Scan all the things

Project Sonar

Mark Schloesser

@ BlackHat Asia - March 27th 2014

Who Am I

- Mark Schloesser
  - Twitter @repmovsb
  - Security Researcher at Rapid7 Labs
  - Member of The Honeynet Project
  - Core developer for Cuckoo Sandbox
  - Published research on botnets, malware
  - Lots of smaller sideprojects, dexlabs.org (Android), honeypots, protocols

Outline

- Intro / History / Motivation / Ethics / etc
- Project Sonar
- Big-enough-data
- Findings
- Future work
- Conclusion

Large scale scanning
Internet wide data-gathering

Limited history of Internet-wide scanning

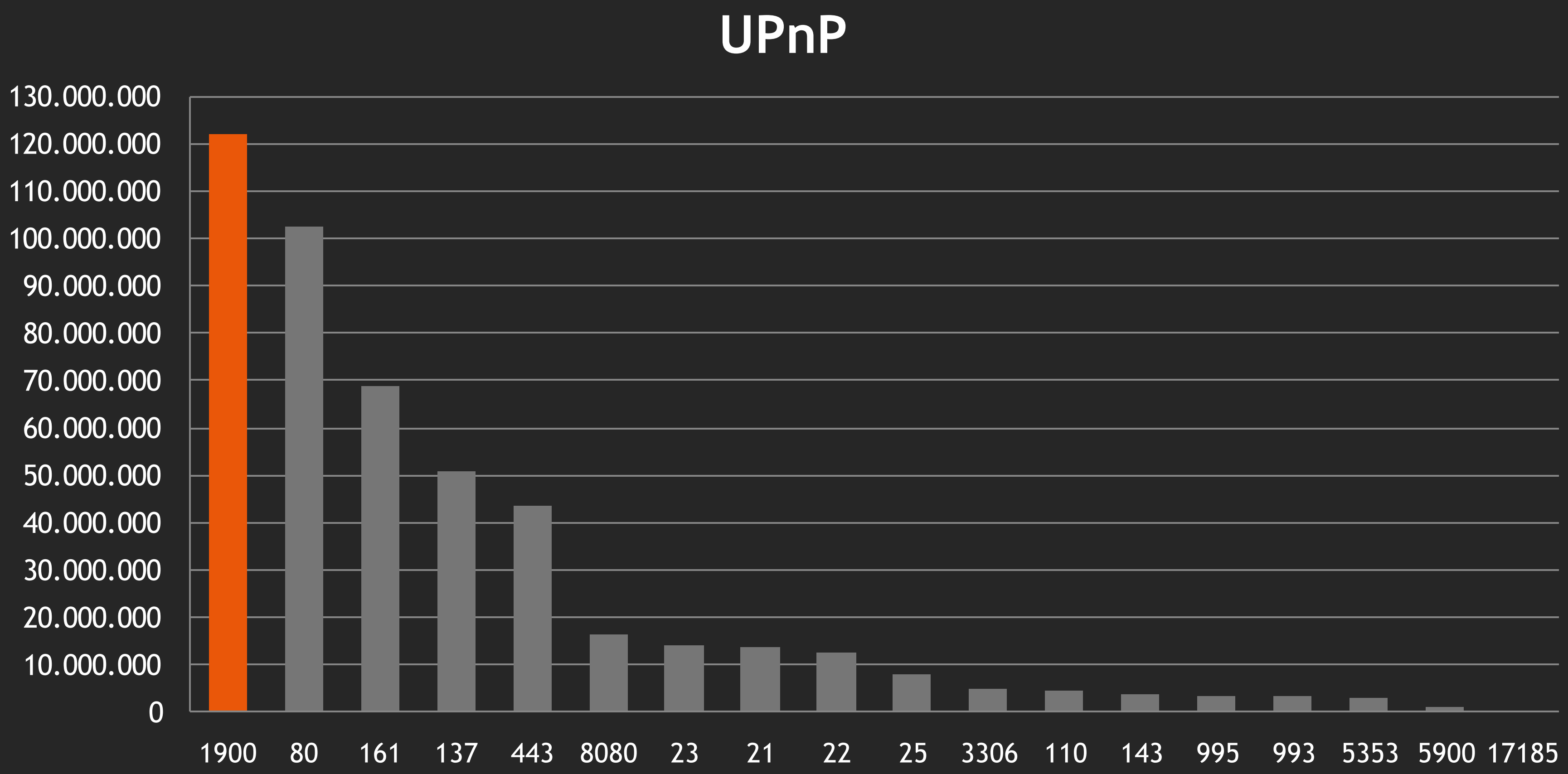
- Internet Mapping Project, 1998
  - Started 1998 at Bell Labs, moved to “Lumeta Corporation”
  - Traceroute-style probes, generating graph visuals
- IPv4 Census 2003-2006
  - University of Southern California, caida.org, ICMP echo
- EFF SSL Observatory 2014
  - Analyzes publicly visible SSL landscape, finds variety of misconfigurations and flaws
  - Weak keys, >600 odd Certificate Authorities, etc.

