

Internet-Wide Scanning and its Measurement Applications

Zakir Durumeric

University of Michigan



Golden Age of Internet Scanning

As of the last year, it is now possible to scan the entire IPv4 address space in minutes thanks to ZMap and Masscan

Measurement Golden Age: full IPv4 scanning available and IPv6 not widely deployed --- most services still available on IPv4

What can we learn using this global perspective?

What can we do to help network operators?

ZMap: The Internet Scanner

an **open-source tool** that can port scan the entire IPv4 address space from just **one machine** in under **45 minutes** with **98% coverage**



```
$ sudo apt-get install zmap
$ zmap -p 443 -o results.csv
34,132,693 listening hosts
(took 44m12s) ←
```

97% of gigabit
Ethernet linespeed

ZMap: Fast Internet-Wide Scanning and its Security Applications (<https://zmap.io>)
Zakir Durumeric, Eric Wustrow, and J. Alex Halderman | 22nd USENIX Security Symposium.

Ethics of Active Scanning

Considerations

- Impossible to request permission from all owners
- No IP-level equivalent to robots exclusion standard
- Administrators may believe that they are under attack

Reducing Scan Impact

- Scan in random order to avoid overwhelming networks
- Signal benign nature over HTTP and w/ DNS hostnames
- Honor all requests to be excluded from future scans

Measurement Case Studies

What can we learn using Internet-wide Internet scanning?

1. Widespread Weak Cryptographic Keys
2. Analysis of HTTPS Certificate Ecosystem
3. The Matter of Heartbleed

Mining Your Ps and Qs

Detection of Widespread Weak Keys in Network Devices

Nadia Heninger, Zakir Durumeric, Eric Wustrow, J. Alex Halderman
Proceedings of the 21st USENIX Security Symposium, August 2012

Public Keys on the Internet

Uncovering weak cryptographic keys and poor entropy collection

We considered the cryptographic keys used by HTTPS and SSH

	HTTPS	SSH
Live Hosts	12,8 million	10,2 million
Distinct RSA Public Keys	5,6 million	3,8 million
Distinct DSA Public Keys	6.241	2,8 million

There are many legitimate reason that hosts might share keys

Hosting providers, large companies (e.g. Google)

Shared Cryptographic Keys

Why are a large number of hosts sharing cryptographic keys?

We find that 5.6% of TLS hosts and 9.6% of SSH hosts share keys in a vulnerable manner

- Default certificates and keys
- Apparent entropy problems

What other, more serious, problems could be present if devices aren't properly collecting entropy?

Factoring RSA Public Keys

What else could go wrong if devices aren't collecting entropy?

RSA Public Key: $n = p \cdot q$, p and q are two large random primes

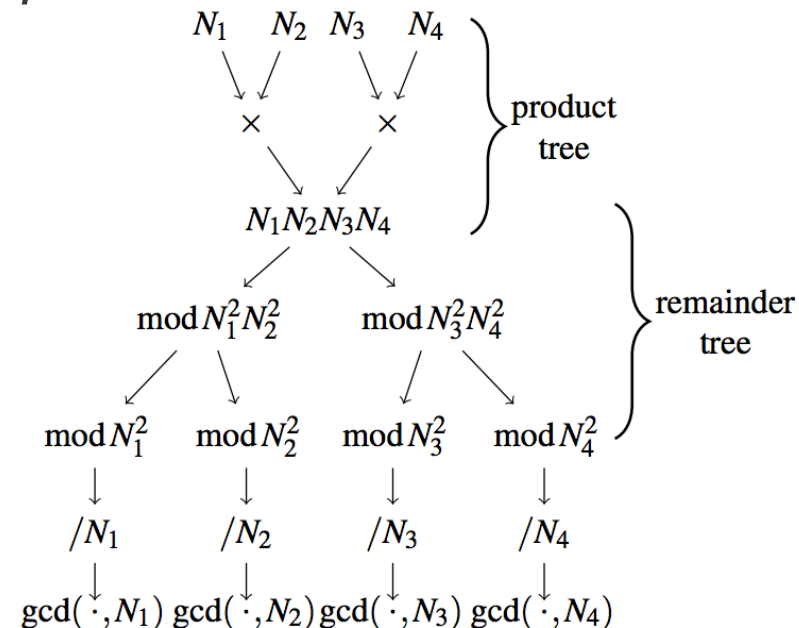
Most efficient known method of compromising an RSA key is to factor n back to p and q

While n is difficult to factor, for

$$N_1 = p \cdot q_1 \text{ and } N_2 = p \cdot q_2$$

we can trivially compute

$$p = \text{GCD}(N_1, N_2)$$



Broken Cryptographic Keys

Why are a large number of hosts sharing cryptographic keys?

