NATIONAL EXPOSURE INDEX

2017

Rapid7 Labs | June 14, 2017

Bob Rudis, Chief Security Data Scientist, Rapid7
Tod Beardsley, Research Director, Rapid7
Jon Hart, Senior Security Researcher, Rapid7
Tom Sellers, Lead Security Researcher, Rapid7

CONTENTS

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?

This year, we continue this investigation into the risk of passive eavesdropping and active attack on the internet, and offer insight into the year-over-year changes involving these exposed services. We've adjusted our ranking algorithm to more accurately reflect overall exposure on a country-by-country basis, enhanced our scanning and geolocation methodologies, and provided evidence that some areas of exposure are, in fact, improving.

Key Findings

- The most exposed regions are Zimbabwe, Hong Kong SAR[1], Samoa, Republic of the Congo, Tajikistan, Romania, Ireland, Lithuania, Australia, and Estonia. No discussion of national exposure would be complete without reference to the three major cyber superpowers: the United States, China, and the Russian Federation. While both the Russian Federation and China are among the top 50 most exposed nations, the U.S. has relatively low exposure in relation to its enormous IPv4 address space.
- The number of IP-addressable, public internet servers offering exposed services in Belgium—the country that took the top spot in the 2016 index—dropped by 250,000 over the year. Belgium no longer ranks in the top 50 most exposed regions at all, due largely to this culling of exposed services.
- Over 1 million endpoints were confirmed exposing Microsoft file-sharing services (SMB, TCP port 445), with 800,000 of them being confirmed[2] Windows systems, spanning virtually the entire product and release version lineage of the company. This made for a target-rich environment for WannaCry, a "ransomworm" that spreads in part through an SMB exploit made public in May of 2017. This vulnerability was also actively sought out in May 2017, with SMB port scan results increasing by 17% (4.7 million to 5.5 million nodes). Blocking port 445 would mitigate potential threats like this.
- Port scanning for telnet (port 23) in 2017 returned just under 10 million responsive nodes, compared to 2016's scan results of over 14.8 million. This 33% drop in apparent telnet services can almost certainly be pinned to two developments: 1) ISP actions, such as closing port 23 in response to the Mirai botnet, and 2) Mirai, BrickerBot, and other botnets knocking nodes offline.

These key findings illustrate two overall themes we explore in this paper. First, the phenomenon of widespread internet exposure makes for an environment attractive to criminals and other malicious actors, as well as increasing the likelihood of accidental breaches. Second, as was the case in Belgium, national technical leadership can absolutely take steps to reduce their regional internet exposure, thereby strengthening their networks and protecting users. National internet service providers around the world, as well as the policy makers and legislators responsible for national security and commerce, can use these findings to reduce their own regional exposure through informed decision-making based on real-world data.

[1] Hong Kong SAR isn't a sovereign nation, but is a member of the IMF. The selection criteria is more fully explained in Appendix C: Selected Countries.

[2] Using Rapid7's open source "Recog" tool — https://github.com/rapid7/recog

REVISITING NATIONAL EXPOSURE

When Rapid7 kicked off this project in the spring of 2016, we had two fairly simple hypotheses we wanted to test: does a country's gross domestic product (GDP) have an appreciable effect on the scale of its adoption of internet technologies (we assumed yes); and do these high GDP, technologically advanced countries have correspondingly high levels of unsafe services that pose major privacy and security risks, such as unencrypted protocols that expose data in cleartext (again, we guessed this was so). Neither of these are particularly novel questions, and most security professionals would believe this to be the case, but what surprised us were two findings in the 2016 National Exposure Index[3].

First, it didn't seem like any other reputable research organization was actively and broadly measuring and publishing some basic metrics about exposure on "the internet," which makes it impossible to answer these fairly fundamental questions. While individual researchers and research organizations do "scan the internet," these activities tend to be ad-hoc, focused on specific services, and rarely are released with full datasets for peer review[4].

Second, our 2016 study found that while national GDPs definitely correlate with fixed-IP internet adoption, we found no correlation between internet exposure and a country's GDP, rate of adoption, or anything else. Many high GDP nations, such as China, France, and the United States, had relatively high rates of exposure, but so do smaller nations, such as Tajikistan, Gabon, and Mozambique. So, while our second hypothesis turned out to be incorrect, the finding of a lack of correlation between internet adoption and exposure level is itself interesting.

Why Keep Looking?

Now in its second year, the National Exposure Index (and the accompanying datasets) continues to serve as an important study into the nature of the internet. We now have access to year-over-year data which can help us develop theories about the trends at work in our shared online world.

Specifically, internet-wide measurements of exposure can help information technology professionals develop and deliver strategies and solutions for their own organizations to measure and manage their exposure and risk. After all, without a sense of what's normal for a given segment of the internet, it can be difficult to make reasoned decisions on how to improve the security posture of a specific organization. We also hope this research will help national policy makers, legislators, and regulators around the world to understand their own baseline of exposure and to consider sensible policy decisions that can promote both internet freedom and internet resiliency.

[3] https://information.rapid7.com/national-exposure-index.html

[4] You can peruse our data at https://github.com/rapid7/data/tree/master/national-exposure

On the methodology side of things, we've updated our regional ranking system algorithm. Last year we based rankings on the exposure percentages of cleartext services across—and as a percentage of—overall exposed IPv4 space of a given country. This year, we based the exposure percentages on overall allocated IPv4 space of a given country in order to better reflect actual total resource exposure. This year we were also able to include more than just potential SYN-scan results exposure, utilizing measured levels of direct exposure to WannaCry with a full protocol exchange. Since we've concluded that GDP does not appear to have a causal effect on exposure, we no longer account for GDP at all.

By breaking out statistics by country and comparing year-over-year standings in the National Exposure Index, we have the opportunity to identify which nations have improved 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, and recent events suggest that foreign nation-state actors are keenly interested in taking advantage of national internet exposure to pursue their own interests.

IMPROVED GEOLOCATION

One of the challenges with internet-wide scans like this is the imperfection of IP geolocation. While blocks of IP addresses are allocated to specific entities by the Internet Assigned Numbers Authority (IANA)[5], those entities are often multinational in nature. Even in cases where IP address ranges are assigned to a specific government, that government often operates internationally. To help with our fidelity, we also rely on the GeoIP2 service from MaxMind[6], a leading provider in IP geolocation services.

However, no geolocation database is perfect, and there are edge cases where smaller regions are sometimes misidentified. For instance, preliminary results this year showed hosts in Antarctica. While it is possible the local penguin population evolved sufficiently to use the operating system that adopted them as its mascot (Linux), upon further, meticulous investigation we discovered that virtually all the addresses geolocated to Antarctica ended up being in Romania. Despite this disappointment, it was a valuable learning experience as to just how difficult it is to geolocate properly at even the continent level[7].

[5] http://www.iana.org/assignments/ipv4-address-space/ipv4-address-space.xhtml

[6] https://www.maxmind.com/en/home

[7] These findings, of course, have some serious implications for incident attribution when basing that attribution on the source IP addresses of attackers, but that's a different topic for a different paper.

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[8]. A client computer, such as a desktop or a smartphone, then makes a TCP/IP connection to that service, and the magic of packet exchanges occurs.

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 30 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 [9].

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 1, to the right, is just about the most accurate map you will find today of "the internet."

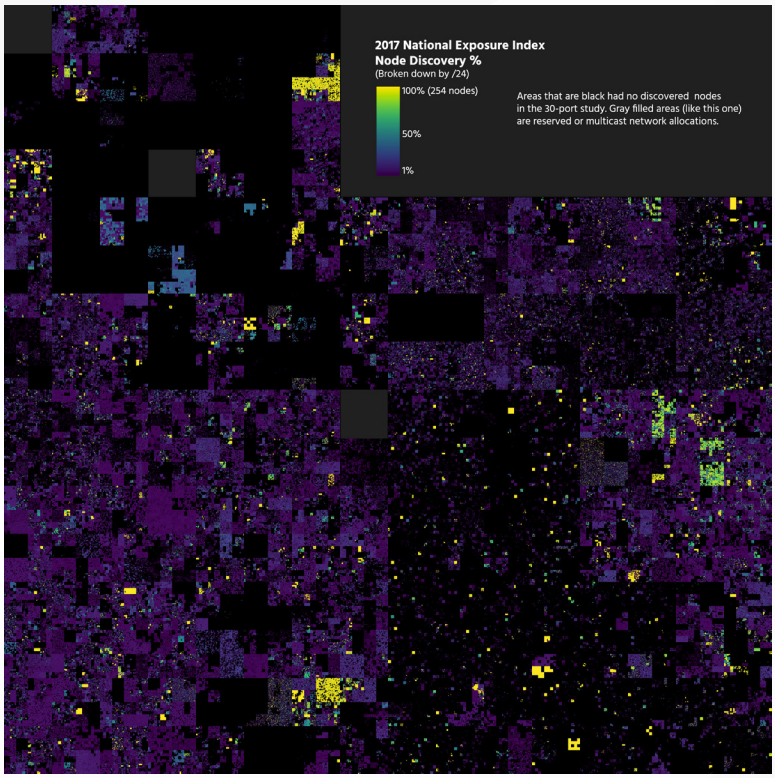


Figure 1: Map of the internet

On this map, where 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 responsive ports, with lighter coloring representing a higher density of services.

