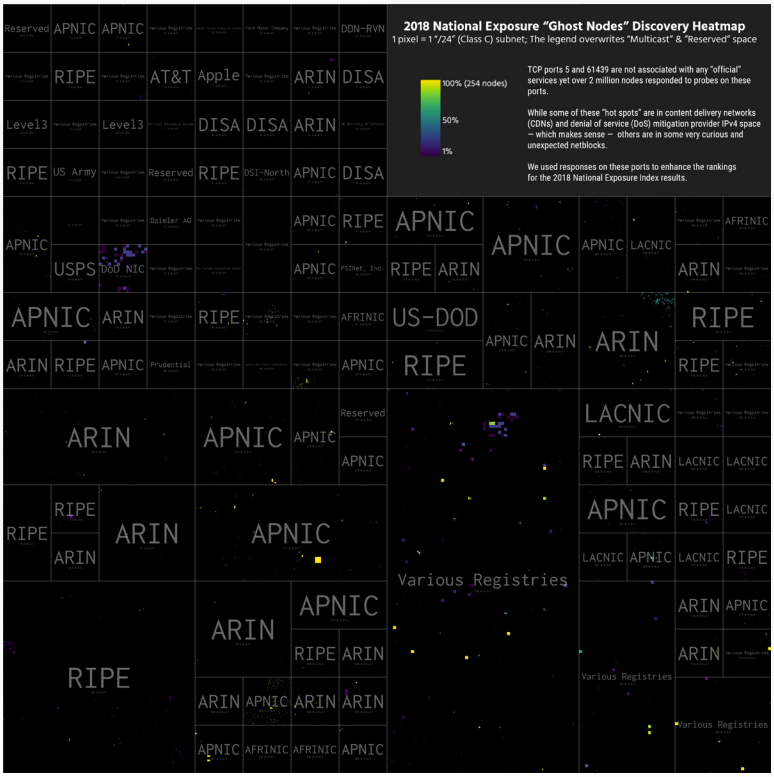
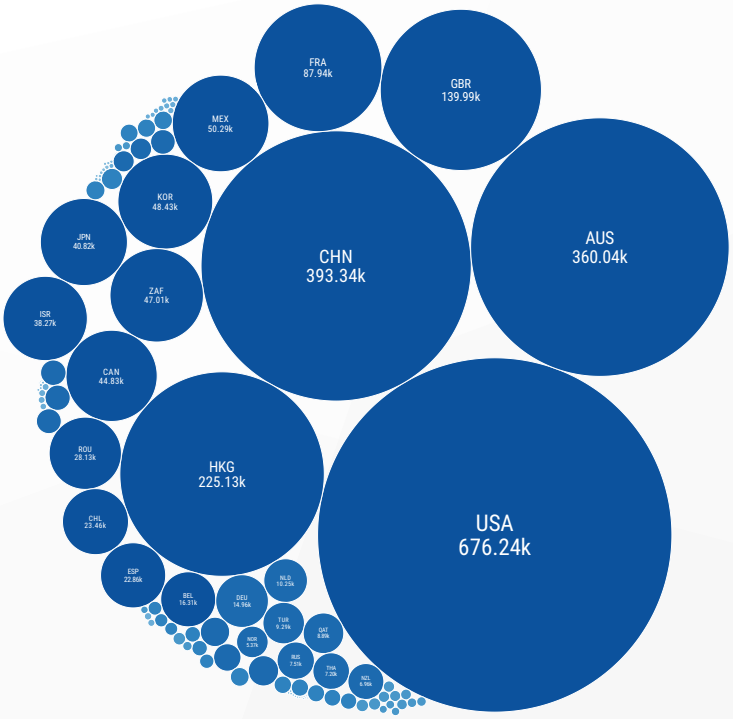


Figure 5: Ghost nodes country distribution

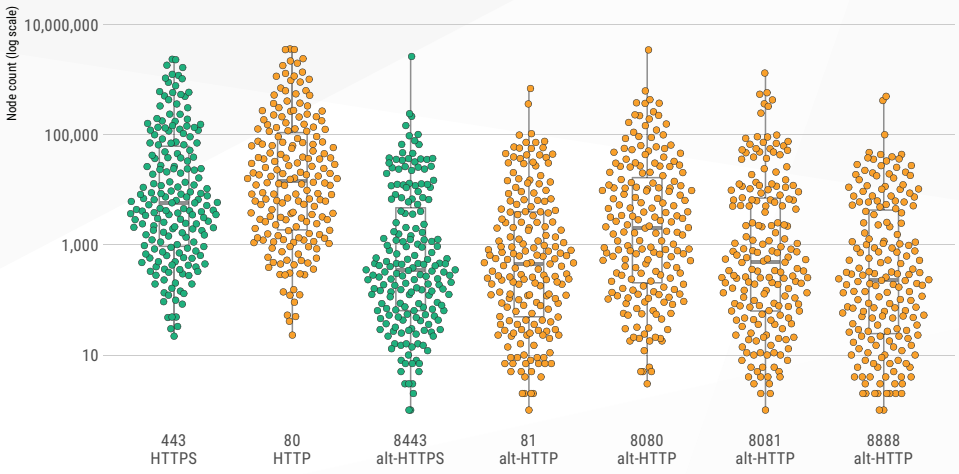


CHARACTERIZING TCP PROTOCOLS

In order to better understand the prevalence of ideal, encrypted services versus their cleartext counterparts, we've plotted out some selected relationships in this section. For this release of the National Exposure Index, we're using a somewhat fancy data visualization technique called a "quasirandom beeswarm" plot. While boxplots provide an overall sense of the distribution, the beeswarm overlay makes the picture of distribution a bit clearer. Boxplots alone can hide both density and outliers, whereas beeswarms let the analyst see the whole data set at a glance. In all cases, the green scatter plots denote encrypted services, while gold plots denote cleartext.

Figure 6: Total distribution of cleartext versus encrypted web services.

Each cluster shows the distributions of the count of number of devices per country exposing that port



Encrypted vs Cleartext Web Ports

The World Wide Web

Unsurprisingly, the World Wide Web (WWW) continues to dominate the internet[21]. Ports 80 and 443 account for 38.4% of all internet services, and when combined with the alternative, unprivileged ports for web hosting (81, 8000, 8080, 8888, and 8443), these services together account for 46.9% of the observed listening services across the internet. This is an overall increase by about 2% over our 2017 study (which did not account for ghost servers), but the number of distinct web **sites**, rather than **servers**, remained essentially static through 2017: there are over 1.78 billion responsive sites, according to the most recent Netcraft study[22].

Globally, there is about a 62% ratio of cleartext HTTP sites to encrypted HTTPS sites. While an all-HTTPS World Wide Web would be ideal from a confidentiality perspective, we're still in the early days of that transition here in 2018. That said, many HTTP sites exist only to forward requests to their HTTPS counterparts; it's uncommon for a website to be available only over HTTPS, even for high-security financial or webmail sites. Today's web browsers will tend to automatically attempt a port 80 connection when given a bare domain name (such as `yourbank.com`) rather than a fully-qualified URL (such as `https://yourbank.com`). It is nearly always better for servers to catch those requests and retry them securely rather than simply drop such requests on the floor. Unfortunately, this can create an opportunity for a man-in-the-middle attack, given the cleartext nature of HTTP[23]. Regardless of this risk, we expect to see legacy HTTP services available on the internet for a long, long time.

[21] For more on the nature of the WWW, please see pp 12-13 of the 2016 National Exposure Index.

[22] https://news.netcraft.com/archives/2018/04/26/april-2018-web-server-survey.html. Note that many sites can reside on one server. Curiously, this number is represents a slight 1.2% decrease in the total number of over 1.8 billion web sites surveyed by Netcraft in April of 2017.

[23] HTTP Strict Transport Security (HSTS), defined in RFC6797 (https://tools.ietf.org/html/rfc6797) does help alleviate some of these MitM concerns, but itself can be subject to abuse under certain circumstances. A complete discussion of the relative benefits and risks of HSTS is beyond the scope of this paper; the most secure method of connecting to a website is to directly navigate to https://example.com, rather than relying on opportunistic encryption.

REDEFINING TRUST ON THE INTERNET

This year will see two significant changes in the HTTPS landscape: the detrusting of certificates generated by Symantec's infrastructure (including certificates issued under the brand names Thawte, VeriSign, Equifax, GeoTrust, and RapidSSL), and the inherent trust placed by web browsers in the now-widespread LetsEncrypt certificates, a free and open certificate authority. Any discussion that positions HTTPS as "more secure" than HTTP would be remiss without also mentioning these developments.

In a way, these two issues are two sides of the same coin: On the one hand, the mechanisms and procedures used by Symantec were identified by the browser developer community to be insufficient, and examples of misused certificates were publicly identified[24]; as a result, Google Chrome and Mozilla-based browsers will, by default, no longer trust Symantec Certificates. The fallout from this Symantec detrusting event is expected to be at least noticeable, and possibly severe: As of April of 2018, about 9.24% of 9 million surveyed websites are encrypted with GeoTrust, Thawte, or Symantec certificates[25].

On the other hand, LetsEncrypt certificates—which are free and open, automatically generated, short-lived, domain-validated certificates[26]—now account for over 50% of all issued certificates across the same 9 million surveyed sites. LetsEncrypt certificates are trusted in all major browsers, yet it is possible (and in fact, common) to automatically generate LetsEncrypt certificates for obviously fraudulent and misleading websites intended for phishing and malvertising campaigns[27].

The difference between LetsEncrypt certificates and older Symantec certificates is not a difference in "security," despite the presence of browser warnings and red or green lock icons; instead, it is a difference in "trust." Both kinds of certificates are equally cryptographically secure at the moment of encrypting web content. LetsEncrypt certificates, however, will quietly and unobtrusively enable web traffic encryption, while detrusted Symantec certificates will trigger a browser failure—despite the fact that the definition of "trust" that might be inferred by a human looking at a green padlock is not the same as the technical definition of "trust" enforced by web browsing software.

In the end, we believe that while LetsEncrypt (and Symantec!) certificate-based encryption is useful and important for defending against pervasive monitoring, the LetsEncrypt and Symantec stories illustrate that "secure," "safe," "trusted," and "encrypted" all cover different, but overlapping, concepts.

---

[24] The thread that started this all was opened in January of 2017, at https://groups.google.com/forum/#!msg/mozilla.dev.security.policy/fyJ3EK2YOP8/yvjS5leYCAAJ

[25] This data provided by Nettrack.info, a free and open web statistics project, here: https://nettrack.info/ssl_certificate_issuers.html

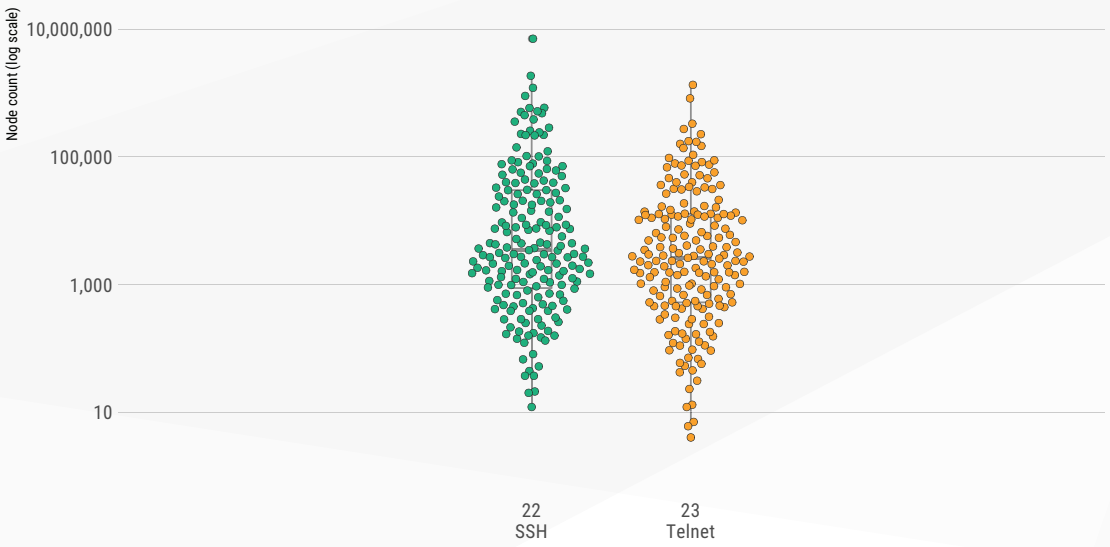
[26] Yes, that is a very long list of descriptors, all of which are important, even if it's unwieldy to read.

[27] See Netcraft's analysis, here: https://news.netcraft.com/archives/2017/04/12/lets-encrypt-and-comodo-issue-thousands-of-certificates-for-phishing.html.