We find 2,134 distinct primes and compute the RSA private keys for **64,081 (0.50%) of TLS hosts**

Using a similar approach for DSA, we are able to compute the private keys for **105,728 (1.03%) of SSH hosts**

Compromised keys are generated by headless or embedded network devices

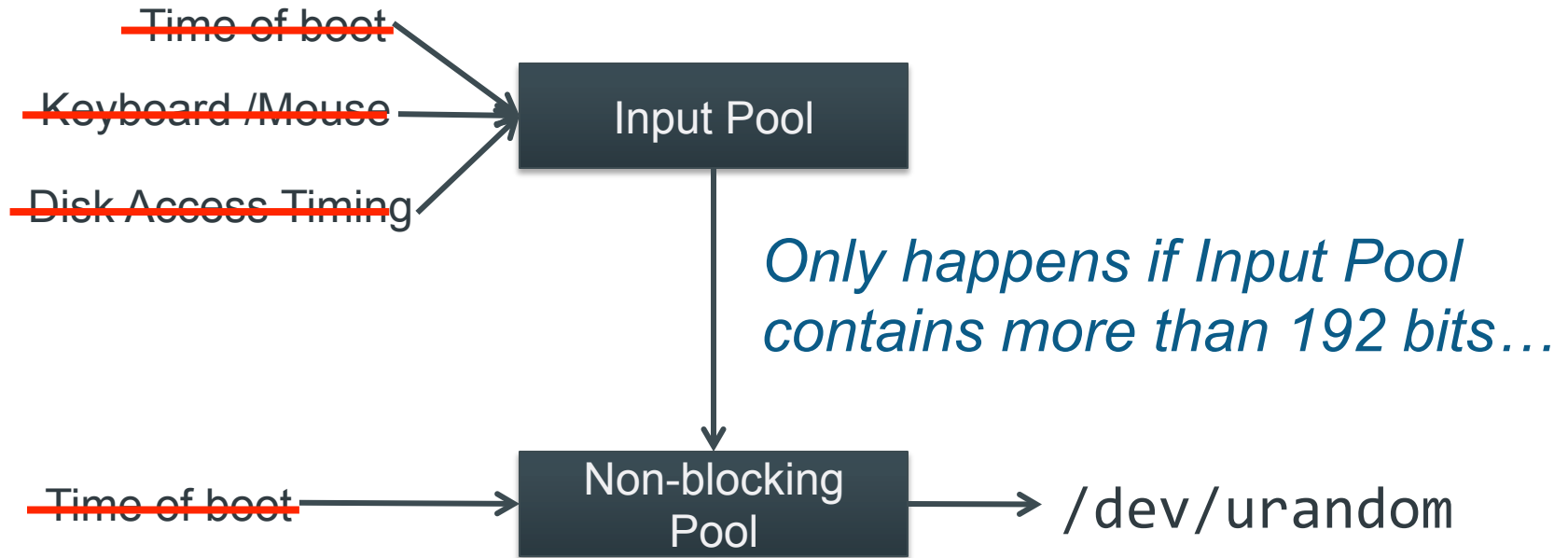
Identified devices from > 40 manufacturers



Linux /dev/urandom

Why are embedded systems generating broken keys?