History cont.

- Internet Census 2012
  - Anonymous individuals conducted with a botnet of infected (mostly embedded) devices
  - Results published online and available for download (9TB unziped)
- Shodan
  - A search engine for Internet-connected devices
- RIPE Atlas
  - Distributed measurements (ping, traceroute, SSL, >6k nodes)
- Critical.IO
  - HD Moore proof-of-concept project, run 2012-2013
  - Provided for research, great outcome (UPnP, IPMI, Serial port servers)

Quick excursion into the most popular service on the Internet

#1 Universal Plug and Play



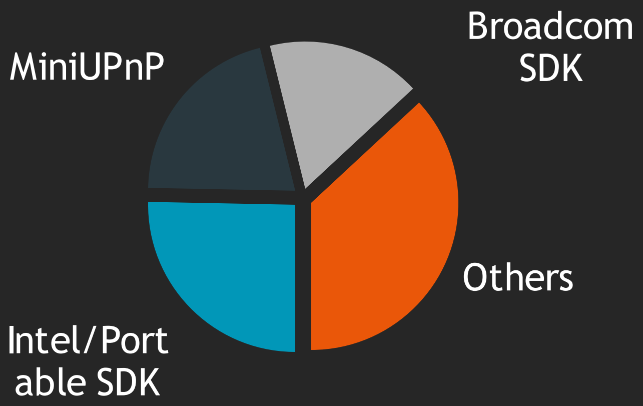
RAPID7

UPnP

- UPnP is a set of network protocols used by devices to discover each other’s presence and establish network services.
- UPnP is most commonly used by consumer devices
  - Routers, personal computers, printers, gaming consoles
- UPnP is also used across enterprise systems
  - Security DVRs, NAS servers, IP cameras, Supermicro IPMI controllers

Pwning UPnP

- The top 3 UPnP stacks are exploitable (63%)
  - Eight distinct buffer overflows in Intel/Portable SDK SSDP code
  - Stack overwrite in MiniUPnP 1.0 SOAP action processor
  - Format string in the Broadcom SOAP IGD service