[8] 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).

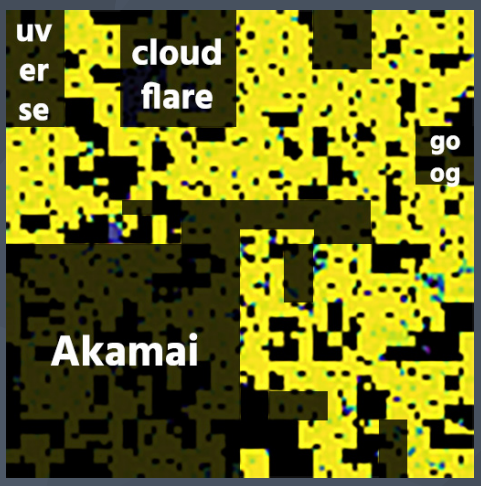
[9] https://sonar.labs.rapid7.com/

SIDEBAR: ISLANDS OF LIGHT

To take one example, the large, bright yellow region near the bottom is the 104.0.0.0/8 netblock, managed by ARIN containing 16,777,216 IPv4 addresses. Nearly half of those (7,177,760 to be exact) are "owned" by the following recognizable cloud providers, content delivery networks (CDNs), internet service providers (ISPs), and corporations:

Figure 2: Providers in the densely populated 104/8 netblock

CLOUD/CDN/ISP/COMPANY	COUNT
Akamai	4,194,304
Cloudflare	1,048,576
Azure	699,424
AT&T UVerse	524,288
Google	393,216
Digital Ocean	131,072
Rackspace	102,656
Microsoft	40,960
SoftLayer	27,904
Linode	15,360



Web-oriented services are the most prolific in this address space, which is not surprising given the presence of CloudFlare and Akamai. All 30 ports in the study managed to appear in some portion of that 104.0.0.0/8, and 57 distinct organizations responded to SYN packets on all 30 ports:

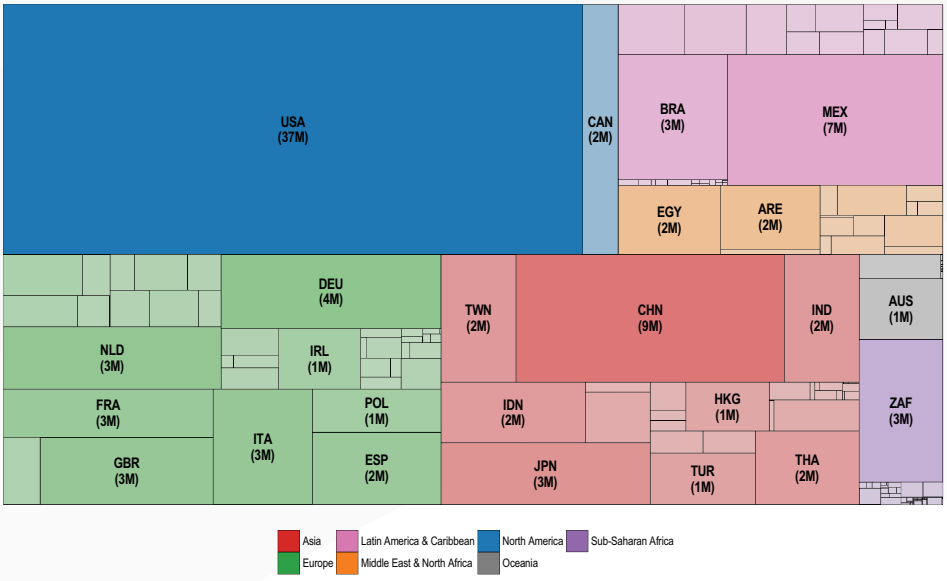
123.Net, Activision Publishing, Axia Connect Limited, Black Oak Computers Inc - Miami, C Spire Fiber, California Internet, L.P., Choopa, LLC, ClearDDoS Technologies, Contina, CorKat Data Solutions, LLC, Digital Ocean, Envision Healthcare Holdings, Enzu, Eonix Corporation, Frontier Communications, Global Frag Networks, Google Cloud, GorillaServers, HighWire Press, HostHatch, IT7 Networks, Kix Media, Lightower Fiber Networks I, LLC, Linode, Merchant-Link, LLC, Microsoft Azure, Netago, Psychz Networks, QuadraNet, QuickPacket Atlanta, LLC, Rackspace Hosting, Reliablehosting.com, Sharktech, SkyWire Fiber LLC, Telepacific Communications, Time Warner Cable, US Electrodynamics, Versaweb, LLC, Virtuzo, VolumeDrive, Wave Broadband, Access One, AT&T Internet Services, ColoCrossing, FranTech Solutions, Hostwinds LLC., Incero LLC, Jaguar Communications, JUCE Communications, KW Datacenter, MTS Allstream, Multacom Corporation, Psychz Networks Dallas, Rackspace Cloud Servers, Roya Hosting LLC, VegasNAP, LLC, Velcom

Many of those organizations—such as ClearDDoS, Psychz Networks, Global Frag Networks, and CloudFlare—provide denial-of-service mitigation services, so it makes some sense for them to interact with all SYNs. Others are hosting providers and some are smaller ISPs that may either have infrastructure configuration issues or have customers who are providing some "interesting" services.

Sorting simply by IP address does not give us the view we're after, though; we want to see the utilization of IP address space by country. For that, we've generated a view of the internet, colored by region and split by country code, in Figure 3.

Figure 3: National Exposure 2017 - Country Level distribution of active IPv4s

Labels rounded to nearest million



It's not surprising to see the United States and China in the lead as they are leaders in providing internet services to the globe. Only 27 countries expose services from more than one million IPv4s. Remember, these are nodes that responded to our SYN scans on one or more ports. Mexico responds to more probes than Japan (7 million versus 3 million), and we expected booming economies such as India to have more exposed nodes than 2 million, especially given their progressive technology sector policies.

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, or (b) offering a service on the internet that is unsuitable for public access. If either (or both) 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 both 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 twentieth century, dominated by terminals physically wired to mainframes, because 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[10]. Even if an adversary controls one of the routers between you and **yourbank.com**, you have assurances built into 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 national security organizations in some countries are capable of conducting large scale, passive monitoring of internet activity, and that the Internet Engineering Task Force proposed in 2014 that "Pervasive Monitoring Is an Attack" in RFC7258, an official memorandum with that title[11].

While we acknowledge that 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.

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

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

While we acknowledge that 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[12]. 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 amateur cyber criminals camped out on a local Wi-Fi access point.

Scanned Cleartext and Encrypted Services

For the purpose of this investigation, the cleartext services listed below were chosen as scan targets. A complete list of scan targets can be found in Appendix A.

Figure 4: Scanned cleartext services

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
80	73,637,628	HTTP	HyperText Transfer Protocol, used to serve web pages and web applications
25	18,297,775	SMTP	Simple Mail Transfer Protocol, used to send email
21	16,980,464	FTP	File Transfer Protocol, used to send and receive data and text files; FTPS, SSH, and HTTPS are all encrypted alternatives
8080	13,428,979	http-alt0	A common alternative port for HTTP, usually used for websites and web proxy services
53	11,829,288	DNS	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
23	9,995,851	telnet	Telnet, a remote command shell interface, one of the oldest protocols on the internet; SSH is an encrypted alternative
143	8,919,856	IMAP	Internet Message Access Protocol, used to receive email
110	8,820,647	POP3	Post Office Protocol version 3, used to receive email
3306	8,279,501	MySQL	MySQL, used to communicate with the (usually) open source MySQL Server published by Oracle
3389	7,279,527	RDP	Remote Desktop Protocol, a graphical user interface to remotely administer (usually) Microsoft Windows servers and desktops
8081	7,170,947	http-alt1	A common alternative port for HTTP, usually used for websites and web proxy services
587	7,056,310	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
111	5,953,599	rpcbind	Remote Procedure Call port mapping service, usually used on Unix-like operating systems, usually for NFS file sharing
445	5,547,284	SMB	Server Message Block, a file transfer and remote administration protocol for (usually) Microsoft operating systems
135	5,413,613	MS-RPC	Microsoft Remote Procedure Call, usually used on Microsoft operating systems for distributed computing

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

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
5000	5,017,418	uPNP	Universal Plug-and-Play, a protocol for machine-to-machine discovery and configuration
139	4,027,291	NBSS	NetBIOS Session Service, used in NetBIOS over TCP/IP, usually on Microsoft operating systems for file and print sharing
5900	3,543,818	RFB	Remote Frame Buffer, a remote GUI for desktop administration, usually used by VNC (Virtual Network Computing)
1433	3,402,532	MSSQL	Microsoft SQL Server service, used to communicate with Microsoft database servers of the same name
389	2,990,559	LDAP	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
9100	2,951,880	jetdirect	HP JetDirect, a printer control service used to manage print jobs

While many of these services do offer **opportunistic encryption**[13], 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[14].