Nearly everything uses /dev/urandom

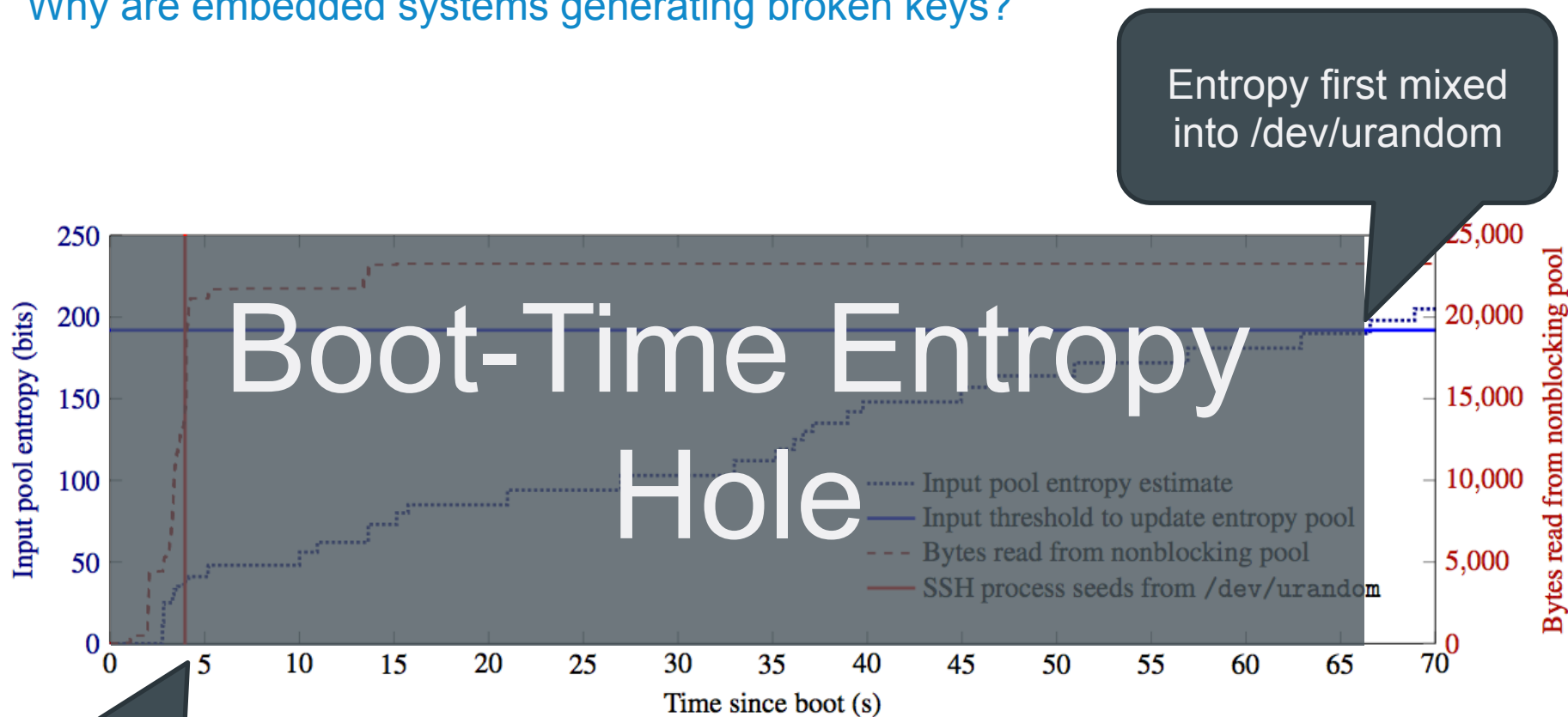


Problem 1: Embedded devices may lack all these sources

Problem 2: /dev/urandom can take a long time to “warm up”

Typical Ubuntu Server Boot

Why are embedded systems generating broken keys?



Entropy first mixed into /dev/urandom

OpenSSH seeds from /dev/urandom

/dev/urandom may be predictable for a period after boot.

Analysis of the HTTPS Certificate Ecosystem

Zakir Durumeric, James Kasten, Michael Bailey, J. Alex Halderman
Proceedings of the 13th Internet Measurement Conference