Telnet vs. SSH

Figure 7: Distribution of SSH versus telnet services

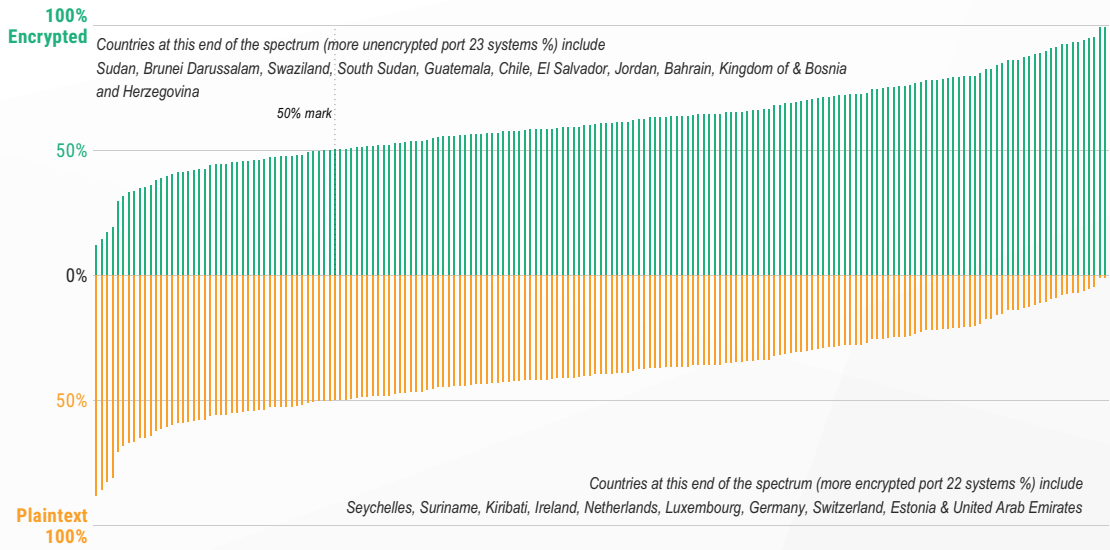
Each cluster shows the distributions of the count of number of devices per country exposing that port



Telnet and SSH

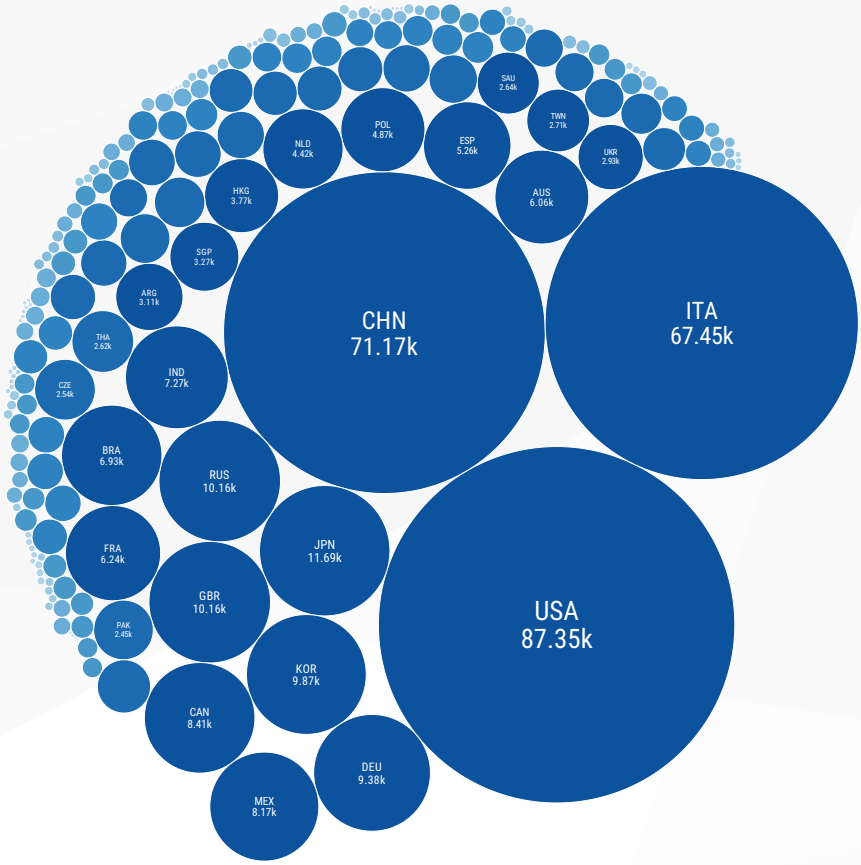
Figure 8: Distribution of plaintext and encrypted systems with shell access (ports 23 and 22)

Each column is a single country with the % of encrypted systems with shell access above the Y-axis and the % of plaintext systems with shell access below the Y-axis



This year, we're taking the kid gloves off and issuing an ultimatum: Telnet is seriously going to cost you next year in the rankings. For now, we've kept telnet at the same level of exposure as the previous years in the hopes (yet again) that the regular pwnage of IoT devices, routers, and legacy systems would have been a sufficient wake-up call to convince folks to do the right thing and stop enabling telnet. Yet, our optimism is regularly unwarranted, including this year. While the 50% mark (more counties with more nodes running SSH than telnet) is tacking port (a good thing), there is still way too much telnet out there (over 5 million nodes). If organizations cannot properly configure their internet-facing Cisco gear when there are repeated Cisco Smart Install vulnerabilities[28], then how can we expect an even more ubiquitous and handy service such as telnet to be removed?

Figure 9: SSH version 1.99 servers, by country



At least the story is a bit better on the SSH side. Over 97% of the SSH servers responding to full protocol probes are using SSH version 2.0. However, that still leaves nearly half-a-million nodes running version 1.99, with Italy taking the number two spot (Figure 9).

The Myth of Fingerprints[29]

Don't let those solid "2.0" numbers fool you into thinking all is well in SSH-land, though. An improperly configured SSH server can be as bad or worse than using telnet ("worse" in the sense that you think you're "safe" when you really aren't). Configuring things "securely" is hard, and SSH is no exception. There's a reason organizations like Mozilla publish guidelines[30] on how to properly configure systems. Yet, even with such guidelines there can be "gotchas." One easily overlooked "gotcha" is duplicate SSH host keys[31]. Because keys are part of the process of establishing trust between SSH clients and servers, security best practices state that each host should have a distinct, unique host key. Unique host keys prevent man-in-the-middle attacks before connections are made. There are some cases in internal host clusters where it might be OK to use duplicate keys, but there are many toolkits and frameworks around that make it possible to generate and manage SSH host keys at-scale.

[28] https://blog.rapid7.com/2017/09/20/cisco-smart-install-exposure/

[29] http://www.paulsimon.com/track/all-around-the-world-or-the-myth-of-fingerprints-2/

[30] https://infosec.mozilla.org/guidelines/openSSH

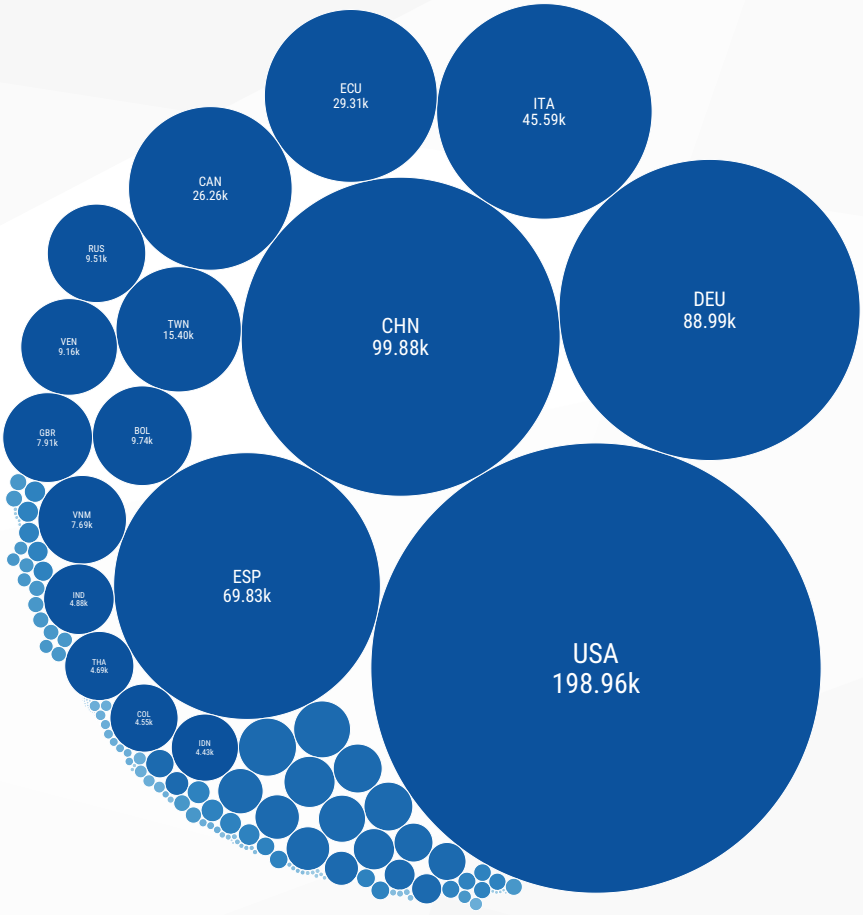
[31] https://www.SSH.com/SSH/host-key

Now, the vast majority of SSH 2.0 hosts in our study are running with unique host keys, but the ones that aren't definitely have an identity crisis. Unfortunately, sensible key management seems to be beyond the reach of nearly six million internet-accessible SSH hosts, since they all are dealing with some level of host key duplication. One of the most egregious examples of this is the host key, dc:14:de:8e:d7:c1:15:43:23:82:25:81:d2:59:e8:c0. This key was documented back in 2015 on the Shodan blog[32] as a problem, due to its replication (back then) on more than a quarter million hosts. Today, we found it still operational on about 62,000 hosts. More interestingly, we found another single SSH host key active on slightly more hosts, and this one doesn't appear to have been documented or reported to the vendor (or anyone else).

We're treating this finding of a massively duplicated host key—along with the 28 other SSH keys that have more than 10,000 instances exposed to the internet— as a vulnerability discovery, and are now in the process of coordinating disclosure of this issue through Rapid7's vulnerability disclosure[33] process. In the meantime, we can see in Figure 10 which countries have the worst problem with these egregious duplicate host identities.

Patterns that account for some of these bad configurations include the usual IoT suspects, such as routers, cameras, doorbells, and other associated toasters; cloud provider-sourced images with automation that do not regenerate the host key on creation; and users following those recipes for initial server configuration that are ill-written or woefully out of date.

Figure 10: Duplicate host keys, by country



[32] https://blog.shodan.io/duplicate-SSH-keys-everywhere/

[33] https://www.rapid7.com/security/disclosure#zeroday

Figure 11: Total distribution of exposed email ports

Each cluster shows the distributions of the count of number of devices per country exposing that port

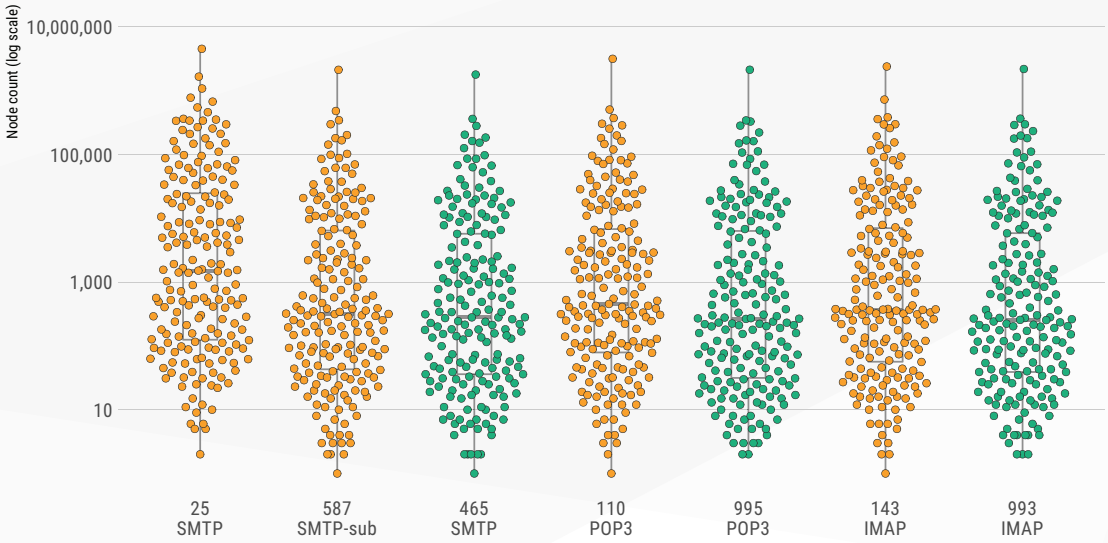


Figure 12: Distribution of plaintext and encrypted mail access (POP) systems (ports 110 and 995)

Each column is a single country with the % of encrypted mail access (POP) systems above the Y-axis and the % of plaintext mail access (POP) systems below the Y-axis

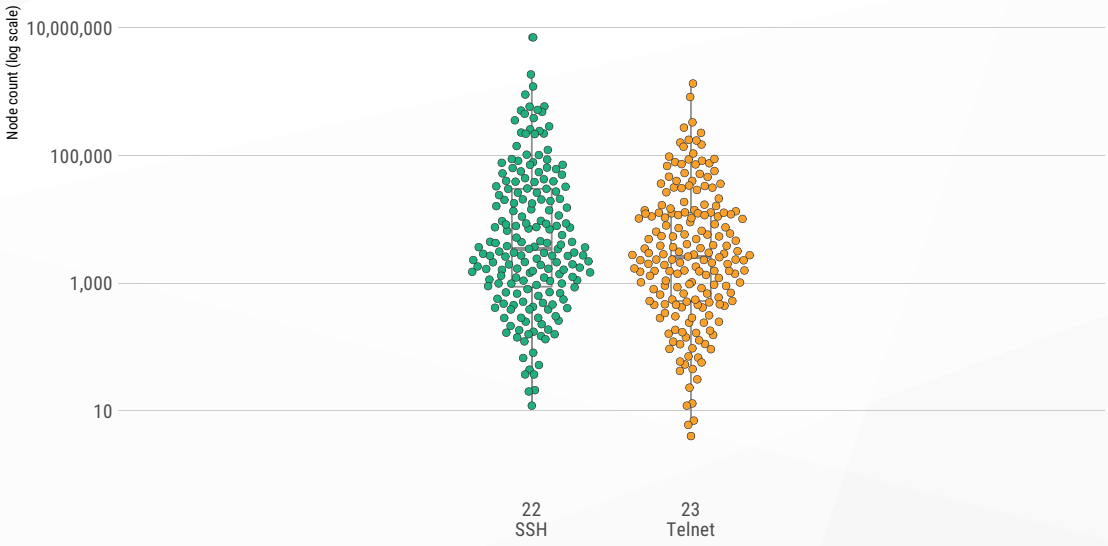


Figure 13: Distribution of plaintext and encrypted mail access (IMAP) systems (ports 143 and 993)