Figure 5: Scanned encrypted services

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
443	49,762,185	HTTPS	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	20,213,618	SSH	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
1723	9,607,708	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers.
993	7,027,891	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	6,983,958	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
465	6,603,840	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
8443	5,429,002	https-alt	A common alternative port for HTTPS, usually used for test websites
990	3,235,369	FTPS	Secure FTP, an encrypted-by-default alternative to FTP

[13] 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.

[14] 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 "reasonable," and more modern VPN solutions with more robust encryption standards should be used in place of PPTP.

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 and encrypted, unrestricted open access to the service creates exposure.

For example, we surveyed the internet for the database service ports associated with Microsoft SQL Server (port 1433) and MySQL (port 3306). Both of these database systems offer perfectly adequate authentication protocols and encryption guarantees, but the services 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 SQL statements provided by amateur DBAs would cause. Databases should always be mediated by a simplified, restricted front end, like a web application[15].

In addition to their sensitivity to denial-of-service conditions, 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, and it has been an attractive target for attackers for decades precisely due to the likelihood of vulnerabilities in its complex implementation[16]. Exposing an SMB service on the internet is simply asking for trouble.

Machines that expose these inappropriately open services are, therefore, necessarily exposed to increased risk more than machines that offer only appropriately curated, internet-ready services. This increase of attack surface is irrespective of any controls that would limit access, access attempts, or allow usage of these services.

[15] 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 SQL server, then surely a population of open SQL servers should be even more troubling.

[16] For example, EternalBlue, an SMB exploit, was disclosed in April of 2017 as part of the ShadowBrokers dump, and subsequently picked up by the WannaCry ransomware worm in May, 2017. SMB vulnerabilities are alive and well.

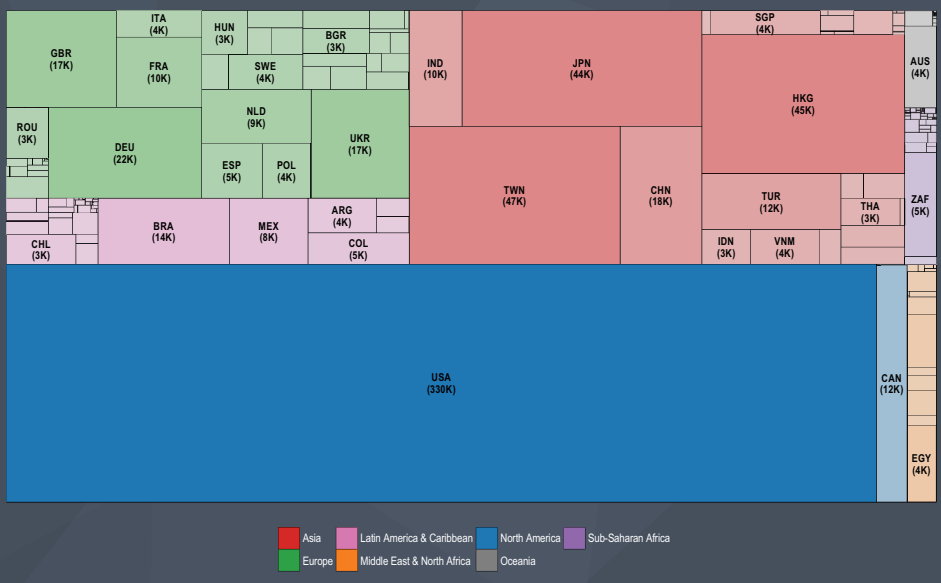
SIDEBAR: SO EXPOSED YOU'LL #WANNACRY

The production of this report coincided with the ransomworm outbreak dubbed "WannaCrypt"/"WannaCry." While our SYN scans picked up over 5 million nodes responding to port 445 SMB SYN scans, we went a bit further and configured Project Sonar to try to fully negotiate SMB and pull back information on server type, operating system, and version. These nodes did not prevent us from discovering their configuration (we even captured their Windows domain name).

Over 800,000 of the 5+ million nodes were verified to be running Windows. Figure 6 provides an overview of the distribution of nodes by country:

Figure 6: Country-level distribution of Windows nodes exposing SMB (port 445)

Hosts were classified using Rapid7's 'Recog' utility - https://github.com/



Source: Rapid7 Project Sonar

These results were prolific enough to cause us to augment our rating algorithm by adding the results of this deeper, full protocol scan to it with a higher rating.

The emergence of this ransomworm further emphasizes the need for individuals and organizations to be far more mindful of what they're exposing to the internet. There is no reason to fully expose SMB without firewalling or restricting access in some other way.

[17] https://community.rapid7.com/community/infosec/blog/2017/05/12/wanna-decryptor-wncry-ransomware-explained

Scanned Inappropriate and Appropriate Services

The inappropriate services listed below were chosen as scan targets. A complete list of scan targets can be found in Appendix A.

Figure 7: Scanned inappropriate services

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
21	16,980,464	FTP	File Transfer Protocol, used to send and receive data and text files; FTPS, SSH, and HTTPS are all encrypted alternatives
23	9,995,851	telnet	Telnet, a remote command shell interface, one of the oldest protocols on the internet; SSH is an encrypted alternative
1723	9,607,708	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers.
3306	8,279,501	MySQL	MySQL, used to communicate with the (usually) open source MySQL Server published by Oracle
3389	7,279,527	RDP	Remote Desktop Protocol, a graphical user interface to remotely administer (usually) Microsoft Windows servers and desktops
111	5,953,599	rpcbind	Remote Procedure Call port mapping service, usually used on Unix-like operating systems, usually for NFS file sharing
445	5,547,284	SMB	Server Message Block, a file transfer and remote administration protocol for (usually) Microsoft operating systems
135	5,413,613	MS-RPC	Microsoft Remote Procedure Call, usually used on Microsoft operating systems for distributed computing
5000	5,017,418	uPNP	Universal Plug-and-Play, a protocol for machine-to-machine discovery and configuration
139	4,027,291	NBSS	NetBIOS Session Service, used in NetBIOS over TCP/IP, usually on Microsoft operating systems for file and print sharing
5900	3,543,818	RFB	Remote Frame Buffer, a remote GUI for desktop administration, usually used by VNC (Virtual Network Computing)
1433	3,402,532	MSSQL	Microsoft SQL Server service, used to communicate with Microsoft database servers of the same name
389	2,990,559	LDAP	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
9100	2,951,880	jetdirect	HP JetDirect, a printer control service used to manage print jobs

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

Figure 8: Scanned internet-appropriate services

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
80	73,637,628	HTTP	HyperText Transfer Protocol, used to serve web pages and web applications
443	49,762,185	HTTPS	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	20,213,618	SSH	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
25	18,297,775	SMTP	Simple Mail Transfer Protocol, used to send email
8080	13,428,979	http-alt0	A common alternative port for HTTP, usually used for websites and web proxy services
53	11,829,288	DNS	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
1723	9,607,708	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers
143	8,919,856	IMAP	Internet Message Access Protocol, used to receive email
110	8,820,647	POP3	Post Office Protocol version 3, used to receive email
8081	7,170,947	http-alt1	A common alternative port for HTTP, usually used for websites and web proxy services
587	7,056,310	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
993	7,027,891	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	6,983,958	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
465	6,603,840	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
8443	5,429,002	https-alt	A common alternative port for HTTPS, usually used for test websites
990	3,235,369	FTPS	Secure FTP, an encrypted-by-default alternative to FTP
8888	2,545,753	http-alt8	A common alternative port for HTTP, usually used for websites and web proxy services

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 Port Scanning Results 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, or one that is made up primarily by those protocols that are already both appropriate for the internet and natively encrypted, such as the following.

Figure 9: Scanned services which are either encrypted or otherwise appropriate for internet use.