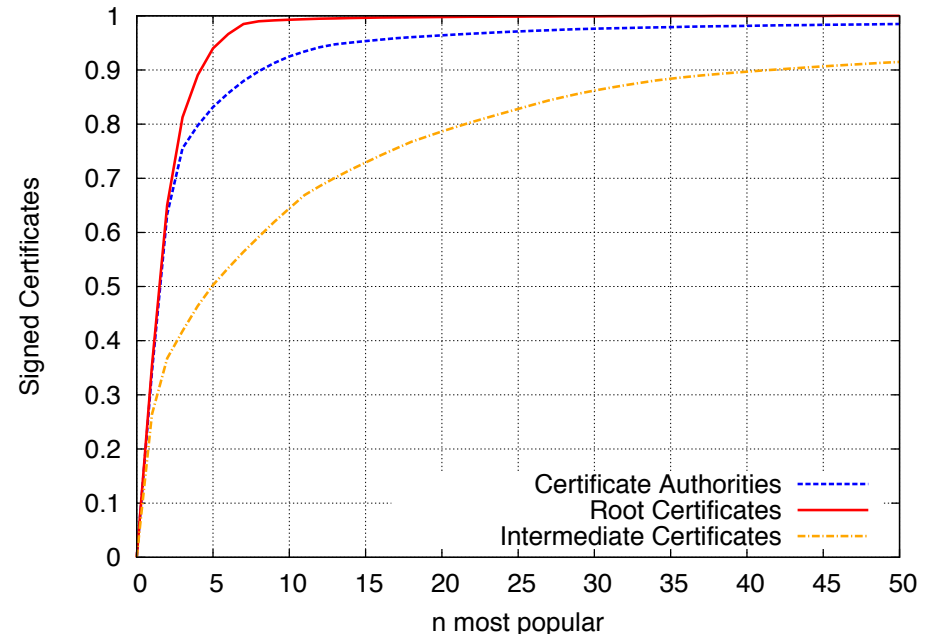
Rampant Certificate Authorities

Daily scans found 88 million total certificates, 9.4 million browser trusted certificates over the last two years

Identified 1,800 CA certificates belonging to 683 organizations

All major roots are selling intermediates to organizations without any constraints

26% of sites are signed by a single certificate!



Ignoring Foundational Principles

What are authorities doing that puts the ecosystem at risk?

We classically teach concepts such as *defense in depth* and the *principle of least privilege*

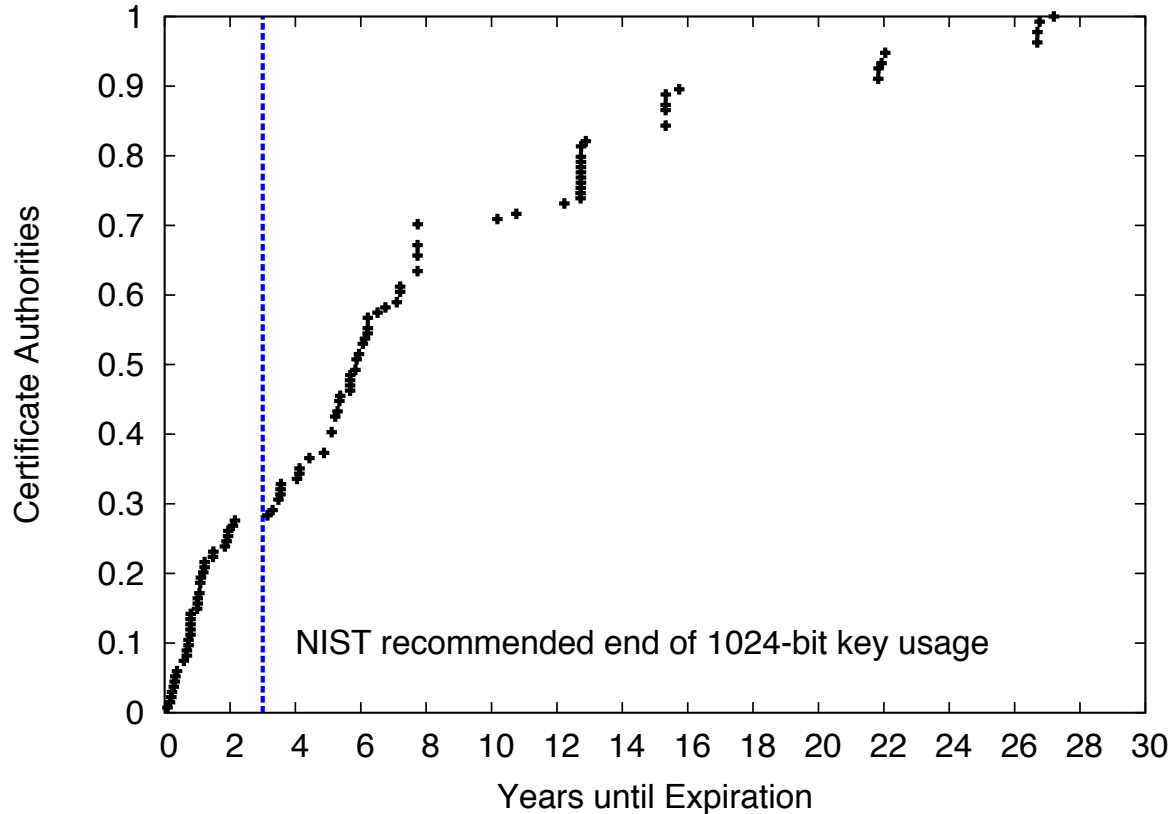
We have methods of constraining what CAs can sign for, yet all but 7 of the 1,800 CA certs we found can sign for anything