Each column is a single country with the % of encrypted mail access (IMAP) systems above the Y-axis and the % of plaintext mail access (IMAP) systems below the Y-axis

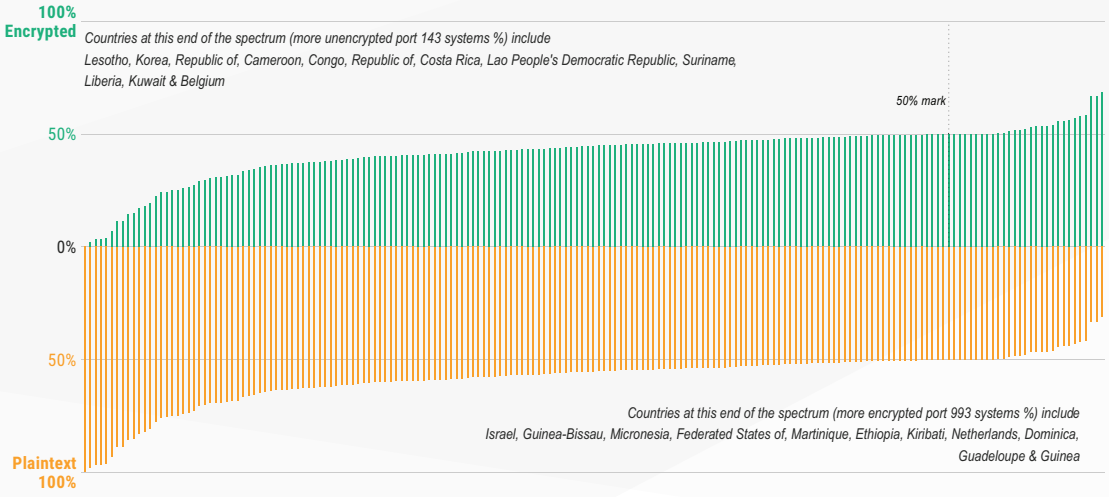
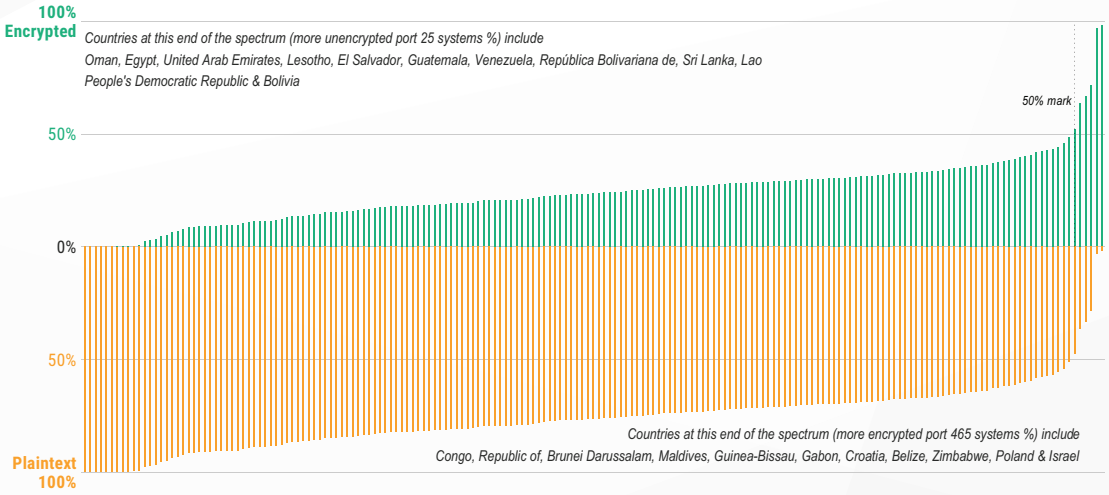


Figure 14: Distribution of plaintext and encrypted mail-oriented systems (ports 25 and 465)

Each column is a single country with the % of encrypted mail-oriented systems above the Y-axis and the % of plaintext mail-oriented systems below the Y-axis



Email Services

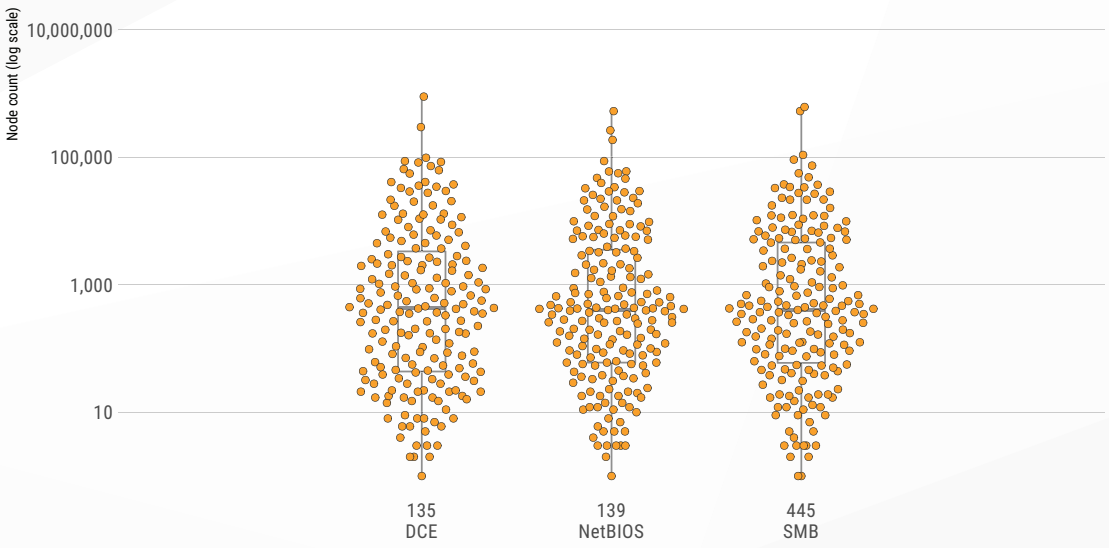
The set of ribbon graphs and beeswarm distribution delivers one strong message about natively encrypted, TLS-secured email servers: A few countries with relatively small internet footprints are taking email transport security very seriously, while most regions of the world are exposing much of their email infrastructure to passive monitoring. Email is an especially important online service for identity management; nearly every website of significance features an "I forgot my password" button that's linked directly to an email inbox. So, an adversary that is able to passively or actively read your email transmissions is almost certainly able to take control of virtually any other account you may rely on for your business, social, or cultural needs. This aspect of email as an identity skeleton key, alone, should be enough to warrant serious work in ensuring the transport encryption of email is available to its users.

As far as personal email goes—that is, emails that are not password resets, spam, legitimate marketing, or mailing list traffic—it is of course possible to integrate application-level security to ensure message integrity and source authentication through technologies like OpenPGP and S/MIME. Unfortunately, actually using these technologies consistently and correctly is notoriously difficult, and we need look no further than the "EFail"[34] disclosure dust-up in May of 2018 to understand just how hard it is to implement reasonable, user-controlled security into something as ubiquitous as email. While the giants of email service providers like Gmail and Outlook.com do employ transport layer security to protect the confidentiality and integrity, the fact that the vast majority of email servers on the internet are still stuck in the middle 1990s suggest that most small and medium-sized organizations are exposing their users to unnecessary exposure.

Microsoft SMB Services

Figure 15: Total distribution of exposed 'Microsoft' services

Each cluster shows the distributions of the count of number of devices per country exposing that port



We penned this section soon after the one-year anniversary of WannaCry[35] and near the anniversary of NotPetya[36]. Both exploits wrought havoc to organizations, municipalities, and individuals across the globe, and they continue to do so even into 2018. Organizations have lost millions of dollars in sales, productivity, and physical damages (yes, that's right … these attacks hurt manufacturing production lines as well as office systems). One would think, or at least hope, that we'd find little-to-no traces of active SMB on the internet after a banner year of attacks. Sadly, this is not the case. Figure 16 shows there was a dramatic decrease in exposure following the release of the Shadow Brokers treasure trove of exploits[37] and a further decrease post-WannaCry. Now, we're in a holding pattern with the total daily counts hovering around 500,000 nodes.

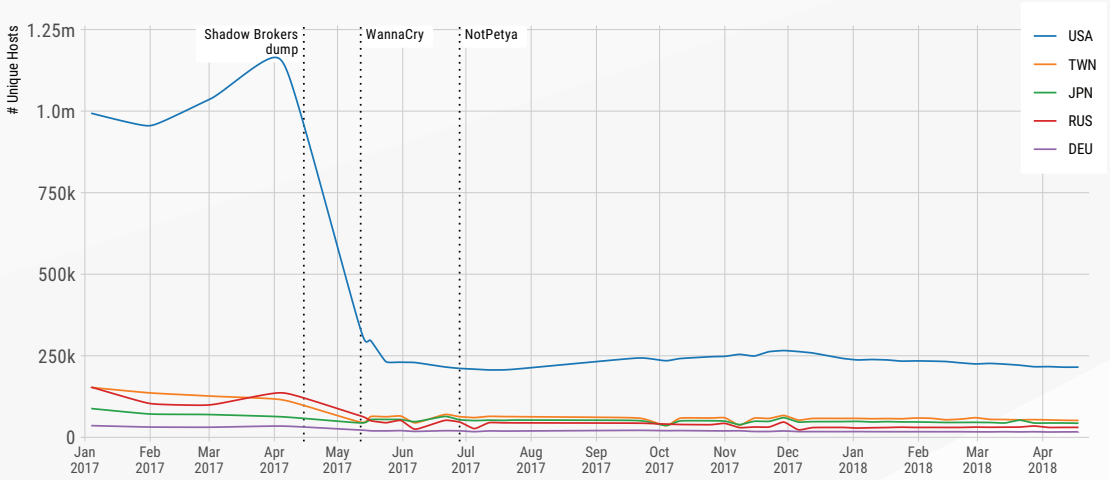
[34] https://efail.de

[35] https://www.rapid7.com/security-response/wanna-decryptor/

[36] https://blog.rapid7.com/2017/06/27/petya-ransomware-explained/

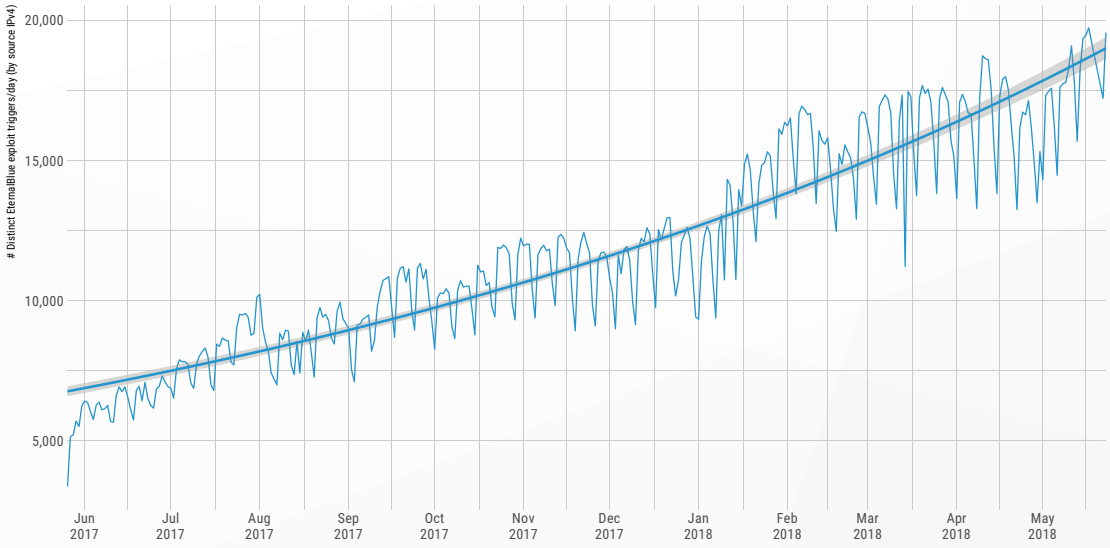
[37] https://blog.rapid7.com/2017/04/18/the-shadow-brokers-leaked-exploits-faq/

Figure 16: Open Microsoft SMB (TCP port 445) servers—Top 5 countries



On the surface, a reduction of this magnitude would seem to be a positive step in the right direction. However, the SMB-based attacks persist, as seen in Figure 17. We've tuned Project Heisenberg[38], our global honeypot network, to watch for EternalBlue-based exploit attempts and have also used enhanced intelligence from GreyNoise[39] to filter out known non-malicious "researcher" attacks. We see a steady increase in malicious tool use across more unique IPv4 addresses each week. As a result of the lack of progress in clamping down further on exposed SMB and the increased desirability of SMB by attackers, we've chosen to weight SMB exposure very highly in the 2018 rankings computation.