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	DESCRIPTION
80	73,637,628	HTTP	HyperText Transfer Protocol, used to serve web pages and web applications
443	49,762,185	HTTPS	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	20,213,618	SSH	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
25	18,297,775	SMTP	Simple Mail Transfer Protocol, used to send email
8080	13,428,979	http-alt0	A common alternative port for HTTP, usually used for websites and web proxy services
53	11,829,288	DNS	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
1723	9,607,708	PPTP	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers
143	8,919,856	IMAP	Internet Message Access Protocol, used to receive email
110	8,820,647	POP3	Post Office Protocol version 3, used to receive email
8081	7,170,947	http-alt1	A common alternative port for HTTP, usually used for websites and web proxy services
587	7,056,310	SMTP submission	SMTP submission service, usually used by endpoint mail clients to send email
993	7,027,891	IMAPS	Secure IMAP, an encrypted-by-default alternative to IMAP
995	6,983,958	POP3S	Secure POP3, an encrypted-by-default alternative to POP3
465	6,603,840	SMTPS	Secure SMTP, an encrypted-by-default alternative to SMTP
8443	5,429,002	https-alt	A common alternative port for HTTPS, usually used for test websites
990	3,235,369	FTPS	Secure FTP, an encrypted-by-default alternative to FTP
8888	2,545,753	http-alt8	A common alternative port for HTTP, usually used for websites and web proxy services

Canary Ports

Along with the 30 service ports 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[18], since there is no well-known service associated with them. Yet, we picked up responses from approximately 3.2 million devices from these two ports. 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. Usually, this is a misconfiguration, and a relatively harmless one at that. We also looked at which IP addresses responded on both canary ports. Over 2.3 million responded to both probes, and they're spread out across the vast internet landscape as seen in figure 10 on the following page.

[18] 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.

Figure 10: Canary Ports

Nodes in only 133 of the 183 countries in the study responded to the dual canary port probes. The list of countries not on that list is a bit more interesting than those on it because they include larger, more modern countries such as Iran, South Korea, and the Russian Federation. Portions of the responding address space may be taken up by DDoS prevention companies[19], but there are far too many individual canary ports to track with any level of accuracy, and official ASN or CIDR attribution ranges may not accurately identify the smaller companies that negotiate for the usage of IPv4 space without actually taking ownership of it.

It's worth noting that we did not include the canary port responses in the overall ranking algorithm.

[19] Especially the suspiciously dense spot in the 104/8 subnet toward the bottom left quadrant, covered earlier.

CHARACTERIZING 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.

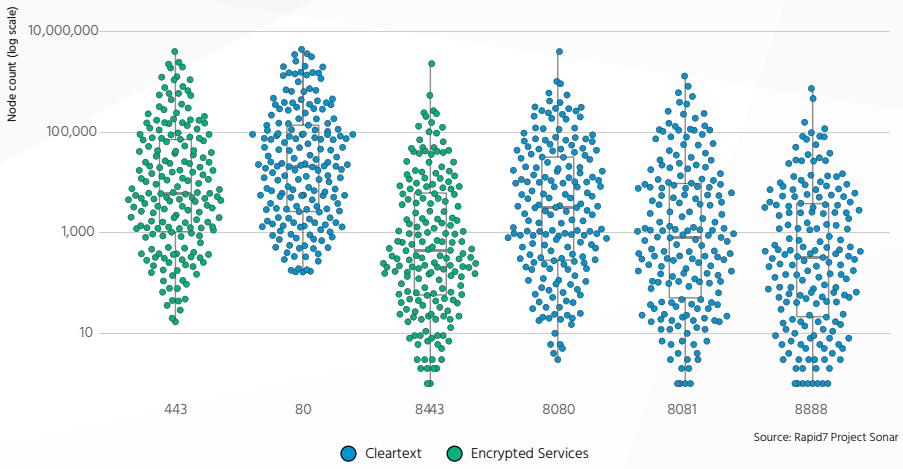
QUASIRANDOM BEESWARMS

In the 2016 National Exposure Index, we introduced box plots as a way to visualize scan data[20]. For this release, we've dressed up these box plots with more relevant data represented as a scatter plot overlay called a "quasirandom beeswarm" plot. While boxplots provide an overall sense of the distribution, the beeswarm overlay makes the picture of the distribution a bit clearer (boxplots can hide both density and outliers, while beeswarms let you see the whole dataset). In all cases, the green scatter plots denote encrypted services, while blue denote cleartext.

Encrypted vs Cleartext Web Ports

Figure 11: Total distribution of encrypted and cleartext web ports

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



Year-Over-Year: The World Wide Web

Unsurprisingly, the World Wide Web (WWW) continues to dominate the internet[21]. Ports 80 and 443 account for 36% of all internet services, and when combined with the alternative, unprivileged ports for web hosting (8000, 8080, 8888, and 8443), these services account for 44% of the listening services across the internet. While there is no doubt that the number of distinct websites continued to grow through 2017—there are over 1.8 billion responsive sites, according to an April 2017 Netcraft study[22]—we did notice a slight drop of about 3 million individual web **servers** since the last National Exposure survey; there are about 2.2 million fewer TCP/80 servers and about 740,000 fewer TCP/443 servers. Out of a universe of about 123 million total web servers across HTTP and HTTPS, this is only about a 2.3% reduction from last year's scans, just barely in the realm of statistically significant. While reducing the number of addressable web servers is a generally positive move for security, even better is the fact that the HTTPS/HTTP ratio has ticked up considerably. When including ports 8000, 8080, 8888, and 8443, we see a 4.3% increase in this ratio.

[20] Please see page 12 of the 2016 National Exposure Index. https://information.rapid7.com/national-exposure-index.html

[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/2017/04/21/april-2017-web-server-survey.html. Note that many sites can reside on one server.

Telnet vs SSH

Figure 12: Total distribution of exposed SSH and telnet services

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

Year-Over-Year: Telnet and Mirai

Much has happened since the publication of the last National Exposure Index report. The Internet of Things (IoT) came to life in the form of the devastating botnet known as Mirai[23]. It took down Twitter and DynDNS and has impacted many other applications, services, and networks as it quickly evolved into a DDoS-botnet-for-hire.

During the 2016 U.S. presidential election, Rapid7 Labs noticed[24] that a measurable number of internet service providers were neutralizing the impact of Mirai by blocking port 23, which is otherwise listening for telnet. This led us to hypothesize that we should see a reduction in nodes responding to TCP port 23 SYN probes this year. As it turns out, there were 5 million fewer nodes responding to our telnet scan than in 2016. This further emphasizes the need for a year-over-year view of the makeup of the internet, if only to see how the distribution of services ebbs and flows with the changing threat and innovation landscape.

SSSHHHH! Don't Tell Telnet, but Port 22 Is Taking Over

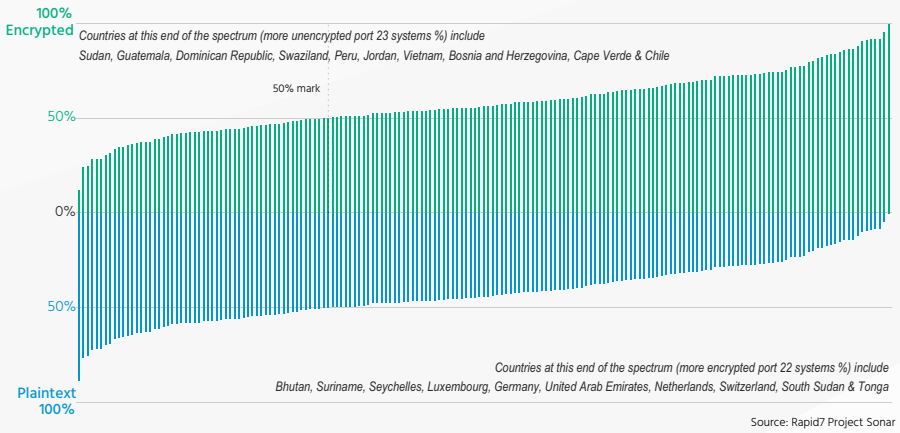
There are more countries with a greater balance—near 50% or higher—of encrypted SSH (port 22) versus unencrypted telnet this year. We've provided one possible reason for that already: the neutering of insecure IoT devices to help mitigate the Mirai botnet. Cloud service providers also make it pretty difficult to expose telnet these days, since the standard configurations exclude it from the default installed packages. Plus, let's face it, SSH is just a more capable protocol that enables ~~sneaky developers to take advantage of~~ both shell access and traffic tunneling. We expect to see the balance continue to tick in the favor of SSH as we press forward with our annual review of internet exposure.