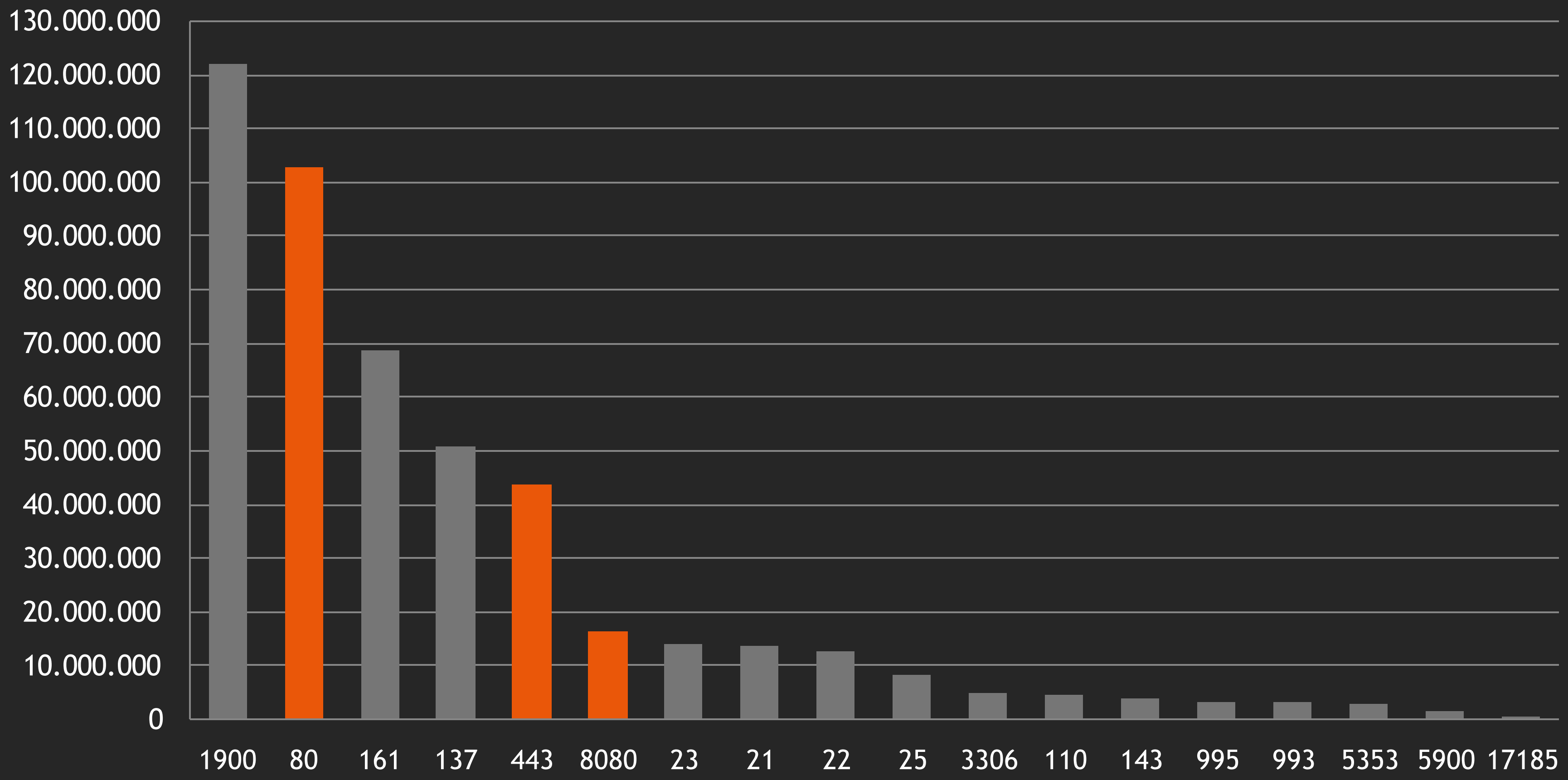
RAPID7

- UPnP - critical vulnerabilities
- Raising awareness
- Hopefully reducing attack surface

What else is out there?

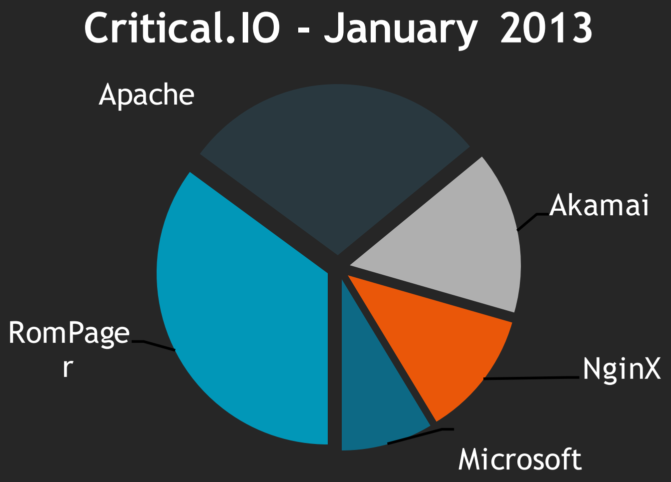
WELCOME TO THE INTERNET!