Lack of constraints allowed a rogue CA certificate in 2012, but in another case prevented 1,400 invalid certificates

Almost 5% of certificates include local domains, e.g. localhost, mail, exchange

Cryptographic Reality

What are authorities doing that puts the ecosystem at risk?



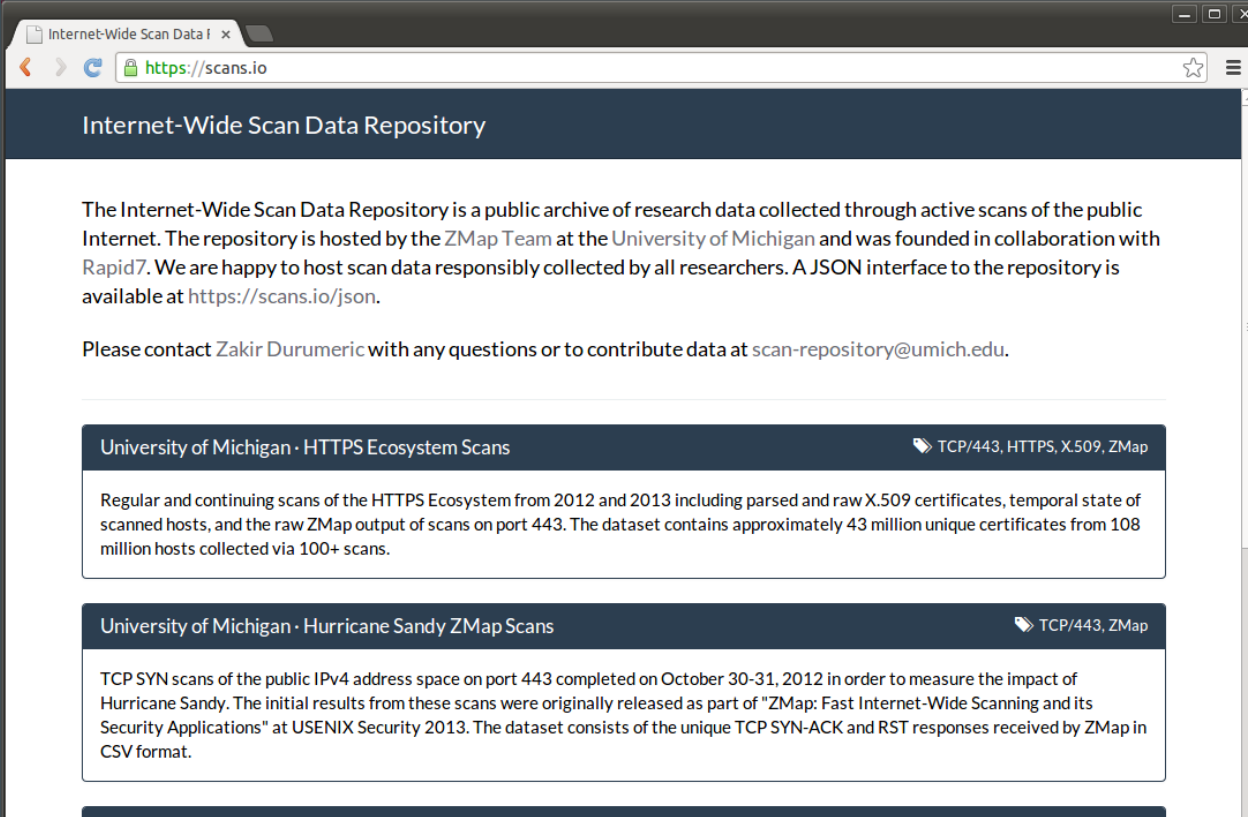
90% of certificates use a 2048 or 4096-bit RSA key

50% of certificates are rooted in a 1024-bit key

More than 70% of these roots will expire after 2016

Scans.IO Data Repository

How do we share all this scan data?