[23] https://community.rapid7.com/community/infosec/blog/2016/10/25/mirai-faq-when-iot-attacks

[24] https://community.rapid7.com/community/infosec/blog/2016/11/08/election-day-tracking-the-mirai-botnet

Figure 13: Distribution of encrypted and plaintext systems with shell access (ports 23 & 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

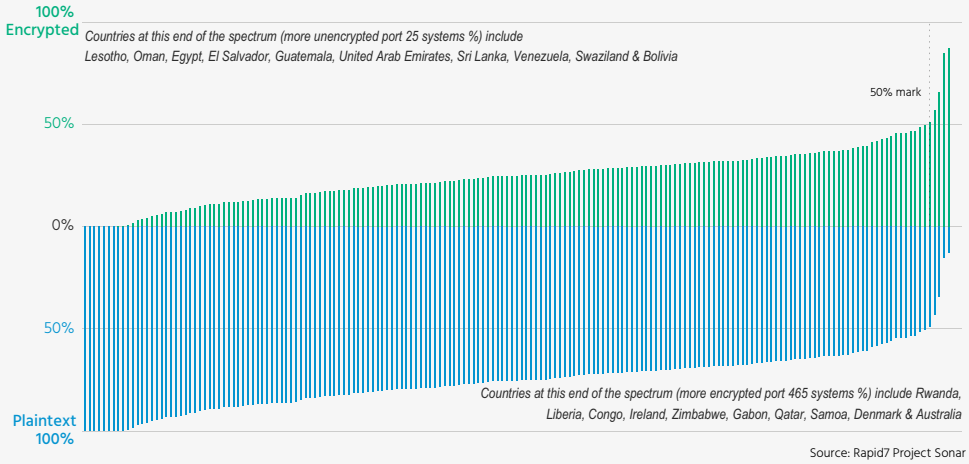


Source: Rapid7 Project Sonar

Email

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

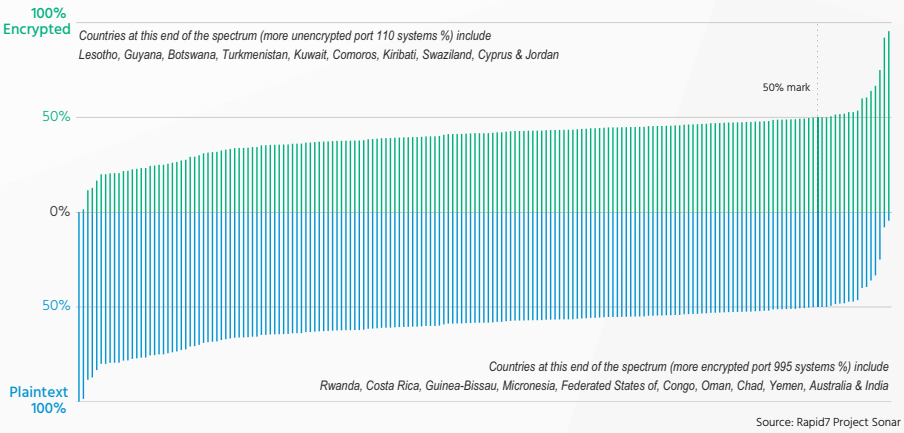
Each column is a single country with the % of encrypted mail-oriented systems above the Y-axis and the % of plaintext mail-oriented systems below the Y-axis



Source: Rapid7 Project Sonar

Figure 15: Distribution of encrypted and plaintext mail access (POP) systems (ports 110 & 995)

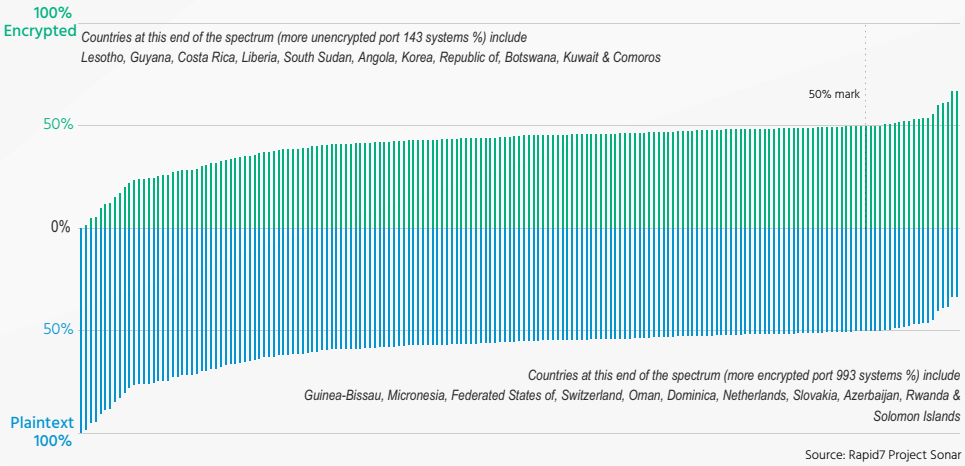
Each column is a single country with the % of encrypted mail access (POP) systems above the Y-axis and the % of plaintext mail access (POP) systems below the Y-axis



Source: Rapid7 Project Sonar

Figure 16: Distribution of encrypted and plaintext mail access (IMAP) systems (ports 143 & 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



Source: Rapid7 Project Sonar

Year-Over-Year: Cleartext Email, Still a Thing?

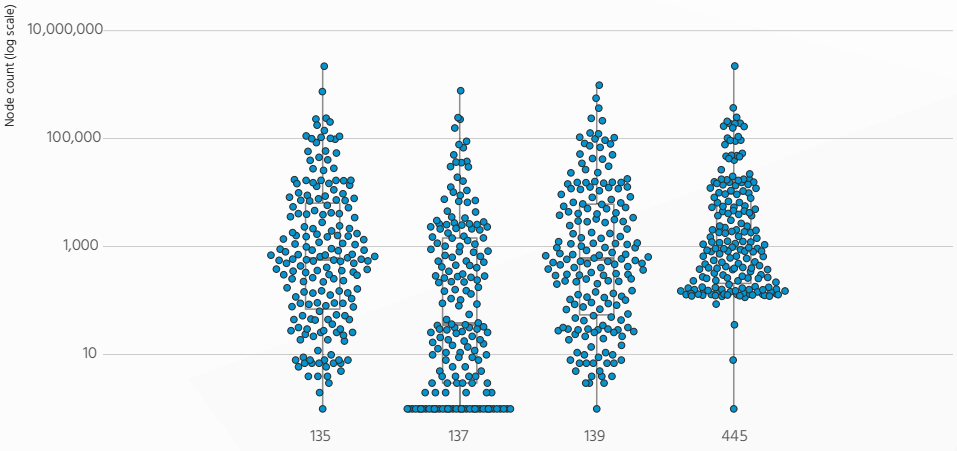
Use of encrypted SMTP mail distribution services had a slight increase across the board when compared with the 2016 study. Virtually every country ticked the balance of encrypted versus plaintext SMTP toward encryption.

The same holds true for POP mail access services, but we noticed a distinct tick in the wrong direction for IMAP services, with fewer countries emerging over the 50% encrypted mark. There's no obvious explanation for this as it makes far more sense to encrypt credentialed access to mail stores versus protecting cat pictures while in transit. Also, there are more centralized mail options—such as Office 365, Google Mail, and others—which make it easy to get mail services up and running while also making it difficult to run them without using encryption. IMAP could be being used as an actual service messaging protocol since it's not restricted to just email usage, but that's pure conjecture at this point.

Microsoft Services

Figure 17: Total distribution of exposed 'Microsoft' services

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



Source: Rapid7 Project Sonar

Year-Over-Year: Clouds vs Big Metal?

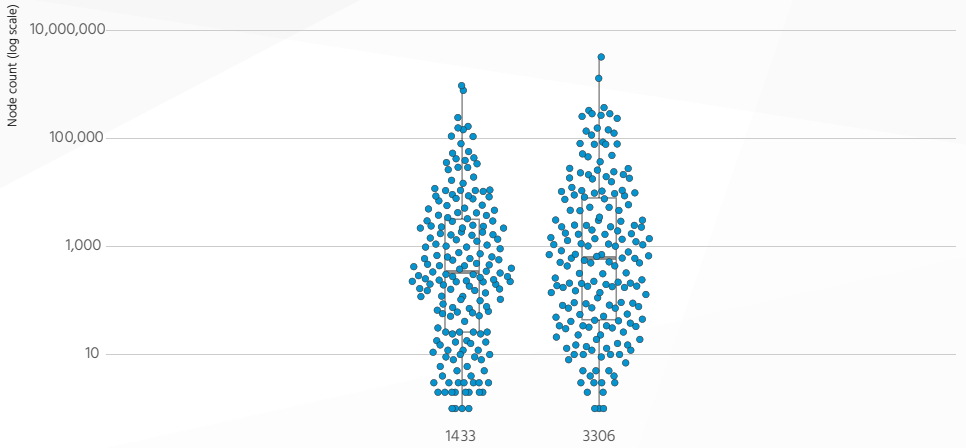
Things did not get better when it comes to exposing Microsoft services. The distributions are nearly identical to 2016, with port 445 even gaining ground.