Figure 17: Daily unique EternalBlue exploit attempts



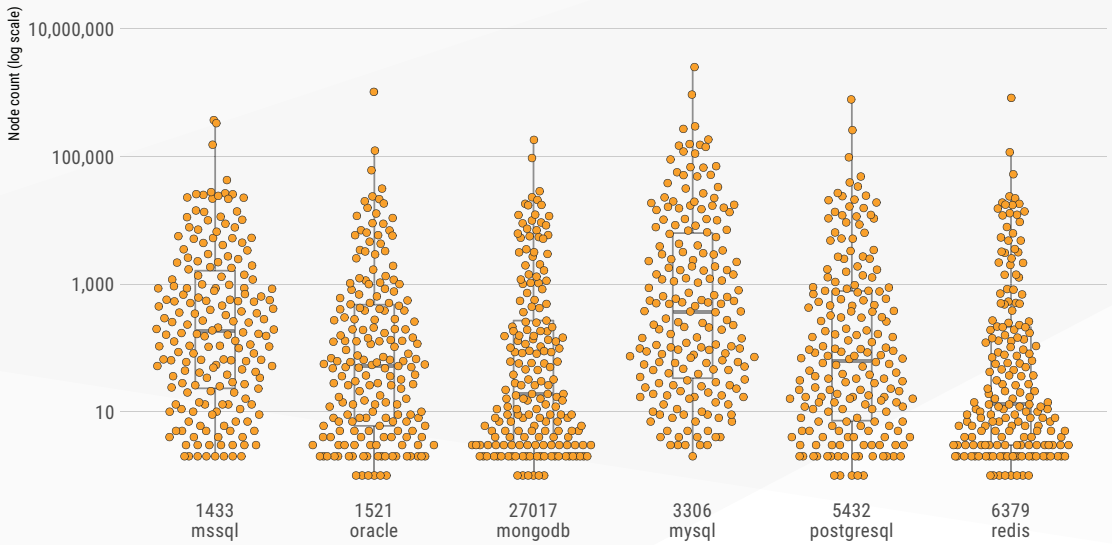
[38] https://opendata.rapid7.com/heisenberg.cowrie/

[39] https://greynoise.io/

Databases

Figure 18: Total distribution of exposed database ports

Each cluster shows the distributions of the count of number of devices per country exposing that port



Database Exposure

This year, Oracle DB[40], MongoDB[41], PostgreSQL[42], and Redis[43] all join MySQL and Microsoft SQL on our list of database services we probe for. Our expectation was that each of these solutions have their own, distinct legacies of risk and exposure that should be sufficient harbingers of doom as to ensure a miniscule representation in the National Exposure measurements. Alas, we've found these technologies are alive and ... well ... exposed. Sure, it's possible to set up certificate-based authentication over encrypted channels to most of these database servers, but every database administrator we talked to advised against it; if you must reach your database over the internet, you're far better off setting up an SSH tunnel or other VPN service in front of it. In addition, many of these internet-exposed databases appear to be legacy, unmaintained systems: well over 100,000 Microsoft SQL Server nodes were willing to tell us they were running on versions ranging from v7 to v14 (the most recent version).[44] Similar stories exist for MongoDB and other databases with an unhealthy mix of extremely old and very current versions running fully exposed for anyone to access or even ransom[45].

In September of 2017, we saw a fairly major ransomware attack that targeted unsecured, exposed MongoDB servers, and to our knowledge, this was the first major database-based attack since the SQL Slammer incident of January, 2003[46]. However, we can see on the distribution graph above there are plenty of other database targets, any one of which can lead to a tremendous loss of data through ransoming, breach, or deletion. For this reason, we weight the exposure of these ports fairly high in our ultimate exposure scoring.

---

[40] http://www.oracle.com/technetwork/database/enterprise-edition/overview/index.html

[41] https://www.mongodb.com/

[42] https://www.postgresql.org/

[43] https://redis.io/

[44] We'll never perform unauthorized, invasive probes or use credentials, well known or otherwise, to gain access to systems to determine these version levels—these results were solely from a general protocol exchange.

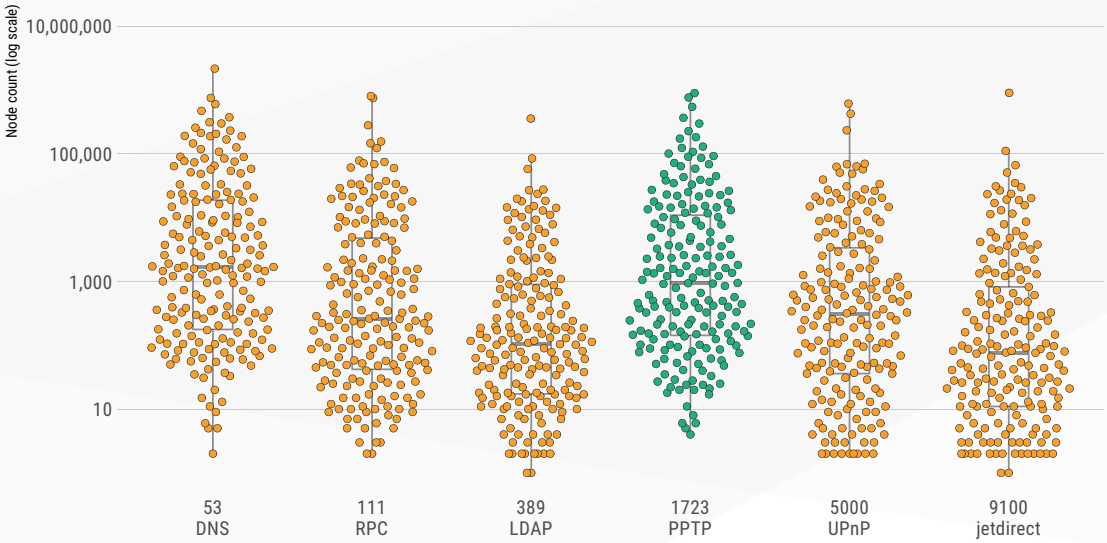
[45] https://blog.rapid7.com/2017/01/30/the-ransomware-chronicles-a-devops-survival-guide/

[46] Well, not since 2003, anyway, when SQL Slammer was released over the American Superbowl weekend.

Everything Else TCP

Figure 19: Total distribution of other TCP services

Each cluster shows the distributions of the count of number of devices per country exposing that port



We've plotted out the remaining TCP ports on the figure above, and with the exception of DNS, none of these services have good reasons for being exposed to the public internet. For example, PPTP is an old, vulnerable VPN service that relies on weak, easily cracked encryption standards. One positive change to report is that the PPTP distribution we've been measuring for three years has finally started to drop significantly. We expect this is a side effect of general infrastructure upgrades: as more and more enterprises retire their old infrastructure in favor of more modern VPN solutions, or migrate entirely to cloud providers, the old PPTP endpoints are getting powered off, permanently.

Even though UDP is stateless, underlying application protocols can implement their own state counters and therefore offer delivery correctness guarantees on responses that exceed the nominal size of a UDP packet.

CHARACTERIZING UDP PROTOCOLS

UDP services break down into two categories: services that are inappropriate for internet exposure (as with the TCP services, above), and a new classification of exposure known as **amplification potential**.

As mentioned earlier, UDP services differ from TCP in one crucial aspect, in that they do not require an establishing connection phase. Instead, clients wishing to use UDP services simply fire off a request and await a response (possibly several) without any guarantee of delivery on either end.

In the early days of internet engineering, this design choice seemed like a reasonable mechanism to implement lightweight, single-message services in an efficient way. In DNS, the Domain Name Service, clients typically asked for one peice of data (the human-readable name of an internet-connected IPv4 node) and got one reply back with that name (or the lack of a record). This two message conversation could be done quickly and easily without the bother of connection setup, monitoring, and teardown.

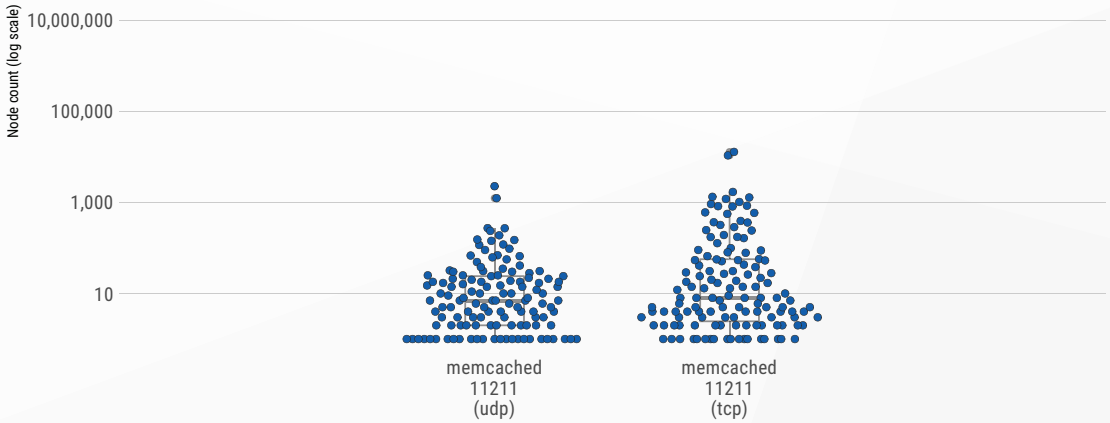
The problem with this design is that it does not account for malicious actors with the ability to forge their own source addresses with a technique called "spoofing." When an attacker spoofs their IP address as another node on the internet, that node will get the reply to any request, no questions asked. In addition, even though UDP is stateless, underlying application protocols can implement their own state counters and therefore offer delivery correctness guarantees on responses that exceed the nominal size of a UDP packet.

These two features make it possible to create application protocols that expect small, client-side queries, and large, server-side responses—responses which are sent to someone other than the original requestor. This sets up a service's amplification potential, where an attacker can induce a server on the internet to throw a bunch of data at an unsuspecting third party for relatively low cost to the original attacker.

Memcached: Amplification, Amplified

Figure 20: Total distribution of Memcached services (TCP and UDP)

Each cluster shows the distributions of the count of number of devices per country exposing that port



To better understand amplification potential as an exposure, imagine the prank of ordering 10 pizzas to be delivered to your frenemy's house. The only way the delivery driver knows where to go is because you told the restaurant—there's no need to prove it, and no way for you to know with certainty if or when the pizzas will arrive at your given "source" address. At the cost of a short phone call, you can create a fairly dramatic response for someone else to deal with.

At the end of February of 2018, researchers from Cloudflare[47] noticed a mechanism to order tens of thousands of cyber-pizzas from thousands of cyber-restaurants: Memcached, a popular caching service that (surprisingly) also operates over UDP to provide large replies of encoded data to small requests. Most amplification attacks use UDP replies that are only two or three times the size of the original request, but memcached can offer a ~137,000 byte-sized reply to a mere ~15 byte request under certain circumstances.

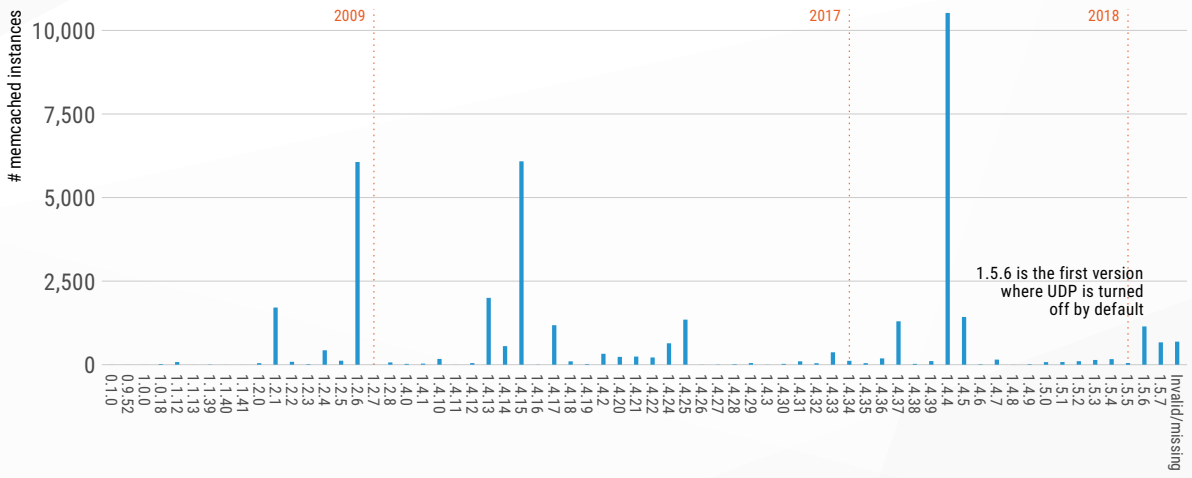
Thanks to this event, we concluded that amplification potential is serious enough to warrant counting it against the overall exposure of a country's IP space[48]. After all, these are resources that can be abused by malicious actors, either to cause distributed denial of service (DDoS) events on their own, or in conjunction with more serious attacks. If some countries are more prone than others to be the source of amplification attacks, it might be reasonable for their neighbors to pre-emptively protect against this threat through blacklisting, which, in turn, reduces the utility of the offending country's internet presence.

Memcached over TCP

The fact that memcached operates over UDP at all was a surprise to many security researchers; the normal use case for memcached is as a TCP service, and normally situated nearby is a client web server in the same layer as a database server, usually unreachable over the internet directly. Naturally, while looking at the clear and present danger of memcached over UDP over the internet, we also took a look at memcached over TCP and got some fairly worrisome results.