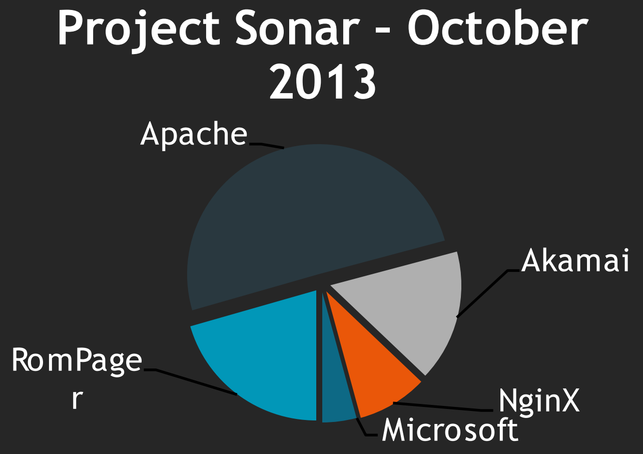
#2 HTTP



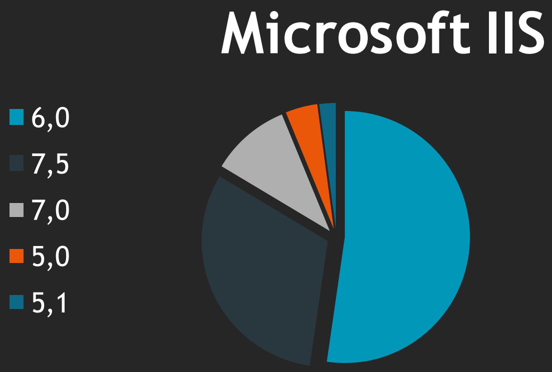
RAPID7

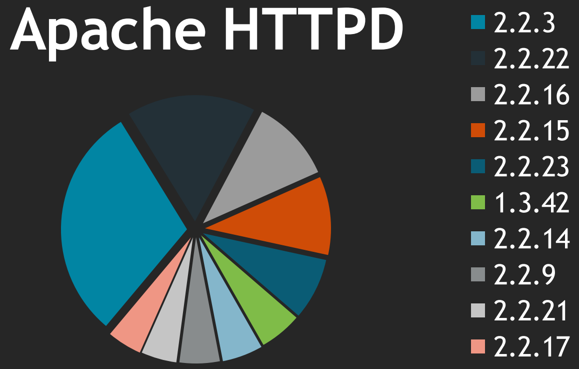
HTTP statistics


Critical.IO - January 2013

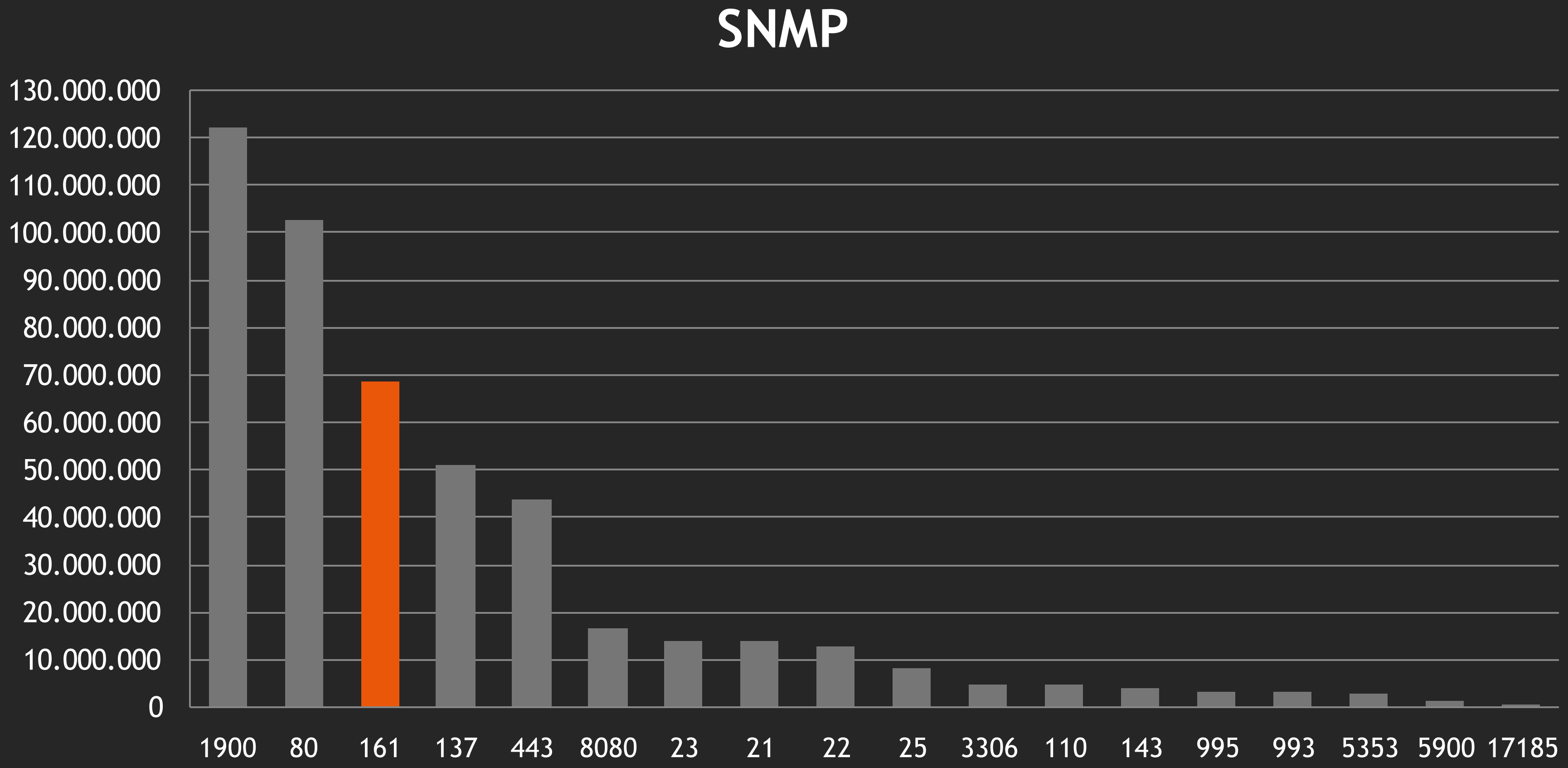

Project Sonar - October 2013

- Most common versions of Apache and IIS are outdated


Microsoft IIS


Apache HTTPD

#3 Simple Network Management Protocol



RAPID7

SNMP

Devices

- Home Routers
- Printers
- Modems
- Switches
- Enterprise Gear

Features

- Routes, addresses, listening ports
- Running processes and services
- Installed software and patches
- Accounts and group names
- DDoS via amplification