Virtually every cloud service provider publishes guidance on how to secure Windows servers that sit directly on the internet, yet these ports remain exposed. As mentioned earlier, this exposure caught up with us this year as attackers weaponized[25] government-developed exploits that concentrated heavily on the Microsoft ecosystem. The "WannaCry" ransomworm went after exposed SMB services on port 139 and 445, and our sidebar (see "So Exposed You'll #WANNACRY" on page 15) shows just how serious this exposure truly is.

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



Source: Rapid7 Project Sonar

Year-Over-Year: Databases and Ransomware

Even though our study doesn't focus on the "NoSQL" flavor of databases, we hypothesized that the rash of ransomware attacks that targeted servers[26] in the beginning of 2017 would possibly have a dampening effect on the number of exposed MySQL and SQL server nodes. Unfortunately, this was not the case; the malicious encryption attacks on exposed, unauthenticated NoSQL servers had little impact on the more traditional Transact-SQL style of databases. While we would never wish ill on enterprise exposed databases, it's becoming clear that taking these services offline, en masse, might be prompted only by a widespread attack involving these services directly; a Mirai or WannaCry that's database-centric will certainly cause quite a bit of damage, but will likely ultimately depress this population of exposed databases.

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

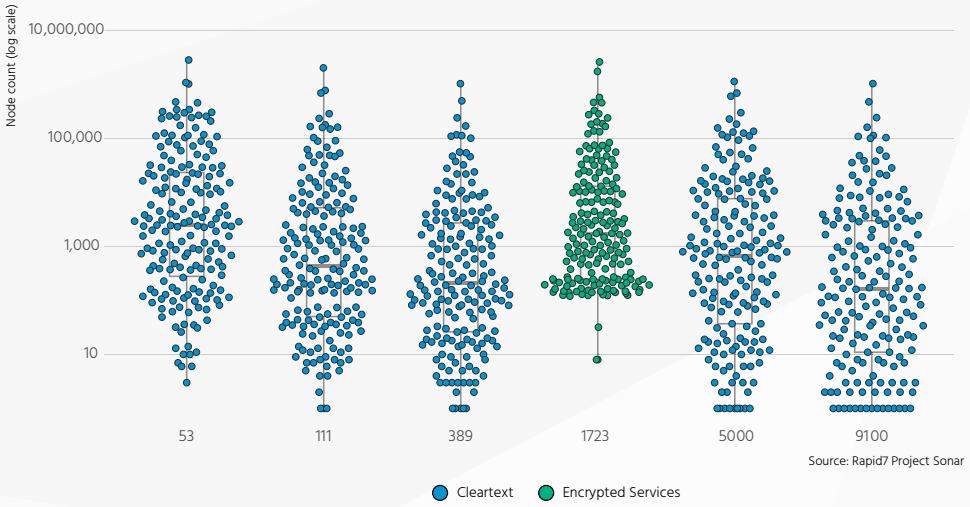
[26] https://community.rapid7.com/community/infosec/blog/2017/01/30/the-ransomware-chronicles-a-devops-survival-guide

Everything Else

We've plotted out the remaining ports on the chart 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. The PPTP endpoints, in particular, form an unusual pattern: if a country has PPTP exposed to the internet, they tend to start off with exposing several hundred of them. We're not sure why this is, but regardless, organizations that still rely on PPTP for VPN services should seriously consider upgrading their deployments to a more secure alternative.

Figure 19: Total distribution of remaining exposed ports

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



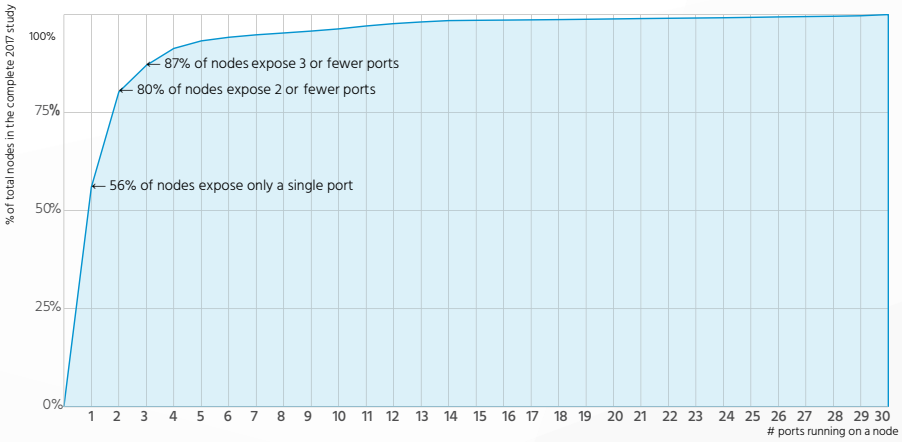
Ports Per Address

Year-Over-Year: Something to Hang This On?

We have one more positive statistic to report: more nodes are running fewer services. There's around a 3% increase in the number of nodes running fewer total ports. Some of this is likely due to the decrease in telnet exposure, but it may be that individuals and organizations are beginning to be more careful about what they expose to the world.

Figure 20: A trend in the right direction

More nodes are running fewer services this year, but only by a hair (3%)



When we did our deep dive into the 104.0.0.0/8 netblock (see the "Islands of Light" sidebar on page 9), the prevalence of DDoS mitigation services struck a chord with our research team. We'll be on the lookout for those nodes in particular next time around and may even try to find a way to rate them more positively since they are performing a vital service to the internet as a whole. We highly suspect they—along with honeypots—are the major contributors to the fraction of systems hanging on to the long tail in figure 20.

NATIONAL EXPOSURE INDEX

The ranking system used for this study is more completely described in Appendix D, but briefly, those countries 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.

For example, Zimbabwe has a total of 88,576 possible IP addresses allocated to it, and 3,295 of those addresses appear responsive to port 23 (telnet), for an overall rate of 3.7% responsiveness to an exposed service. Hong Kong SAR, on the other hand, has almost 200,000 apparent telnet servers, but this is out of a total allocation of over 11.8 million IP addresses. This 1.7% rate (along with the other exposure metrics) is enough to push it to the number two most-exposed spot, even though it has more countable exposed port activity than the more diminutive space of Zimbabwe.

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—whereas the least exposed regions tend to expose well under one percent of their IP address space.

1. Zimbabwe
2. Hong Kong SAR
3. Samoa
4. Congo, Republic of
5. Tajikistan
6. Romania
7. Ireland
8. Lithuania
9. Australia
10. Estonia
11. Denmark
12. Latvia
13. Turkey
14. France
15. Qatar
16. Liberia
17. El Salvador
18. India
19. Rwanda
20. Grenada
21. China
22. Yemen
23. Singapore
24. St. Kitts and Nevis
25. Taiwan
26. Bulgaria
27. South Sudan
28. Mongolia
29. Peru
30. Thailand
31. Spain
32. Dominica
33. Croatia
34. Russian Federation
35. Uzbekistan
36. Guatemala
37. United Kingdom
38. Bangladesh
39. Czech Republic
40. Mozambique
41. Turkmenistan
42. Albania
43. Canada
44. Portugal
45. Ukraine
46. New Zealand
47. Marshall Islands
48. Gabon
49. Jordan
50. Azerbaijan

Year-Over-Year: Country Rerankings

Last year, after we measured all the unique IPv4 addresses exposing one or more of 30 services, we calculated the exposure percentages of each service based on that total number of unique responders. This year, the percentages are based on total allocated IPv4 space. The rationales for this change were manifold.

First, we again compared our SYN responses to scans.io ICMP responses and CAIDA 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. Finally, while our SYN scans measure potential exposure, we had a unique opportunity to dig deep into the Windows SMB protocol and measure actual exposure to the WannaCry threat for each country. We believe these adjustments better reflect individual exposure and cause less skew in either direction.

CONCLUSIONS

At the beginning of the report we noted that Belgium had a significant reduction in exposure (250,000 fewer nodes-per-port on average). That knocked Belgium both out of the top spot and out of the top 50 altogether, even when accounting by the 2016 ranking method. Wealthy regions, such as Australia, China, Hong Kong SAR, France, and the United Kingdom remained on the top 50 list and shifted places a bit.

Zimbabwe claimed the top spot this year, despite their smaller CIDR allocations (only about 90,000 IPv4s). When we peered into the data, one aspect that drove Zimbabwe into the top spot was consistent percentages (versus one or two ports having large outlier percentages). Appendix D on methodology covers the ranking algorithm configuration and changes that also contributed to the year-over-year differences. One major component of that change is that countries that have larger percentages of cleartext services across more overall IPv4 allocated space incur a greater penalty than they did the previous year. This is fairer, overall. All rankings have an unfair component, but this rewards not having all IPv4 space consumed with cleartext services and that seems like a good tradeoff.