Figure 21: Version distribution of TCP memcached services

Over 7% of these nodes are accessible via UDP (which means they are capable of being mindless ampli-bot nodes)

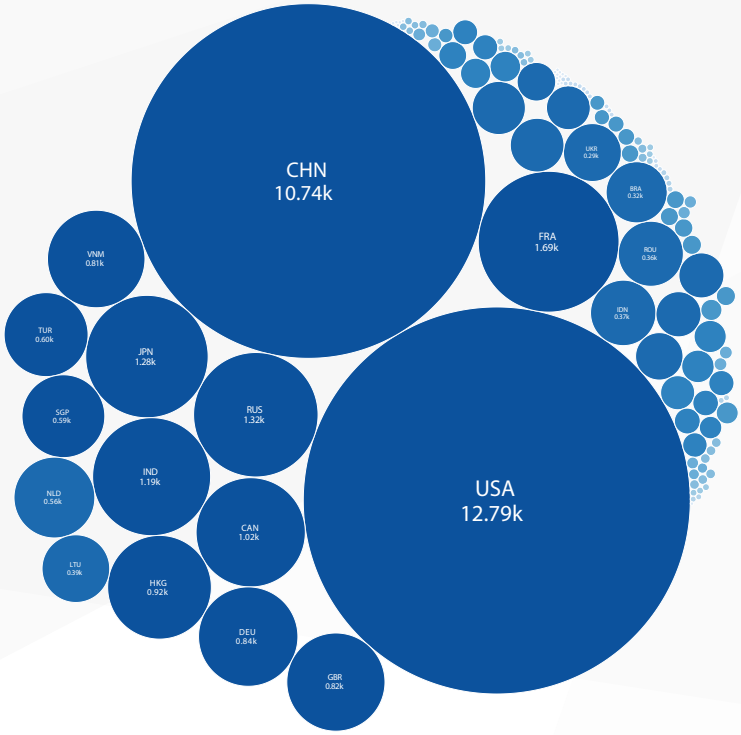


[47] https://blog.cloudflare.com/memcrashed-major-amplification-attacks-from-port-11211/

[48] For a more complete discussion of memcached and other UDP amplification services, see the 2018 Q1 Threat Report by Rebekah Brown, Kwan Lin, and Bob Rudis, available at https://www.rapid7.com/info/threat-report/2018-q1-threat-report/.

While there remain only about 3,500 memcached UDP servers on the public internet, there is an order of magnitude more apparent memcached TCP servers[49]. What's more, there are some distinct clusterings of versions in production today, with significant populations running versions 1.2.1, 1.2.6, 1.4.13, 1.4.15, and 1.4.4. Recall, the UDP threat was not clearly understood until version 1.5.6, so it's likely these memcached servers are both unmaintained and present a risk for enabling the UDP functionality if it's not yet enabled today, in addition to a number of other vulnerabilities that have impacted memcached over the years. Memcached servers, with weak and unmaintained configurations, are a significant, ongoing threat to the stability of the internet. The regions with the largest live deployment of these servers, as shown in the circlepack graph in Figure 22, are urged in the strongest terms to deal with these DDoS-A landmines with all due haste.

Figure 22: Memcached country distribution

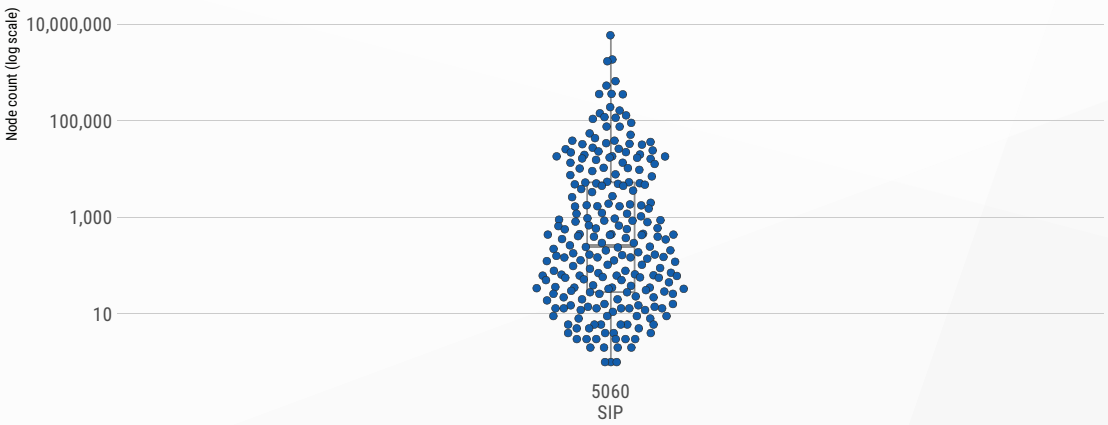


Cleartext SIP

Far and away, the most popular UDP protocol we captured in our survey is 5060, associated with the Session Initiation Protocol (SIP), most commonly used in a Voice over IP (VoIP) application. There are over 14 million nodes[50] responsive to our probes, almost twice as many as the next most common UDP protocol, DNS. While it is common to expose this port to the internet to support an enterprise's external phones, the usual advice today is to limit access to this port through network ACLs or VPN tunnels. Exposing VoIP signaling data directly over UDP over the internet creates exposure through passive monitoring for usernames, address books, pins, and recorded voicemail, as well as active monitoring for traffic analysis purposes. Today, all popular SIP implementations offer native TLS encryption over port 5061, which offers the usual cryptographic guarantees of authenticity and confidentiality.

Figure 23: Total distribution of SIP (5060) services

Each cluster shows the distributions of the count of number of devices per country exposing that port



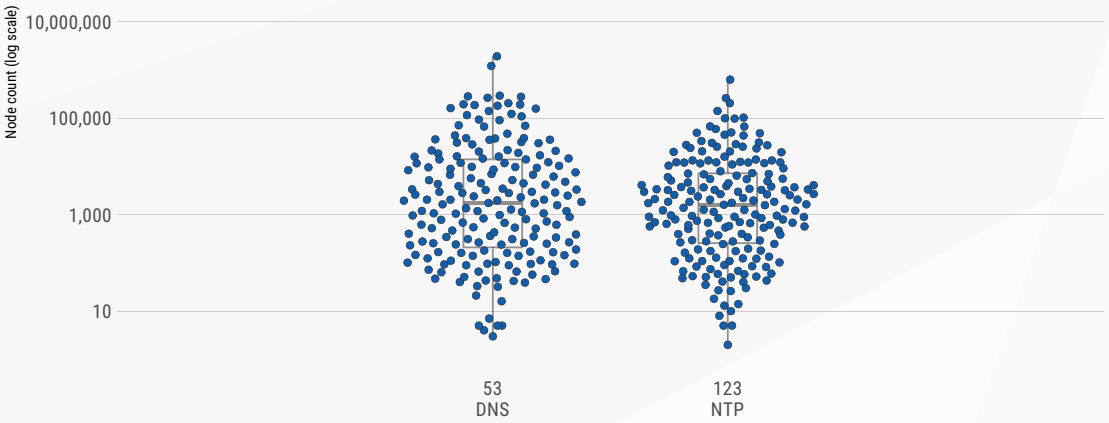
[49] The top two outliers in the TCP beeswarm are the United States with 12,791 nodes, and China with 10,741 nodes. Similarly, in the UDP beeswarm, we see China with 2,259 nodes and the U.S. hosting 1,234 nodes.

[50] The top three outliers in the SIP beeswarm are Germany, with 5.9 million nodes, Japan with 1.9 million nodes, and Saudi Arabia, with 1.7 million nodes.

DNS and NTP

Figure 24: Total distribution of DNS and NTP (UDP) services

Each cluster shows the distributions of the count of number of devices per country exposing that port

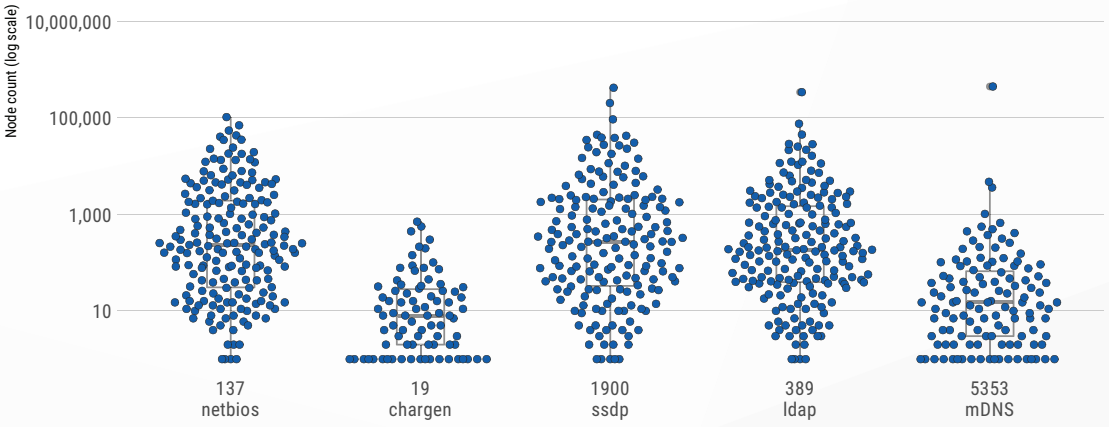


Unfortunately, we are stuck with DNS and NTP. For all their faults, we cannot reasonably expect a national internet infrastructure to implement trusted, secure timekeeping or domain name systems. That said, an alternative name service protocol (such as DNSCrypt[51] and DNS-over-HTTPS [52] is being pursued. These laudable engineering efforts from OpenDNS, Cloudflare, and other large-scale domain name service providers are signs of hope for eventual and ubiquitous reasonable encryption for this critical service. Even though they're long shots today, these projects demonstrate at least some desire to get out of the cleartext woods with DNS. Depressingly, there seems to be no such coordinated and funded effort today to replace NTP with a secure alternative.

The Rest of UDP

Figure 25: Total distribution of remaining UDP ports

Each cluster shows the distributions of the count of number of devices per country exposing that port



For completeness, Figure 25 plots the rest of the UDP-based protocols we scan for. As with other inappropriate services, we can think of no practical reason to expose these particular services on the internet. For example, multicast DNS, on port 5353, is designed specifically for small, local networks and doesn't even work in an internet context. And yet, South Africa has the highly suspicious finding that it is making mDNS available to the tune of 447,000 nodes.

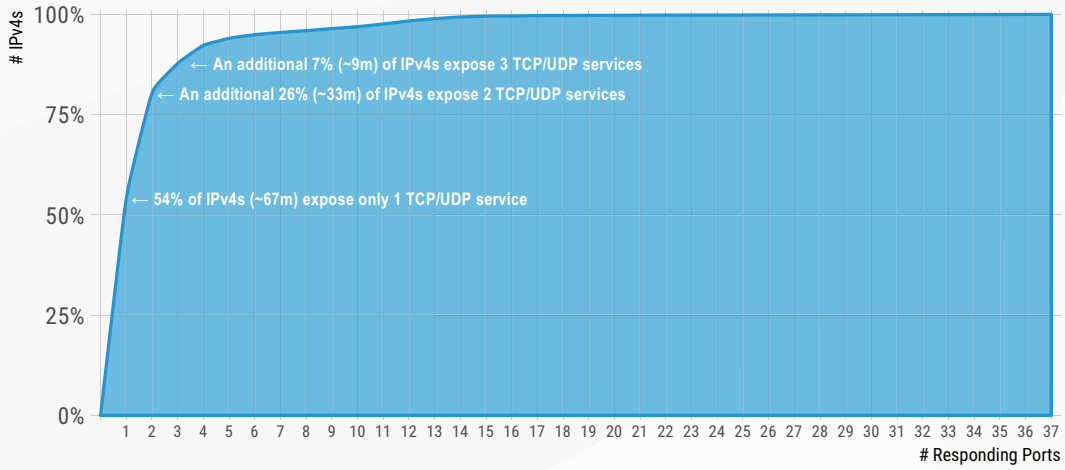
For the rest of the topmost outliers, we find these populations of inappropriate and amplification-prone services hosted in the United States, China, and Russia. This offers even more evidence to contribute to their top exposure rankings.

[51] OpenDNS DNSCrypt <https://www.opendns.com/about/innovations/dnscrypt/>

[52] Cloudflare DNS-over-HTTPS <https://developers.cloudflare.com/1.1.1.1/dns-over-https/>