SNMP - list processes, get credentials

username=sa password=Masterkey2011 LicenseCheck=Defne
DSN=sms;UID=XXX;PWD=XXXsys; DSN=GeoXXX;UID=XXX;PWD=XXXsys; 8383
password h4ve@gr8d3y
--daemon --port 8020 --socks5 --s_user Windows --s_password System
XXXX /ssh /auth=password /user=admin /passwd=admin_p@s$word
http://a.b.c/manage/retail_login.php3?ms_id=14320101&passwd=7325
a.b.c.d:3389 --user administrator --pass passw0rd123

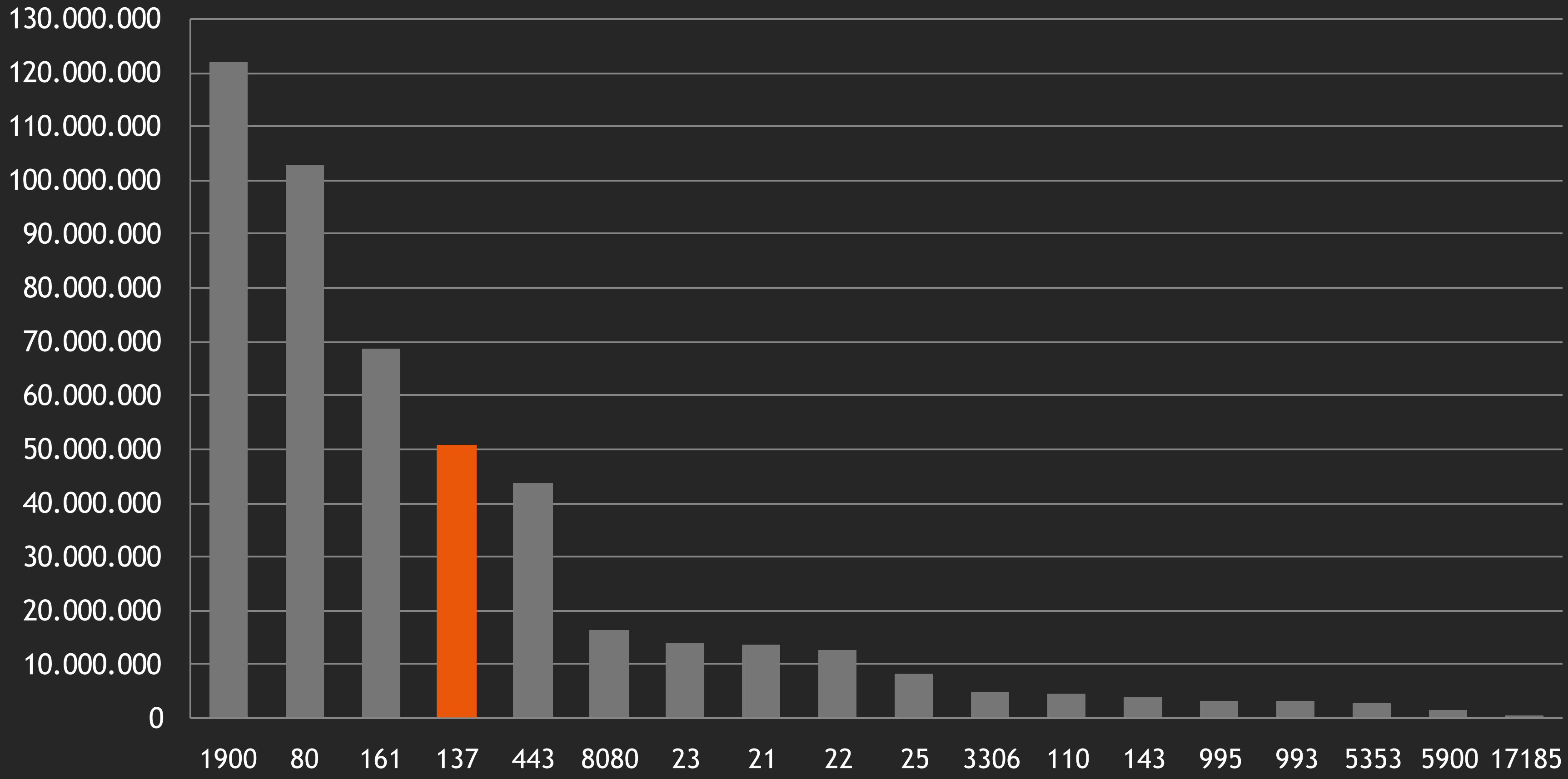
SNMP Write

- Authentication based on “community string”
- Defaults to “public” read-only and “private” read-write on a lot of devices
- Some devices also allow write with “public”
  - Reconfigure functionality across 11 millions devices?
  - Thousands of routers can be instantly compromised