The screenshot shows a web browser window with the URL <https://scans.io>. The page title is "Internet-Wide Scan Data Repository". The main content area contains the following text:

The Internet-Wide Scan Data Repository is a public archive of research data collected through active scans of the public Internet. The repository is hosted by the ZMap Team at the University of Michigan and was founded in collaboration with Rapid7. We are happy to host scan data responsibly collected by all researchers. A JSON interface to the repository is available at <https://scans.io/json>.

Please contact Zakir Durumeric with any questions or to contribute data at scan-repository@umich.edu.

Below this text are two data collection entries:

- University of Michigan · HTTPS Ecosystem Scans** (TCP/443, HTTPS, X.509, ZMap)
Regular and continuing scans of the HTTPS Ecosystem from 2012 and 2013 including parsed and raw X.509 certificates, temporal state of scanned hosts, and the raw ZMap output of scans on port 443. The dataset contains approximately 43 million unique certificates from 108 million hosts collected via 100+ scans.
- University of Michigan · Hurricane Sandy ZMap Scans** (TCP/443, ZMap)
TCP SYN scans of the public IPv4 address space on port 443 completed on October 30-31, 2012 in order to measure the impact of Hurricane Sandy. The initial results from these scans were originally released as part of "ZMap: Fast Internet-Wide Scanning and its Security Applications" at USENIX Security 2013. The dataset consists of the unique TCP SYN-ACK and RST responses received by ZMap in CSV format.



The Matter of Heartbleed

Zakir Durumeric, James Kasten, J. Alex Halderman,
Michael Bailey, Frank Li, Nicholas Weaver, Bernhard Amann,
Jethro Beekman, Mathias Payer, Vern Paxson

Preventing the Spread of Misinformation

<https://zmap.io/heartbleed>

Heartbleed Bug Health Report

The Heartbleed Bug is a vulnerability in the OpenSSL cryptographic library that allows attackers to invisibly read sensitive data from a web server. This potentially includes cryptographic keys, usernames, and passwords. More information and frequently asked questions can be found in the initial disclosure. Information on popular websites that were impacted, but are no longer vulnerable can be found on Mashable's [The Heartbleed Hit List: The Passwords You Need to Change Right Now](#). If you are concerned that a specific website is vulnerable, you can test that website using the [Qualys SSL Server Test](#). If you are a Systems Administrator, the EFF has published [Heartbleed Recovery for System Administrators](#) with information on how to protect services.

Most Popular Vulnerable Domains

Below, we list the top 1,000 most popular web domains and mail servers that remain vulnerable to the heartbleed vulnerability as of 4:00 PM EDT on April 16, 2014. More comprehensive lists of vulnerable [web servers](#) and [mail servers](#) are also available.

Web Servers

Rank	Domain	Vulnerable
1829	gi-akademie.com	vulnerable
1863	prezentacya.ru	vulnerable
1873	wallstcheatsheet.com	vulnerable
1907	semalt.com	vulnerable
2700	gazzetta.gr	vulnerable
3159	protothema.gr	vulnerable
3428	text.ru	vulnerable
3451	haodf.com	vulnerable

Mail Servers

Rank	Domain	Vulnerable
727	turbobit.net	vulnerable
1700	nmisr.com	vulnerable
2100	boerse.bz	vulnerable
2951	ubi.com	vulnerable
3277	filmifullizle.com	vulnerable
3992	uline.com	vulnerable
4081	elektroda.pl	vulnerable
5186	memecenter.com	vulnerable