Ports Per Address

Figure 26: Distribution of the number of open ports per address



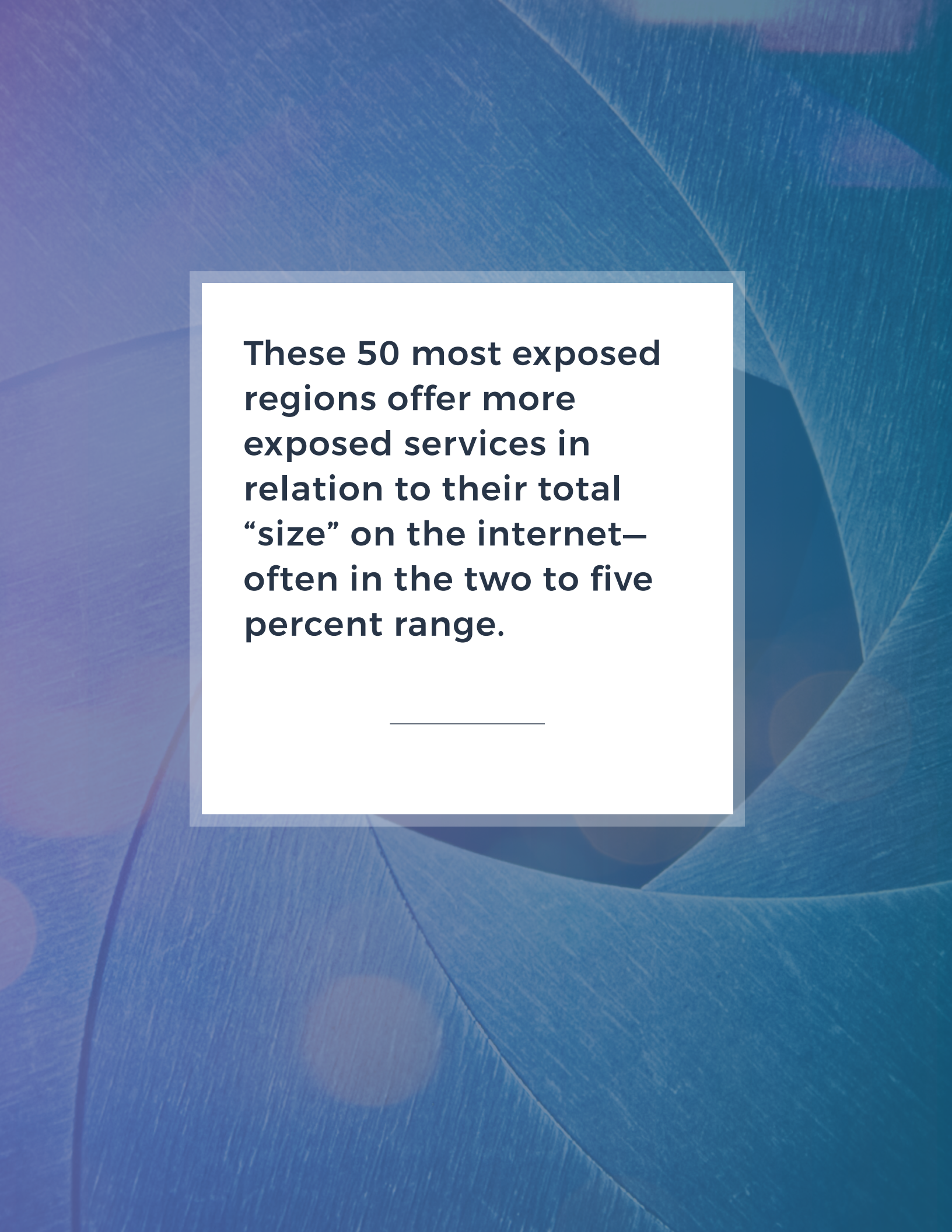
When a node runs more than one service, the complexity of that system increases, as does the attack surface and threat event potential. Rapid7's Project Heisenberg sees daily, malicious probes and attacks across every port/service identified in the National Exposure Index, along with a multitude of other ports/protocols/services. Sure, modern systems are capable of handling additional and increasingly diverse workloads more than ever before, but that doesn't mean we should be increasing the complexity of individual nodes at the expense of node or service resilience/safety.

It is difficult to relay the year-over-year difference for this "complexity exposure" due to our enhanced discovery methods and increased port/service coverage, as mentioned under "Changes to Methodology" at the start of this paper. That said, readers familiar with prior National Exposure Index releases will recall that systems exposing only a single port in the 2017 scans constituted 56% of the overall complexity distribution. This year, that number fell to 50%, while we only added 7 new TCP ports to the scan mix.

Now, the vast majority of single port systems are running HTTP/HTTPS/mail/DNS services, and the dual-port/service systems are mostly running combinations such as HTTP/HTTPS, SMTP/(POP[S]|IMAP[S]), or DNS/[some other service]—in other words, service pairs you would reasonably expect to see in production. The complexity is fairly low for such setups, but all of those individual components require solid configurations and regular patching. Once a single node is hosting three, four, or five plus services, the complexity exposure of that system rises dramatically; not only are such systems extremely difficult to patch and maintain with regularity, but an exploit on one service can lead to an impact of all services hosted on that system, either directly through a compromised root shell or through the access gained in a non-root context. This risk is especially relevant when considering that the Spectre and Meltdown[53] vulnerabilities can expose kernel memory pages inapporpriately to unprivileged, but compromised, processes.

That said, the fact that approximately 80% of systems are running low-complexity and fairly "safe" services or service pairs should be considered an overall positive general indicator. Organizations can use this complexity exposure metric on their own perimeters to gain an understanding of what additional risks they may be facing.

[53] https://meltdownattack.com/



These 50 most exposed regions offer more exposed services in relation to their total “size” on the internet—often in the two to five percent range.

NATIONAL EXPOSURE INDEX

The ranking system used for this study is fully described in Appendix C, but briefly, a country with a higher **percentage** of exposed services in relation to its total allocated IP address space will tend to score higher on National Exposure. In addition, those countries that have **confirmed** Microsoft SMB exposed to the internet are weighted even higher. Finally, we've added more complete coverage for database ports, and as a result, we're weighting those results as well; databases tend to hold many keys to many kingdoms, so exposing those services is a serious issue that any national plan for internet management should address.

In short, these 50 most exposed regions offer more exposed services in relation to their total "size" on the internet—often in the two to five percent range—and often offer more exposed services in absolute terms as well. The least exposed regions tend to expose well under one percent of their IP address space, and also offer small target populations of exposed services in absolute terms.

Table 8: Top 50 most exposed countries

TOP 1-10	TOP 11-20	TOP 21-30	TOP 31-40	TOP 41-50
United States	Australia	Spain	Sweden	Malaysia
China	Brazil	Finland	Czech Republic	Portugal
Canada	Hong Kong	Romania	Belgium	Venezuela
South Korea	Russian Federation	Thailand	Denmark	El Salvador
United Kingdom	Poland	Singapore	Colombia	New Zealand
France	South Africa	Indonesia	Egypt	Saudi Arabia
Netherlands	India	Ireland	United Arab Emirates	Kazakhstan
Japan	Turkey	Vietnam	Norway	Brunei Darussalam
Germany	Israel	Ukraine	Pakistan	Bangladesh
Mexico	Iran	Chile	Austria	Peru

Country Re-Rankings

In the first year of the National Exposure Index, after we measured all the unique IPv4 addresses exposing one or more of thirty services, we calculated the exposure percentages of each service based on that total number of unique responders. Last year, percentages were instead based on total **allocated** (rather than merely possible) IPv4 space. This year, we retain this notion of allocated-vs-total, but we further added an "anti-weighting" for nodes that respond to canary ports, described earlier in this paper. The rationales for these changes over the three studies were manifold. First, we again compared our SYN responses to Censys ICMP[54] responses and CAIDA[55] estimates, and came up with the same results that our probes—while measuring exposure of certain services—do not capture all in-use devices on country IPv4 networks. Second, large cloud service providers cause very dynamic usage patterns of IPv4 space within the networks they occupy. One "bad" day for a given provider (i.e. a large number of exposed servers being spun up) in a given region could easily skew results in either direction.

[54] https://censys.io/data/0-icmp-echo_request-full_ipv4

[55] http://www.caida.org/

While our SYN scans measure potential exposure, we had an opportunity to dig deep into the Windows SMB protocol and measure actual exposure to the various EternalBlue-based threats for each country. Since we also had more protocol-level scans this year and a better understanding of how the canary ports factor in to skewing the results plus errors inherent in geoip coding, we used these conditions to come up with an observational error factor of 10% to help ensure a fairer overall ranking system that better emphasized likely exposure. If the observed number of IPv4 endpoints for a given scan was less than 10% of total found IPv4s for a given country, then we lowered the weight of the ranking for that particular service for that particular country. Those services are still factored into the rankings, but only after more reliable results for other countries have been factored in. We believe these adjustments better reflect individual exposure and cause less skew in either direction. Unless something radical happens with our thinking in the coming years, we expect this approach to measuring national exposure will endure and we can begin truly long-term longitudinal studies of the internet server ecosystem.

CONCLUSIONS

At the beginning of this report, we made especial note that we've updated our ranking strategies, algorithms, and contributing data, as well as expanded our definition of "exposure" to account for the risk of services misuse to contribute to devastatingly amplified denial of service attacks. While these updates make it impossible to make reasonable comparisons to past National Exposure results, we're confident that readers can use this year's ranking as a basis for immediate action, as well as long-term, year-over-year comparisons going forward. The fact that the United States and China are the top two most-exposed countries is probably unsurprising; that said, the authors of this paper believe that the data-rich ranking system described here is a remarkably useful tool to come to this conclusion, especially for the top 50 most exposed countries.

Globally, we continue to see some disturbing trends in internet exposure, the most significant being that even headline-grabbing attacks against inappropriate services such as Windows SMB, database services, and powerful amplification services are not enough to truly zero out their ongoing risk to attack and misuse. Even as there are engineering efforts to bolster the domain name system and bring it to modern levels of encryption and security, we still see millions of poorly maintained, misconfigured computers, ready to be abused by intelligence and espionage agencies, sophisticated criminal organizations, and casual, unsophisticated threat actors.

With that bit of doomsaying, we are encouraged by the reception of the National Exposure Index. After all, it's impossible to solve these problems without first measuring them, and this report continues to be an important tool in conducting those measurements. If you are a part of the technical leadership for your region's local internet, either in an engineering or a policymaking capacity, we'd like to invite you to contact us directly at **research@rapid7.com** to discuss the topics raised in this paper. At Rapid7, we strongly believe that, given the right tools and education, we can rebuild our networks to be more safe, more stable, and more secure media for commerce and culture.

APPENDIX A: PROJECT SONAR

Project Sonar[56] started in 2013 as a security research project with the goal of helping the larger information security community understand global exposure to security vulnerabilities. Sonar conducts frequent surveys or studies using publicly available information, and from this information one can infer security vulnerabilities, misconfigurations, or simply protocol/product usage on a global basis.

The vast majority of the work that Sonar does is in 'active' studies, where the primary goal is to inspect a given service on every public IPv4 address and collect intelligence about each endpoint. In the simplest of cases, this intelligence might simply be that the port is or is not open. In more complex cases, this intelligence gathering process might attempt to negotiate the protocol expected on the endpoint in question and perform additional reconnaissance. For example, for studies of the HTTP protocol, Sonar will attempt to negotiate a TCP connection to the respective HTTP endpoint on every public IPv4 address, and over that connection a HTTP request will be sent. The resulting HTTP response is saved, and from this information all manner of intelligence can be gained.

Sonar currently performs studies of over 70 different services. These studies are performed on a regular basis; some with a weekly cadence, others with a lower frequency.

Every study that Sonar performs is done with an additional goal of ensuring that the study is legal and as non-disruptive as possible. This means two things. First, every Sonar study must stay within the bounds of United States law. More specifically, this means that no Sonar study will attempt to circumvent or bypass any technical controls in the course of its collection activities. Second, recognizing that reconnaissance activities like this on the part of good-willed researchers and organizations might be confused for malicious activity or otherwise be disruptive to an organization's security operations, Sonar has established a process by which organizations can be excluded from Sonar's activities.

As previously hinted at, one of Sonar's primary goals is to provide data to the larger information security community. While some individuals may be able to obtain this data on their own, oftentimes the data acquisition process can be time-consuming, costly, and legally risky for the unprepared. To this end, Sonar strives to publish as much data from these studies as possible through Rapid7's Open Data[57].

[56] https://sonar.labs.rapid7.com/

[57] https://opendata.rapid7.com/

Complete Port Scan Target List

The choices of these 37 TCP ports and 9 UDP ports was guided by both the nmap services list and the collective wisdom of Rapid7 researchers.

The top 15 TCP protocols are one-for-one matches with the most frequent protocols identified by a series of private nmap scans of the internet conducted in 2008 and updated in 2017, while the remaining 22 are protocols that we hypothesized should occur fairly routinely and speak directly to exposure. As for UDP, we are limited to those protocols from which we can elicit a response on a request packet; of those, we selected the nine that, like TCP, are most likely to indicate something security-relevant about the target host.

Table 9: All ports scanned for National Exposure Index

PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	APPROPRIATE?	ENCRYPTED?	DESCRIPTION
5	2,854,356	canary-low	N/A	N/A	A low-number unassigned port that should not normally respond to SYN/ACK
19	3,756	Chargen (UDP)	FALSE	FALSE	Chargen, a service that echos a list of characters
21	13,359,961	FTP	FALSE	FALSE	File Transfer Protocol, used to send and receive data and text files; FTPS, SSH, and HTTPS are all encrypted alternatives
22	19,061,180	SSH	TRUE	TRUE	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
23	5,814,024	telnet	FALSE	FALSE	Telnet, a remote command shell interface, one of the oldest protocols on the internet; SSH is an encrypted alternative
25	15,664,213	SMTP	TRUE	FALSE	Simple Mail Transfer Protocol, used to send email
53	8,832,463	DNS	TRUE	FALSE	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
53	7,352,839	DNS (UDP)	TRUE	FALSE	Domain Name Service, used to resolve human-memorable names to IP addresses
80	62,656,633	HTTP	TRUE	FALSE	HyperText Transfer Protocol, used to serve web pages and web applications
81	2,464,657	http-alt	TRUE	FALSE	A common alternative port for HTTP, usually used for web sites and web proxy services
110	7,114,795	POP3	TRUE	FALSE	Post Office Protocol version 3, used to receive email
111	3,375,227	rpcbind	FALSE	FALSE	Remote Procedure Call port mapping service, usually used on Unix-like operating systems, usually for NFS file sharing
123	2,738,152	NTP (UDP)	TRUE	FALSE	NTP, the Network Time Protocol
135	2,437,524	MS-RPC	FALSE	FALSE	Microsoft Remote Procedure Call, usually used on Microsoft OSes for distributed computing

PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	APPROPRIATE?	ENCRYPTED?	DESCRIPTION
137	737,185	NBSN (UDP)	FALSE	FALSE	NetBIOS Name Service, used in NetBIOS over TCP/IP, usually on Microsoft OSes for file and print sharing
139	1,934,357	NBSS	FALSE	FALSE	NetBIOS Session Service, used in NetBIOS over TCP/IP, usually on Microsoft OSes for file and print sharing
143	6,668,963	IMAP	TRUE	FALSE	Internet Message Access Protocol, used to receive email
389	862,686	LDAP	FALSE	FALSE	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
389	810,656	LDAP (UDP)	FALSE	FALSE	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
443	44,849,191	HTTPS	TRUE	TRUE	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
445	3,507,183	SMB	FALSE	FALSE	Server Message Block, a file transfer and remote administration protocol for (usually) Microsoft operating systems
465	4,416,327	SMTPS	TRUE	TRUE	Secure SMTP, an encrypted-by-default alternative to SMTP
587	5,262,246	SMTP submission	TRUE	TRUE	SMTP submission service, usually used by endpoint mail clients to send email
990	1,046,579	FTPS	TRUE	TRUE	Secure FTP, an encrypted-by-default alternative to FTP
993	5,155,630	IMAPS	TRUE	TRUE	Secure IMAP, an encrypted-by-default alternative to IMAP
995	4,997,893	POP3S	TRUE	TRUE	Secure POP3, an encrypted-by-default alternative to POP3
1433	1,369,495	MSSQL	FALSE	FALSE	Microsoft SQL Server service, used to communicate with Microsoft database servers of the same name
1521	1,489,749	oracle	FALSE	FALSE	Oracle Database listening service, used to communicate with the T-SQL server of the same name
1723	5,334,237	PPTP	FALSE	FALSE	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers.
1900	1,289,184	SSDP (UDP)	FALSE	FALSE	Simple Service Discovery Protocol, used with UPnP
3306	6,087,830	MySQL	FALSE	FALSE	MySQL, used to communicate with the (usually) open source MySQL Server published by Oracle

PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	APPROPRIATE?	ENCRYPTED?	DESCRIPTION
3389	4,934,495	RDP	FALSE	FALSE	Remote Desktop Protocol, a graphical user interface to remotely administer (usually) Microsoft Windows servers and desktops
5000	2,300,261	uPNP	FALSE	FALSE	Universal Plug-and-Play, a protocol for machine-to-machine discovery and configuration
5060	14,001,928	SIP (UDP)	FALSE	FALSE	Session Initiation Protocol, usually used in Voice over IP applications
5353	463,924	mDNS (UDP)	FALSE	FALSE	Multicast DNS, useful in networks without dedicated name services
5432	1,594,877	PostgreSQL	FALSE	FALSE	PostgreSQL listening service, used to communicate with the T-SQL server of the same name
5900	1,142,393	RFB	FALSE	FALSE	Remote Frame Buffer, a remote GUI for desktop administration, usually used by VNC (Virtual Network Computing)
6379	1,258,944	Redis	FALSE	FALSE	RESP, the Redis Serialization Protocol, used to communicate with Redis, a popular open source database and caching service
8080	9,243,677	http-alt0	TRUE	FALSE	A common alternative port for HTTP, usually used for web sites and web proxy services
8081	5,452,401	http-alt1	TRUE	FALSE	A common alternative port for HTTP, usually used for web sites and web proxy services
8443	4,515,905	https-alt	TRUE	TRUE	A common alternative port for HTTPS, usually used for test web sites
8888	1,855,002	http-alt8	TRUE	FALSE	A common alternative port for HTTP, usually used for web sites and web proxy services
9100	1,536,469	jetdirect	FALSE	FALSE	HP JetDirect, a printer control service used to manage print jobs
11211	39,799	Memcached (TCP)	FALSE	FALSE	Memcached, a distributed memory object caching system
11211	3,777	Memcached (UDP)	FALSE	FALSE	Memcached, a distributed memory object caching system
27017	561,471	Mongo	FALSE	FALSE	Mongo Wire Protocol, used to communicate with MongoDB, a popular open source document database
50000	1,066,201	DB2	FALSE	FALSE	IBM DB2 service, used to communicate with DB2 database servers
61439	3,244,998	canary-high	N/A	N/A	A high-number unassigned port that should not normally respond to SYN/ACK

APPENDIX B: TCP/IP TELEMETRY

While we believe that the National Exposure Index offers the most reliable view of "the internet" to date, there are a few factors that limit our telemetry capabilities.

Technical Considerations

First and foremost, we do not make any attempt to probe the growing IPv6 space. We are concerned totally with IPv4 space only. While the 4 billion-ish addresses that are possible with IPv4 might seem like a lot, IPv6 has an upper limit of about 340 undicillion (340 followed by 36 zeros)—a stupendously large number that is currently impossible to "scan" with any hope of finishing in our species' lifetime. We continue to investigate some technical shortcuts that will give us reasonable visibility in this space and hope to have some solid data in time for 2019.

Another area of "the internet" we cannot measure includes the networks and individual computers behind Network Address Translation (NAT) devices and firewalls. Unlike the Internet Census of 2012[58], Project Sonar and the National Exposure Index operate in a legal and ethical manner.

In fact, this brings up another class of network that we do not account for: the opted-in "blocklist" of networks that have requested that Sonar stop scanning, briefly alluded to above. At the time of our last scan, there were about 51 million IP addresses, or about 1% of the total possible routable addresses, that are consensually off-limits to our scanning[59].

Finally, we cannot scan purely client computers that are nonetheless connected directly to the internet. Any machine that offers no services cannot be "seen" by this study. Our study is strictly focused on the server-side of the internet.

Today, it is very normal for a closed UDP port to silently discard any unsolicited requests. Contrary to the UDP specification, many enterprises suppress their icmp-unreachable responses to unwanted UDP packets. This is a fine security practice, but it makes mapping UDP space a little bit more involved. For our Sonar studies, we always provide a normal and expected datagram, and listen for the subsequent reply. However, this does limit our UDP scanning capabilities; any protocol that requires a password, for example, is off-limits for Sonar, since providing authentication is too close to a United States Computer Fraud and Abuse Act (CFAA) violation.

[58] https://insights.sei.cmu.edu/cert/2013/10/working-with-the-internet-census-2012.html

[59] Hopefully, those opted-out organizations will see the value of this paper and reconsider their decisions.

Political Considerations

While our first National Exposure Index was concerned with countries ranked by GDP, we continue to find no meaningful correlation between GDP and a nation's exposure. Nevertheless, we are keenly interested in correlating virtual IP space to the virtual political space that is our planet's international landscape.

However, geography aficionados will be the first to tell you that the definition of a "country" or "nation" can sometimes be tricky, not to mention fraught with some deeply held political and cultural beliefs (especially by their residents!). For ease of reading, this paper refers to all political regions that are represented by a top-level IANA IPv4 registry as "countries" or "nations" interchangeably, irrespective of their official (or sometimes, disputed) political designations. However, some of these regions are not sovereign entities, such as the Hong Kong Special Administrative Region, and some island states are not represented in our data as independent IANA IPv4 registries, such as Greenland. Finally, if a region is not represented by a valid ISO 3166-1 alpha-3 codes in the country code R package for International Monetary Fund (IMF) country codes at the time of analysis, it will not be represented in this study. Therefore, unless otherwise noted, the data and visualizations presented in this paper are limited to only the 187 nations that represent nearly all of the identifiable internet services offered.

Top 50 Countries Ranked by Exposure

Table 10, below, lists the top 50 "most exposed" countries that also meet the above criteria.

Table 10: Top 50 IMF Countries, Ranked by Exposure

EXPOSURE RANK	ISO-3 CODE	COUNTRY	GDP (BILLIONS)	POPULATION (MILLIONS)	ALLOCATED IPV4S
1	USA	United States	$19,390.6	325.89	1,605,538,816
2	CHN	China, P.R.: Mainland	$12,014.6	1390.08	340,344,064
3	CAN	Canada	$1,652.4	36.66	70,245,632
4	KOR	Korea, Republic of	$1,538.0	51.45	112,449,024
5	GBR	United Kingdom	$2,624.5	66.05	125,988,760
6	FRA	France	$2,583.6	64.80	83,149,520
7	NLD	Netherlands	$825.7	17.08	49,262,304
8	JPN	Japan	$4,872.1	126.75	203,815,936
9	DEU	Germany	$3,684.8	82.71	120,768,552
10	MEX	Mexico	$1,149.2	123.52	28,879,360
11	AUS	Australia	$1,379.5	24.76	48,337,664
12	BRA	Brazil	$2,055.0	207.68	84,486,656
13	HKG	China, P.R.: Hong Kong	$341.7	7.41	11,920,896
14	RUS	Russian Federation	$1,527.5	143.99	45,286,528
15	POL	Poland	$524.9	37.97	20,981,320
16	ZAF	South Africa	$349.3	56.52	29,188,096
17	IND	India	$2,611.0	1316.90	41,681,664
18	TUR	Turkey	$849.5	80.81	16,545,280

EXPOSURE RANK	ISO-3 CODE	COUNTRY	GDP (BILLIONS)	POPULATION (MILLIONS)	ALLOCATED IPV4S
19	ISR	Israel	$350.6	8.71	7,728,384
20	IRN	Iran, Islamic Republic of	$431.9	81.42	12,781,056
21	ESP	Spain	$1,314.0	46.33	30,841,920
22	FIN	Finland	$253.2	5.50	13,588,928
23	ROU	Romania	$211.3	19.64	8,346,368
24	THA	Thailand	$455.4	69.10	9,064,704
25	SGP	Singapore	$323.9	5.61	12,228,096
26	IDN	Indonesia	$1,015.4	261.99	18,279,168
27	IRL	Ireland	$334.0	4.73	6,519,888
28	VNM	Vietnam	$220.4	93.64	15,927,552
29	UKR	Ukraine	$109.3	42.33	11,460,064
30	CHL	Chile	$277.0	18.38	10,244,864
31	SWE	Sweden	$538.6	10.12	30,370,408
32	CZE	Czech Republic	$213.2	10.58	9,344,384
33	BEL	Belgium	$494.7	11.35	28,527,488
34	DNK	Denmark	$324.5	5.75	12,454,760
35	COL	Colombia	$309.2	49.29	17,349,120
36	EGY	Egypt	$237.1	94.80	22,823,424
37	ARE	United Arab Emirates	$377.4	10.14	3,948,672
38	NOR	Norway	$396.5	5.29	15,997,328
39	PAK	Pakistan	$304.0	197.26	5,403,648
40	AUT	Austria	$416.8	8.82	11,668,320
41	MYS	Malaysia	$314.5	32.05	6,668,544
42	PRT	Portugal	$218.1	10.31	6,633,248
43	VEN	Venezuela, República Bolivariana de	$210.1	31.43	6,800,128
44	SLV	El Salvador	$28.0	6.37	656,896
45	NZL	New Zealand	$201.5	4.84	7,069,952
46	SAU	Saudi Arabia	$683.8	32.38	9,054,976
47	KAZ	Kazakhstan	$160.8	18.19	2,998,016
48	BRN	Brunei Darussalam	$12.7	0.43	208,384
49	BGD	Bangladesh	$261.4	163.19	1,536,512
50	PER	Peru	$215.2	31.83	3,184,896

The table below covers the rest of our country set, also sorted by exposure rank. However, it is important to understand that it is impossible to state which country is truly the "least" exposed—while #187 place happens to be occupied by the Federated States of Micronesia, it is not meaningfully more or less exposed than East Timor or Montserrat. Firstly, these country statistics have not been subjected to the more refined ranking algorithm described in Appendix D, and secondly, the differences in exposure among the lowest ranking countries are generally not statistically significant enough to warrant any real praise or derision.

Table 11: Less exposed countries

EXPOSURE RANK	ISO-3 CODE	COUNTRY	GDP (BILLIONS)	POPULATION (MILLIONS)	ALLOCATED IPV4S
51	ITA	Italy	$1,937.9	60.59	57,546,880
52	CHE	Switzerland	$678.6	8.42	20,665,032
53	MAC	China, P.R.: Macao	$49.8	0.64	334,080
54	BGR	Bulgaria	$56.9	7.06	4,435,968
55	HRV	Croatia	$54.5	4.15	2,172,928
56	HUN	Hungary	$152.3	9.81	5,908,992
57	PHL	Philippines	$313.4	105.31	5,553,664
58	ECU	Ecuador	$102.3	16.78	2,629,376
59	GRC	Greece	$200.7	10.77	5,596,416
60	LTU	Lithuania	$47.3	2.83	2,344,960
61	LVA	Latvia	$30.3	1.95	1,749,504
62	SVK	Slovak Republic	$95.9	5.43	2,670,592
63	EST	Estonia	$26.0	1.31	1,270,000
64	NGA	Nigeria	$376.3	188.69	2,511,616
65	SRB	Serbia, Republic of	$41.5	7.03	2,286,592
66	PAN	Panama	$61.8	4.10	1,838,336
67	CRI	Costa Rica	$58.1	4.97	2,589,952
68	SVN	Slovenia	$48.9	2.07	2,598,656
69	DOM	Dominican Republic	$75.0	10.17	1,582,592
70	GTM	Guatemala	$75.7	16.92	614,656
71	BOL	Bolivia	$37.1	11.07	1,148,672
72	KWT	Kuwait	$120.4	4.41	1,955,584
73	MDA	Moldova	$8.1	3.55	1,331,456
74	KEN	Kenya	$79.5	46.73	5,552,896
75	BLR	Belarus	$54.4	9.45	1,837,312
76	MAR	Morocco	$109.8	34.85	10,687,488
77	CYP	Cyprus	$21.3	0.85	1,102,144
78	LUX	Luxembourg	$62.4	0.59	1,433,600
79	SYC	Seychelles	$1.5	0.09	7,978,496
80	TUN	Tunisia	$40.3	11.52	6,538,240