Traffic Amplification

- As recently seen in the DDoS attacks utilizing NTP (monlist command)
- DNS is bad enough
- SNMP is worse

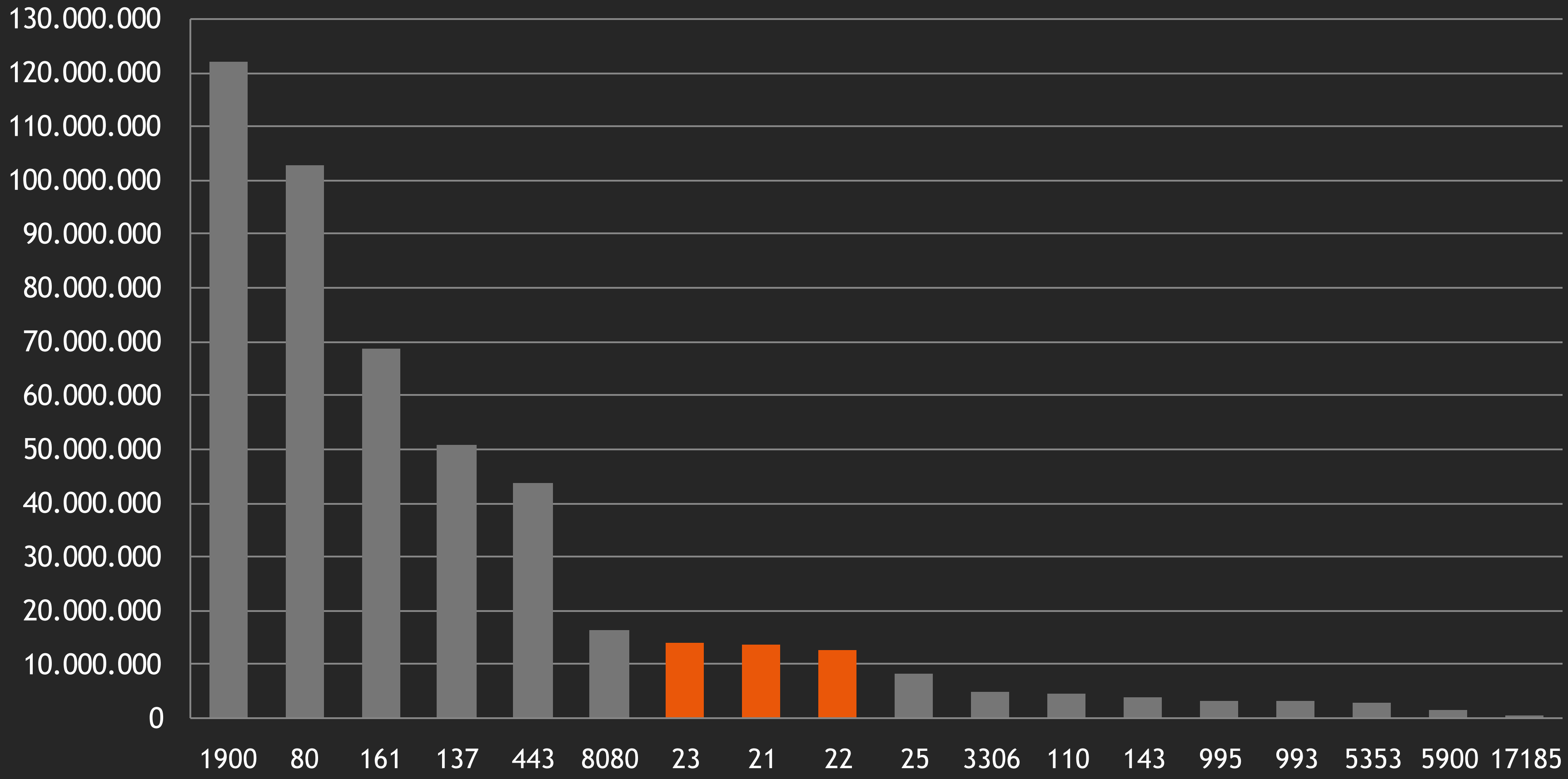
#4 NetBIOS



130.000.000
120.000.000
110.000.000
100.000.000
90.000.000
80.000.000
70.000.000
60.000.000
50.000.000
40.000.000
30.000.000
20.000.000
10.000.000
0
1900
80
161
137
443
8080
23
21
22
25
3306
110
143
995
993
5353
5900
17185