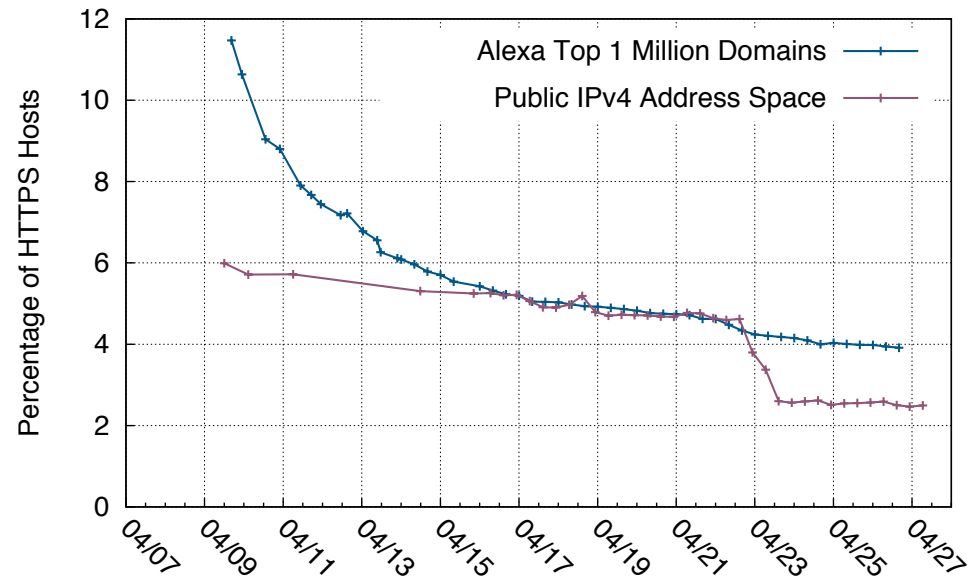
Patching Observations

11% of servers remained vulnerable after 48 hours

Patching plateaued at 4%

Only 10% of sites vulnerable in our first scan replaced their TLS certificates

15% of sites that replaced certificates used vulnerable cryptographic keys



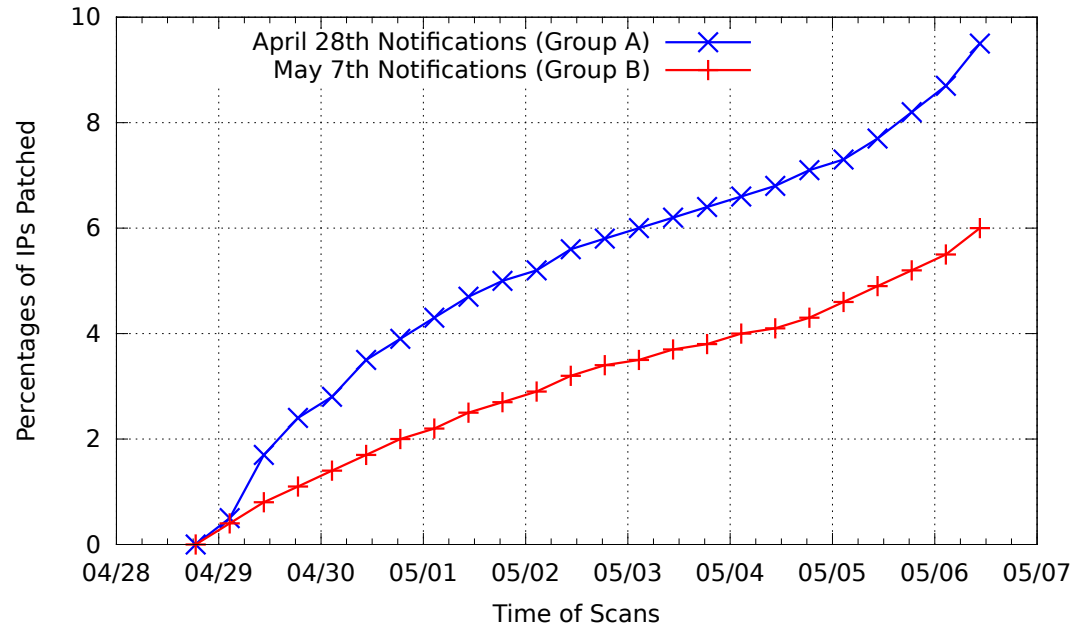
Heartbleed Vulnerable Hosts

Vulnerability Notifications

We notified remaining vulnerable organizations after 2 weeks

Statistically significant impact on patching

Out of 59 human responses: 51 positive, 3 neutral, 2 negative



Impact of Notification

Conclusion

Living in a unique period

IPv4 can be quickly, exhaustively scanned

IPv6 has not yet been widely deployed

ZMap lowers barriers of entry for Internet-wide surveys

Now possible to scan the entire IPv4 address space from **one host** in under **45 minutes** with **98% coverage**

Explored three applications of high-speed scanning

Ultimately hope that ZMap enables future research

Internet-Wide Scanning and its Measurement Applications

ZMap: <https://zmap.io>

Weak Keys: <https://factorable.net>

Public Data: <https://scans.io>

Heartbleed: <https://zmap.io/heartbleed>

Zakir Durumeric, University of Michigan

zakir@umich.edu | [@zakirbpd](https://twitter.com/zakirbpd)