EXPOSURE RANK	ISO-3 CODE	COUNTRY	GDP (BILLIONS)	POPULATION (MILLIONS)	ALLOCATED IPV4S
81	LKA	Sri Lanka	$87.6	21.44	545,024
82	COG	Congo, Republic of	$8.5	4.35	116,224
83	AZE	Azerbaijan, Republic of	$40.7	9.82	757,760
84	JOR	Jordan	$40.5	7.13	684,928
85	DZA	Algeria	$178.3	41.54	4,791,040
86	MNG	Mongolia	$11.1	3.06	233,472
87	GEO	Georgia	$15.1	3.69	1,213,440
88	OMN	Oman	$74.3	4.13	927,744
89	QAT	Qatar	$166.3	2.74	835,840
90	BIH	Bosnia and Herzegovina	$18.1	3.51	802,560
91	NPL	Nepal	$24.5	29.34	531,456
92	URY	Uruguay	$58.4	3.49	2,442,240
93	PRY	Paraguay	$29.6	6.95	1,091,584
94	HND	Honduras	$23.0	8.31	528,896
95	KHM	Cambodia	$22.3	16.01	344,576
96	ARM	Armenia, Republic of	$11.5	2.99	618,528
97	IRQ	Iraq	$197.7	38.86	663,040
98	MUS	Mauritius	$12.4	1.27	2,079,744
99	MKD	Macedonia, FYR	$11.4	2.08	700,160
100	TZA	Tanzania	$51.7	50.05	1,047,552
101	ISL	Iceland	$23.9	0.34	884,224
102	LBN	Lebanon	$51.5	4.51	594,688
103	UGA	Uganda	$26.3	37.67	880,384
104	PSE	West Bank and Gaza	$10.0	4.55	685,824
105	GHA	Ghana	$47.0	28.28	2,268,160
106	ALB	Albania	$13.2	2.88	369,664
107	MDV	Maldives	$4.5	0.36	70,912
108	AFG	Afghanistan, Islamic Republic of	$20.9	35.53	158,720
109	BLZ	Belize	$1.9	0.39	167,424
110	BHR	Bahrain, Kingdom of	$34.9	1.45	449,536
111	NIC	Nicaragua	$13.7	6.22	405,760
112	MLT	Malta	$12.5	0.46	627,200
113	VUT	Vanuatu	$0.9	0.28	17,152
114	AGO	Angola	$124.2	28.18	1,209,088
115	GAB	Gabon	$15.2	1.91	421,888

EXPOSURE RANK	ISO-3 CODE	COUNTRY	GDP (BILLIONS)	POPULATION (MILLIONS)	ALLOCATED IPV4S
116	LAO	Lao People's Democratic Republic	$17.0	6.68	75,264
117	CIV	Cote d'Ivoire	$40.4	24.96	1,165,824
118	KGZ	Kyrgyz Republic	$7.2	6.26	277,504
119	TTO	Trinidad and Tobago	$21.6	1.37	542,720
120	MOZ	Mozambique	$12.7	29.54	440,064
121	YEM	Yemen, Republic of	$16.5	29.98	135,168
122	ZWE	Zimbabwe	$17.5	14.88	102,144
123	MNE	Montenegro	$4.8	0.62	227,328
124	NAM	Namibia	$12.7	2.34	459,008
125	BHS	Bahamas, The	$11.6	0.37	135,680
126	JAM	Jamaica	$14.4	2.84	215,040
127	CMR	Cameroon	$34.0	24.28	713,984
128	NCL	French Territories: New Caledonia	$9.9	0.27	159,232
129	TJK	Tajikistan	$7.3	8.84	70,144
130	SDN	Sudan	$58.2	40.78	1,357,056
131	COD	Congo, Democratic Republic of	$41.4	86.65	146,944
132	CUW	Curacao	$3.1	0.16	190,208
133	ZMB	Zambia	$25.5	17.24	1,612,544
134	ETH	Ethiopia	$80.9	92.66	361,472
135	MDG	Madagascar	$11.5	25.61	167,680
136	BWA	Botswana	$17.2	2.18	148,736
137	GLP	Guadeloupe	$8.0	0.40	17,152
138	SEN	Senegal	$16.5	15.86	400,384
139	BRB	Barbados	$5.0	0.28	172,800
140	SUR	Suriname	$3.3	0.58	79,872
141	MMR	Myanmar	$66.5	52.65	141,312
142	HTI	Haiti	$8.6	10.98	162,816
143	FJI	Fiji	$5.1	0.89	143,616
144	MWI	Malawi	$6.2	19.17	417,536
145	RWA	Rwanda	$9.1	11.84	348,672
146	TGO	Togo	$4.8	7.80	317,952
147	ABW	Aruba	$2.5	0.10	87,040
148	MTQ	Martinique	$9.6	0.39	13,312
149	BEN	Benin	$9.2	11.13	162,304
150	CYM	Cayman Islands	$2.5	0.06	173,568
151	PNG	Papua New Guinea	$23.6	8.25	62,720

EXPOSURE RANK	ISO-3 CODE	COUNTRY	GDP (BILLIONS)	POPULATION (MILLIONS)	ALLOCATED IPV4S
152	MLI	Mali	$15.3	18.89	85,504
153	ATG	Antigua and Barbuda	$1.5	0.09	64,512
154	GUY	Guyana	$3.6	0.77	68,608
155	BFA	Burkina Faso	$12.6	18.94	293,888
156	LSO	Lesotho	$2.8	1.94	119,808
157	NER	Niger	$8.3	18.76	39,936
158	GMB	Gambia, The	$1.0	2.10	258,560
159	LCA	St. Lucia	$1.7	0.18	18,432
160	BTN	Bhutan	$2.3	0.80	30,976
161	KNA	St. Kitts and Nevis	$0.9	0.06	14,336
162	SWZ	Swaziland	$4.5	1.15	47,360
163	WSM	Samoa	$0.8	0.20	17,920
164	BDI	Burundi	$3.4	10.87	34,816
165	GRD	Grenada	$1.1	0.11	9,728
166	MRT	Mauritania	$5.1	3.88	42,496
167	CPV	Cabo Verde	$1.7	0.54	28,672
168	GIN	Guinea	$9.7	12.97	32,512
169	GNQ	Equatorial Guinea	$10.7	0.84	18,432
170	SSD	South Sudan	$2.9	12.59	14,336
171	DMA	Dominica	$0.6	0.07	11,520
172	DJI	Djibouti	$2.0	1.02	49,664
173	LBR	Liberia	$3.3	4.51	97,792
174	TCD	Chad	$9.9	12.19	20,992
175	SLB	Solomon Islands	$1.3	0.61	11,520
176	TON	Tonga	$0.4	0.11	9,728
177	SLE	Sierra Leone	$3.6	7.41	89,088
178	AIA	Anguilla	$0.2	0.02	8,448
179	VCT	St. Vincent and the Grenadines	$0.8	0.11	8,704
180	SMR	San Marino	$1.6	0.04	34,560
181	KIR	Kiribati	$0.2	0.12	4,608
182	CAF	Central African Republic	$1.9	4.98	7,424
183	PLW	Palau	$0.3	0.02	5,632
184	GNB	Guinea-Bissau	$1.4	1.70	6,144
185	MSR	Montserrat	$0.0	0.01	1,280
186	TLS	Timor-Leste, Dem. Rep. of	$2.6	1.24	15,872
187	FSM	Micronesia, Federated States of	$0.3	0.10	8,192

APPENDIX C: METHODOLOGY

Choosing Ports

We continue to increase our internet measurement capabilities, and—while we strived for study parity between 2016 and 2017—we chose to fully utilize these enhancements to Project Sonar to provide the most robust study data to-date.

First, we have altered the port makeup of our studies for 2018 and have included:

- TCP port 81 (another common HTTP alternative port)
- TCP port 11211 (Memcached)
- TCP port 27017 (MongoDB)
- TCP port 50000 (DB2)
- TCP port 6379 (Redis
- TCP port 5432 (PostgreSQL)

We have also introduced protocol-level scans for SMB (TCP port 445), SSH (TCP port 22), Microsoft SQL Server (TCP port 1433) for both enhanced analysis and use in the final rankings.

Furthermore, the TCP "Canary" ports of 5 and 61439 were chosen to represent low-frequency IANA-recognized and unofficial services (as indicated by the latest nmap-services[60] list).

Unless there is a significant change in future services offered on the internet, the authors of this paper are quite confident this will be the last major change to the National Exposure Index's port profile. We are likely to perform more protocol-level scans as we develop techniques that ensure the safety and accuracy of said scans.

Surveying The Internet

Appendix A discussed the technical underpinnings of Project Sonar. Our scanning blacklist grew to just under 51 million restricted IPv4s, up from just over 50 million in 2017—a roughly 2% increase.

We have also used the results from our protocol-level scans in conjunction with the results of the "canary port" scans to refine the portions of public IPv4 space we incorporated into the study. The TCP SYN scans are still the primary source of the internet survey results as they are the least intrusive type of probes. However, that does not mean that a device responding on, say, TCP port 445 is truly a host running one or more Microsoft protocols. The response itself is significant (i.e. something is listening and responding), but in this period of growing numbers of massive content delivery networks, DDoS mitigation services, cloud providers and highly complex routing and forwarding configurations, we continue to be determined to ensure the results are as representative as possible without becoming more intrusive.

We excluded all IPv4 addresses that responded on both canary ports and used the comparisons between protocol scans and canary results to alter the ranking algorithm (see "Ranking Exposure By Country" below).

[60] https://nmap.org/book/nmap-services.html

Geolocating Countries

The commercial version of MaxMind's geolocation databases was used to match each IPv4 address to a country. In both 2016 and 2017 we noted a larger percentage of geolocation errors in the results than were claimed on MaxMind's site. Some of the most egregious errant results were IPv4s attributed to Antarctica when those nodes were both being routed by infrastructure in other countries and containing content/services that were not based in Antarctica. This year, the geolocation accuracy improved greatly and on parity with CAIDA's observations in their most recent comparative study in 2011[61].

Ranking Exposure By Country

The National Exposure Index was created by aggregating the results of the individual rankings of the following exposed, usually cleartext ports:

- HTTP (80, 81, 8000, 8080, and 8888)
- SMTP
- SIP (UDP)
- FTP
- POP3
- IMAP
- Database ports (MySQL, PostgreSQL, Oracle DB, MSSQL, Redis, DB2, and MongoDB)
- Telnet
- PPTP
- RDP
- Rpcbind
- MS-RPC
- uPNP
- SMB
- NBSS
- HP JetDirect
- SSDP (UDP)
- RFB
- LDAP (TCP and UDP)
- NBSN (UDP)
- mDNS (UDP)
- Memcached (TCP and UDP)
- Chargen (UDP)

We chose these services from the 37 TCP ports and 9 UDP ports covered in the full study scans as there is either a greater likelihood of exposure of sensitive information over cleartext channels with them, or they expose services that have been identified with extensive vulnerabilities over time.

[61] https://www.caida.org/publications/papers/2011/geocompare-tr/geocompare-tr.pdf

As noted throughout the paper, there were three changes to the ranking algorithm for the 2018 report. The first change was the addition of more database ports to the scanning set, and weighting those results more severely due to the enhanced risk posed by exposing those services on the internet. The second change was including a number of UDP-based services that were identified not only as vectors for direct attack, but also for their utility as DDoS amplifiers. Finally, we have treated nodes that respond to both "Canary" ports much differently as noted above. Those nodes do count against the total utilized IP address space, but the apparent protocols they respond to are discarded. We also used the "Canary" measurements along with the protocol-level results to reduce the impact on the weights used in the final ranking algorithm where protocol-based measurements were known.

We generate individual ranked lists on a per-port results basis. If the total number of nodes found for a given port is greater than 10% of the total IPv4s found, then the individual port rankings are based on:

percentage_of_found_nodes_in_that_country X

log2(count_of_found_nodes) X

log2(sonar_found_country_ipv4_total)

For UDP and non-protocol-measured comparisons, the individual rankings are based on:

log2(count_of_found_nodes) X log2(sonar_found_country_ipv4_total)

We adjusted the ranking algorithm this time after seeing the aftermath of exposure in both 2016 and 2017 both in public disclosures and our passive view of the internet with Project Heisenberg. That is, device counts do matter, and the likelihood of node exposure dramatically rises with the volume of nodes within a country. Even a casual review at both the individual rankings results and the final rankings "make more sense" to cybersecurity practitioners.

We used the same weighted, seeded (using the same seed) Cross Entropy Monte Carlo (CEMC) algorithm to generate the index of the 50 countries having the most exposure. Ranking challenges, such as this one, fall into the category of a combinatorial optimization problem, and the CEMC approach provides a stochastic computational means to iterate over each ranked list, perform importance sampling, and derive a final outcome. This year's results further support our belief that the nature of these ranked lists makes CEMC a preferred methodology over others. We used this technique to perform a comprehensive ranking across all represented countries versus the top 50 in previous years.

R[62], RStudio[63], Apache Drill[64], Amazon Athena[65] and an enhanced version of the Measurement Factory's ipv4-heatmap[66] tool were used for all data processing, analysis, and visualizations. Full code, package/tool references, and further details on the analyses will be released on Rapid7 Labs' GitHub repository for the report: https://github.com/rapid7/data/. All relevant study data will be released on Rapid7's Open Data[67] portal.

[62] https://www.r-project.org/

[63] https://rstudio.com/

[64] https://drill.apache.org/

[65] https://aws.amazon.com/athena/

[66] https://github.com/hrbrmstr/ipv4-heatmap

[67] https://opendata.rapid7.com/

ABOUT RAPID7

Rapid7 powers the practice of SecOps by delivering shared visibility, analytics, and automation that unites security, IT, and DevOps teams. The Rapid7 Insight platform empowers these teams to jointly manage and reduce risk, detect and contain attackers, and analyze and optimize operations. Rapid7 technology, services, and research drive vulnerability management, application security, incident detection and response, and log management for organizations around the globe. To learn more about Rapid7 or get involved in our threat research, visit www.rapid7.com.

QUESTIONS

Reach us at research@rapid7.com