The bigger story than the Top 50 index is that we were all witness to the damage exposure can cause since the last report. Server ransomware, ransomworm propagation, insecure Internet of Things, and dozens more headlines reminded us, almost monthly, that the internet is, indeed, a fragile ecosystem that needs deliberate care and attention. Being mindful of both what your organization deploys and how those services are deployed and maintained can have a significant impact on the health of the entire internet. As more individuals and organizations move critical internal and personal services to the cloud, host new applications, and attach more “things” to the public internet, such mindfulness will be more critical than ever if we wish to keep commerce, content, and cat pictures flowing fast and feely.

APPENDIX A: PROJECT SONAR

Project Sonar 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 some 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 40 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 so 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 providing 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 a partnership with the University of Michigan's **scans.io**.

Complete Port Scan Target List

The choices of these 30 ports were guided by both the Nmap services list as well as the collective wisdom of Rapid7 researchers[28].

The top 15 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, while the remaining 15 are protocols which we hypothesized should occur fairly routinely that speak directly to exposure and security.

Figure 21: Ports scanned for National Exposure Index

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	APPROPRIATE?	ENCRYPTED?	% OF TOTAL	DESCRIPTION
80	73,637,628	HTTP	TRUE	FALSE	21.53%	HyperText Transfer Protocol, used to serve web pages and web applications
443	49,762,185	HTTPS	TRUE	TRUE	14.55%	HyperText Transfer Protocol (Secure), an encrypted-by-default means to perform HTTP functions
22	20,213,618	SSH	TRUE	TRUE	5.91%	Secure Shell, an encrypted-by-default alternative to telnet, used for remote administration and protocol tunneling
25	18,297,775	SMTP	TRUE	FALSE	5.35%	Simple Mail Transfer Protocol, used to send email
21	16,980,464	FTP	FALSE	FALSE	4.97%	File Transfer Protocol, used to send and receive data and text files
8080	13,428,979	http-alt0	TRUE	FALSE	3.93%	A common alternative port for HTTP, usually used for websites and web proxy services
53	11,829,288	DNS	TRUE	FALSE	3.46%	Domain Name Service, used to resolve human-memorable names to IP addresses, usually handling longer responses than its UDP counterpart
23	9,995,851	telnet	FALSE	FALSE	2.92%	Telnet, a remote command shell interface, one of the oldest protocols on the internet
1723	9,607,708	PPTP	TRUE	TRUE	2.81%	Point-to-Point Tunneling Protocol, a Virtual Private Network (VPN) service common for older Microsoft Windows servers
143	8,919,856	IMAP	TRUE	FALSE	2.61%	Internet Message Access Protocol, used to receive email
110	8,820,647	POP3	TRUE	FALSE	2.58%	Post Office Protocol version 3, used to receive email
3306	8,279,501	MySQL	FALSE	FALSE	2.42%	MySQL, used to communicate with the (usually) open source MySQL Server published by Oracle
3389	7,279,527	RDP	FALSE	FALSE	2.13%	Remote Desktop Protocol, a graphical user interface to remotely administer (usually) Microsoft Windows servers and desktops

[28] This list is identical to the 2016 National Exposure list, which was selected using source material from both the 2008 Nmap services survey (https://nmap.org/book/nmap-services.html) and a survey of Rapid7's deep bench of data scientists and security researchers.

TCP PORT	OBSERVED COUNT	PROTOCOL/ SERVICE	APPROPRIATE?	ENCRYPTED?	% OF TOTAL	DESCRIPTION
8081	7,170,947	http-alt1	TRUE	FALSE	2.10%	A common alternative port for HTTP, usually used for websites and web proxy services
587	7,056,310	SMTP submission	TRUE	FALSE	2.06%	SMTP submission service, usually used by endpoint mail clients to send email
993	7,027,891	IMAPS	TRUE	TRUE	2.06%	Secure IMAP, an encrypted-by-default alternative to IMAP
995	6,983,958	POP3S	TRUE	TRUE	2.04%	Secure POP3, an encrypted-by-default alternative to POP3
465	6,603,840	SMTPS	TRUE	TRUE	1.93%	Secure SMTP, an encrypted-by-default alternative to SMTP
111	5,953,599	rpcbind	FALSE	FALSE	1.74%	Remote Procedure Call port mapping service, usually used on Unix-like operating systems, usually for NFS file sharing
445	5,547,284	SMB	FALSE	FALSE	1.62%	Server Message Block, a file transfer and remote administration protocol for (usually) Microsoft operating systems
8443	5,429,002	https-alt	TRUE	TRUE	1.59%	A common alternative port for HTTPS, usually used for test websites
135	5,413,613	MS-RPC	FALSE	FALSE	1.58%	Microsoft Remote Procedure Call, usually used on Microsoft operating systems for distributed computing
5000	5,017,418	uPNP	FALSE	FALSE	1.47%	Universal Plug-and-Play, a protocol for machine-to-machine discovery and configuration
139	4,027,291	NBSS	FALSE	FALSE	1.18%	NetBIOS Session Service, used in NetBIOS over TCP/IP, usually on Microsoft operating systems for file and print sharing
5900	3,543,818	RFB	FALSE	FALSE	1.04%	Remote Frame Buffer, a remote GUI for desktop administration, usually used by VNC (Virtual Network Computing)
1433	3,402,532	MSSQL	FALSE	FALSE	1.00%	Microsoft SQL Server service, used to communicate with Microsoft database servers of the same name
990	3,235,369	FTPS	TRUE	TRUE	0.95%	Secure FTP, an encrypted-by-default alternative to FTP
389	2,990,559	LDAP	FALSE	FALSE	0.87%	Lightweight Directory Access Protocol, a directory protocol usually used for authentication and asset lookup
9100	2,951,880	jetdirect	FALSE	FALSE	0.86%	HP JetDirect, a printer control service used to manage print jobs
8888	2,545,753	http-alt8	TRUE	FALSE	0.74%	A common alternative port for HTTP, usually used for websites and web proxy services

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.

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 zeroes), a stupendously large number that is currently impossible to "scan" with any hope of finishing in our species' lifetime. That said, we are working up some shortcuts to get a handle on the IPv6 address space with some reasonable accuracy. Watch for the 2018 report.

Another area of "the internet" we cannot measure are those networks and individual computers that are behind Network Address Translation (NAT) devices and firewalls. Unlike the Internet Census of 2012[29], 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 50 million IP addresses, or about 1% of the total possible addresses, that are consensually off-limits to our scanning[30].

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.

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

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

APPENDIX C: SELECTED COUNTRIES

While last year's 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 opinions (especially by their residents). For ease of reading, this paper refers to members of the International Monetary Fund as "countries" or "nations" interchangeably, irrespective of their official political designations. However, some IMF members are not sovereign entities, such as the Hong Kong Special Administrative Region, and some island states are not IMF members, such as Greenland. This is consistent with IMF documentation, which also uses these terms.

In order to help smooth our investigation and results into a coherent dataset, we first limited our focus to the focus on the 189 members of the IMF[31]. These nations account for well over 99% of all internet activity in the world. In addition, we excluded smaller IMF member countries where our geolocation resources returned more IPv4 addresses than existed in the verified, assigned country CIDR routes at the time of scan results; this reduced the total country count by an additional six: Côte d'Ivoire, Kosovo, Libya, Namibia, St. Lucia, and Tuvalu were all dropped.

Unless otherwise noted, the data and visualizations presented in this paper are therefore limited to only the 183 nations that represent nearly all of the identifiable internet services offered.

IMF MEMBER NATION	TOTAL ALLOCATED IPV4S	WEIGHTED AVERAGE RANK
Zimbabwe	88,576	1
Hong Kong SAR	11,814,400	2
Samoa	17,920	3
Congo, Republic of	53,760	4
Tajikistan	70,144	5
Romania	8,565,504	6
Ireland	6,480,720	7
Lithuania	2,282,752	8
Australia	48,475,904	9
Estonia	1,237,560	10

[31] http://www.imf.org/external/index.htm

IMF MEMBER NATION	TOTAL ALLOCATED IPV4S	WEIGHTED AVERAGE RANK
Denmark	12,573,032	11
Latvia	1,741,824	12
Turkey	16,474,112	13
France	81,346,256	14
Qatar	833,792	15
Liberia	31,232	16
El Salvador	656,640	17
India	41,411,584	18
Rwanda	345,600	19
Grenada	9,728	20
China	338,182,144	21
Yemen	134,144	22
Singapore	6,723,072	23
St. Kitts and Nevis	14,336	24
Taiwan	35,511,552	25
Bulgaria	4,429,056	26
South Sudan	14,336	27
Mongolia	226,048	28
Peru	3,171,072	29
Thailand	9,056,000	30
Spain	30,237,760	31
Dominica	11,520	32
Croatia	2,154,304	33
Russian Federation	45,145,344	34
Uzbekistan	230,912	35
Guatemala	608,512	36
United Kingdom	126,463,384	37
Bangladesh	1,387,008	38
Czech Republic	9,281,920	39
Mozambique	445,184	40
Turkmenistan	17,920	41
Albania	352,000	42
Canada	70,383,360	43
Portugal	6,621,728	44
Ukraine	11,746,656	45
New Zealand	6,855,168	46