Let’s skip NetBIOS / SMB / CIFS ;)

#5 Telnet, FTP, SSH



RAPID7

Telnet: Router Shells

10,000+ Routers don’t even bother with passwords

```
jiuyuan_bt_nm_ah>
jiyougongsi>
jjcaisanxiaoxue>
jjda>
jjdc>
jjgd>
jjlhlianfangzhizao>
jjpzx>
jjshhshengangzhizao>
jjxjy>
jjxy>
jjxz>
jjyljuda>
jkx_sdl>
jnszy_2692>
joelsmith>
jsyh>
jt_net>
jtic>
jx123>
jzglkyzz>
kashiwa>
kbbmetro>
kd-ip>
mp1700-kslp>
mp1700E>
mp1762>
mp2600e>
mp2692>
mp2700>
msk-cat3>
mty-3500-1>
multivoice01>
mvy-rtr-01>
mx-fdc-dmz1>
mx-frtsw01>
mx-frtsw02>
nak2ama-east-ps>
nak2ama-north-ps>
nak2ama-ps>
nak2ama-south-ps>
nak2ama-west-ps>
naldi>
nanchang2621>
nanquc3550-02>
nanshigaosu_A5>
narashino>
nayana2>
telnet@AYRS-CES2k-1>
telnet@AdminVideoSW1>
telnet@BBG>
telnet@BEL-WIFI-1>
telnet@BGLWANSW01>
telnet@BGLWANSW02>
telnet@BI-RX-1>
telnet@BI-Solsi>
telnet@BIGION-CORE-1>
telnet@BR2-NET1-MLXe>
telnet@BRCD-ADX-2>
telnet@BSI01>
telnet@Backbone_Backup>
telnet@BigIron RX-4 Router>
telnet@BigIron RX-8 Router>
telnet@BigIron Router>
telnet@Bloco.A1.Core>
telnet@Bloco.B.Core>
telnet@Border40G-1>
telnet@Brocade_ABA_1>
telnet@CHD-BOU-CO-2>
telnet@CON-LONFESX4801>
telnet@CON-LONFESX4802>
S1-DNS-3560-NSGK>
```

Telnet: Windows CE Shells

3,000+ Windows CE devices drop CMD shells

```
Welcome to the Windows CE Telnet Service on WindowsCE Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on ITP Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on WindowsCE Pocket CMD v 6.00 \>
Welcome to the Windows CE Telnet Service on WindowsCE Pocket CMD v 4.20 \>
Welcome to the Windows CE Telnet Service on PicoCOM2-Sielaff Pocket CMD v 6.00 \>
Welcome to the Windows CE Telnet Service on WindowsCE Pocket CMD v 4.10 \>
Welcome to the Windows CE Telnet Service on G4-XRC Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on HMI_Panel Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on G4-XFC Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on PELOAD Pocket CMD v 6.00 \>
Welcome to the Windows CE Telnet Service on MCGS Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on Db1200 Pocket CMD v 5.0 \>
Welcome to the Windows CE Telnet Service on VEUIICE Pocket CMD v 6.00 \>
Welcome to the Windows CE Telnet Service on Borne Cebus/Horus Pocket CMD v 6.00 \>
```

Telnet: Linux Shells

3,000+ Linux systems drop to root

```
MontaVista(R) Linux(R) Professional Edition 4.0.1 (0502020) Linux/armv5tejl
Welcome telnet root@~#
Local system time:  Sun May 20 04:12:49 UTC 2012 root:#
root@(unknown):/#
root@routon-h1:/#
root@umts_spyder:/ #
root@vanquish_u:/ #
root@smi:/ #
root@dinara_cg:/ #
root@BCS5200:/#
root@edison:/ #
root@umts_yangtze:/ #
root@cdma_spyder:/ #
root@vanquish:/ #
root@scorpion_mini:/ #
root@qinara:/ #
sh-3.00#
~ #
```

Telnet: other stuff

License plate readers, on the internet, via Telnet

ATZ P372 application Aug 29 2008 16:07:45 P372 RAM: 128M @ 128M EPROM: 512k Flex capabilities 003f Camera firmware: 4.34 362 ANPR enabled for: USA Louisiana . Installed options: 00220018 * ... Compact Flash * ... Basic VES with no security * ... USA Licenceplate recognition * **PIPS Technology AUTOPLATE (tm) license plate recognition** * VES - (violation enforcement system)