IMF MEMBER NATION	TOTAL ALLOCATED IPV4S	WEIGHTED AVERAGE RANK
Marshall Islands	4,096	47
Gabon	355,328	48
Jordan	671,616	49
Azerbaijan	755,712	50
Micronesia, Federated States of	8192	51
Eritrea	4096	52
Lesotho	103424	53
Tanzania, United Republic of	1045760	54
Syrian Arab Republic	1123328	55
Iran, Islamic Republic of	12501760	56
St. Vincent and the Grenadines	26112	57
Lebanon	571904	58
Kuwait	1951232	59
Dominican Republic	1574912	60
Bahrain	460800	61
Lao People's Democratic Republic	72192	62
Iraq	658688	63
United Arab Emirates	3929472	64
Indonesia	18025216	65
Mexico	28842752	66
Cambodia	297216	67
Afghanistan	167936	68
South Africa	28208896	69
Antigua and Barbuda	64512	70
Trinidad and Tobago	533504	71
Swaziland	44288	72
Mali	78848	73
Bolivia	1144320	74
Solomon Islands	11520	75
Jamaica	213760	76
Guinea	26368	77
Poland	20809288	78
Comoros	5120	79

IMF MEMBER NATION	TOTAL ALLOCATED IPV4S	WEIGHTED AVERAGE RANK
Palau	5632	80
Kiribati	4608	81
Botswana	136448	82
Equatorial Guinea	16384	83
Netherlands	50012384	84
Belize	161792	85
Timor-Leste	13568	86
Bosnia and Herzegovina	795392	87
Armenia	612640	88
Kazakhstan	2931200	89
Bahamas	136704	90
Papua New Guinea	58368	91
Guyana	60416	92
Guinea-Bissau	6144	93
Tonga	9728	94
San Marino	34560	95
Moldova, Republic of	1318400	96
Slovakia	2651392	97
Macedonia	683776	98
Bhutan	29696	99
Maldives	61440	100
Honduras	512512	101
Central African Republic	7424	102
Costa Rica	2580224	103
Vanuatu	16640	104
Korea, Republic of	112428032	105
Nepal	513792	106
Serbia	2285824	107
Sri Lanka	545792	108
Malaysia	6586624	109
Kyrgyzstan	272640	110
Panama	1875712	111
Cape Verde	29696	112
Sao Tome and Principe	9216	113
Vietnam	15860992	114
Ecuador	2583040	115

IMF MEMBER NATION	TOTAL ALLOCATED IPV4S	WEIGHTED AVERAGE RANK
Suriname	80128	116
Georgia	1204224	117
Niger	39936	118
Hungary	5887488	119
Madagascar	167680	120
Philippines	5467904	121
Israel	7714304	122
Chad	19968	123
Pakistan	5353216	124
Brazil	83370752	125
Argentina	19109376	126
Cyprus	1060160	127
Nicaragua	404736	128
Germany	119650600	129
Barbados	172800	130
Tunisia	5227008	131
Angola	1206272	132
Haiti	162816	133
Chile	10194944	134
Montenegro	225280	135
Congo, Democratic Republic of the	147712	136
United States	1613186816	137
Colombia	17296896	138
Iceland	882176	139
Burundi	35840	140
Austria	11567456	141
Senegal	361984	142
Greece	5634304	143
Brunei Darussalam	204288	144
Slovenia	2602240	145
Cameroon	446720	146
Egypt	22299136	147
Mauritania	42496	148
Malta	620032	149
Luxembourg	1418752	150

IMF MEMBER NATION	TOTAL ALLOCATED IPV4S	WEIGHTED AVERAGE RANK
Sweden	30289512	151
Norway	15924112	152
Switzerland	20365512	153
Belarus	1833216	154
Japan	203276288	155
Italy	57309312	156
Paraguay	1083648	157
Oman	912384	158
Finland	13696704	159
Djibouti	49664	160
Fiji	142336	161
Myanmar	129024	162
Kenya	5425408	163
Nigeria	2512384	164
Burkina Faso	96256	165
Seychelles	7965696	166
Uruguay	2441216	167
Sierra Leone	51200	168
Venezuela	6796544	169
Saudi Arabia	8905984	170
Sudan	1271040	171
Benin	163328	172
Uganda	887040	173
Mauritius	1869056	174
Morocco	9900288	175
Ghana	1134848	176
Malawi	418304	177
Zambia	1096192	178
Belgium	28502400	179
Gambia	258560	180
Algeria	4791040	181
Ethiopia	361472	182
Togo	304128	183

APPENDIX D: METHODOLOGY

Choosing Ports

We chose to lean on the side of consistency versus novelty when choosing port scan targets for this 2017 study, as consistency enables a longitudinal view of the makeup of the internet, at least through the lens of the National Exposure Index.

We did include two "canary" ports (TCP 5 and 61439), which were used in a study control capacity, and we also scanned TCP port 137, but did not incorporate that data into the final rankings.

Surveying the Internet

Appendix A discussed the technical underpinnings of Project Sonar. Our scanning blacklist grew to just over 50 million restricted IPv4s, up from just over 42 million in 2016—a 19% increase. We, again, compared our scan target results to the ICMP survey by the University of Michigan (our scans.io partner) and are still in the 50% reach range when compared to their "ping" results.

Geolocating Countries

The commercial version of MaxMind's geolocation databases was used to match each IPv4 address to a country. As noted, we performed extra validation of the geolocation results and future studies may incorporate geolocation results from other providers to increase accuracy of the final ratings.

Ranking Exposure by Country

The Exposure Index was created by aggregating the results of 16 individual rankings for exposed, usually cleartext ports: **web, MSSQL, MySQL, SMTP, POP, IMAP, LDAP, RDP, RFB, UPnP, jetd, PPTP, rpcbind, NBSS, MS-RPC, SMB.** We chose these services from the 30 ports covered in the full study scans because there is either a greater likelihood of exposure of sensitive information over cleartext channels with them, or because they expose services that have been identified with extensive vulnerabilities over time.

As noted, there were two changes to the ranking algorithm for the 2017 report.

First, the percentage of exposure for each country was changed to be based on exposed IPv4 prevalence across the total possible IPv4 space for a country versus the prevalence in just combined study IPv4 space for a country. After performing additional research on the 2016 data, we believe this provides a more equitable representation of exposure.

If we take just a look at France (in the top 50) and the United States (not in the top 50), one high-level feature is that France is using more of its allocated IPv4 space than the United States is (the individual port exposure percentages can be reviewed in the previous section):

COUNTRY	IN-USE	ALLOCATED
France	2,589,409 (3.18%)	81,346,256
United States	37,096,215 (2.3%)	1,613,186,816

Deriving a percentage of the SYN counts per port from total in-use nodes penalized countries that aren't spreading exposure across their potential set of address ranges. To put it another way, IPv4 space is a precious, finite resource that should be mindfully managed by individuals, organizations, and network providers, aided by thoughtful government policies and guidance.

The second change was the substitution of aggregated service exposure counts with that of the prevalence of confirmed exposure of full Windows Server Message Block (SMB) services by nodes within a country. The WannaCry ransomworm outbreak provided a very real example of the harm exposure can cause, and we believed it was important to include this as a feature of the final rankings.

We used the same weighted, seeded (using the same seed) Cross Entropy Monte Carlo (CEMC) algorithm[32] 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.

R[33], RStudio[34], and Apache Drill[35] were used for all data processing, analysis, and visualizations. Full datasets, code, and further details on the analyses will be released on Rapid7 Labs' GitHub repository for the report: https://github.com/rapid7/data/.

[32] The Cross-Entropy Method for Continuous Multi-Extremal Optimization; Kroese, Porotsky, Rubinstein; DOI 10.1007/s11009- 006-9753-0

[33] https://r-project.org/

[34] https://rstudio.com/

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

ABOUT RAPID7

With Rapid7, technology professionals gain the clarity, command, and confidence to safely drive innovation and protect against risk. We make it simple to collect operational data across systems, eliminating blind spots and unlocking the information required to securely develop, operate, and manage today's sophisticated applications and services. Our analytics and science transform your data into key insights so you can quickly predict, deter, detect, and remediate attacks and obstacles to productivity. Armed with Rapid7, technology professionals finally gain the insights needed to safely move their business forward. To learn more about Rapid7, visit www.rapid7.com.

QUESTIONS

Reach us at research@rapid7.com