Telnet: GPS tracking

GPS tracking systems (Ankara, Turkey)

```
QM Extension: 2012/05/21 13:53:40.343
1067|PESQHandler.c{UE2 } 0x03e8 Last Sync: 1, Current Sync:
2, RTU played: 0 2012/05/21 13:53:40.343
1068|PESQHandler.c{UE2 } 0x03e8 Jitter: -62 0x0380 Qual=1
Valid=YES HDOP=1 PDOP=1.88 Lat=39.96039 Long=32.71275
Satellites=9 Heading=286 Speed=21 Altitude=825
2012/05/21 13:53:40.734 500|GpsNmeaStandar{GPS } 0x0380
```

Stay strong, it’s not over…

6
7
8
9

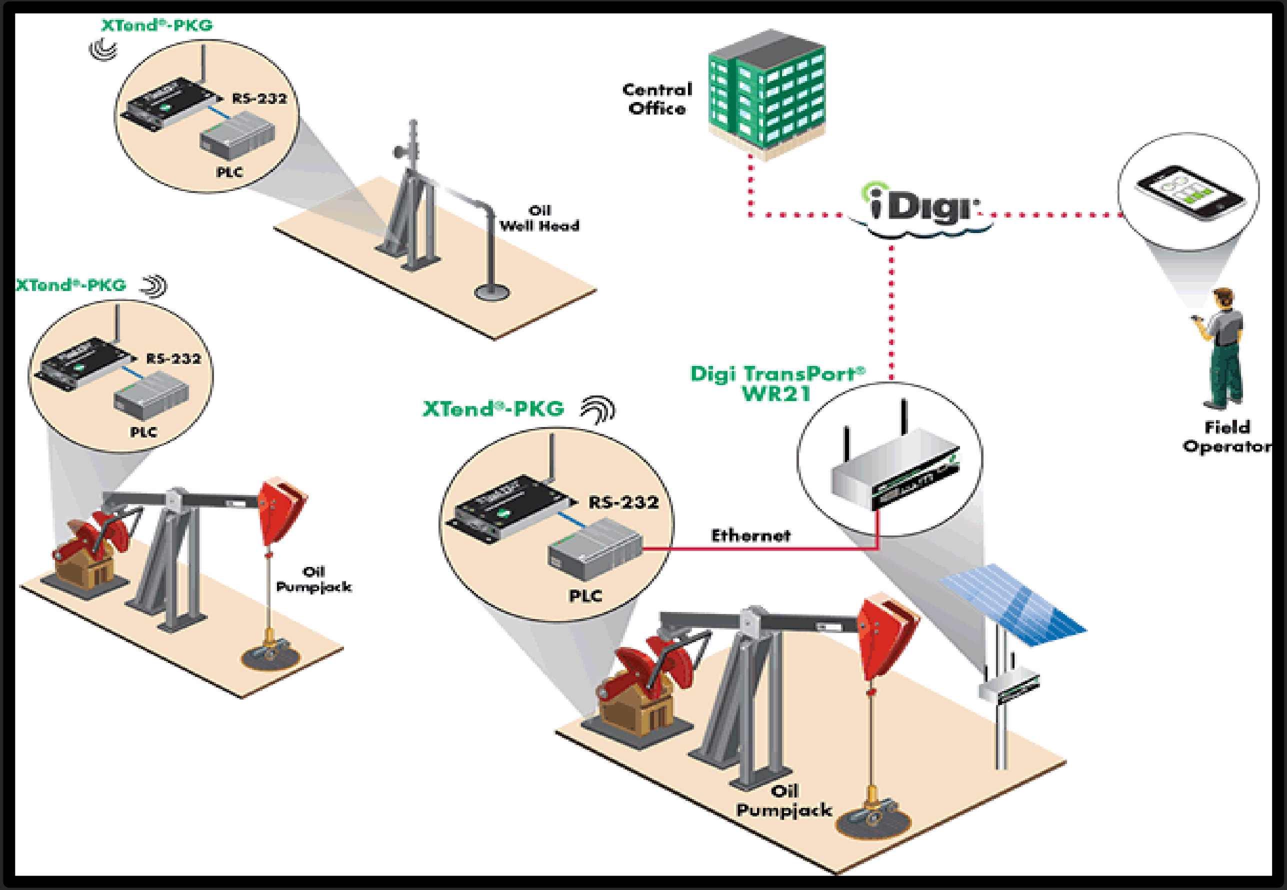
Serial Port Servers

- Devices that make network-disabled devices into network-enabled ones.
- Doesn’t sound like a good idea…

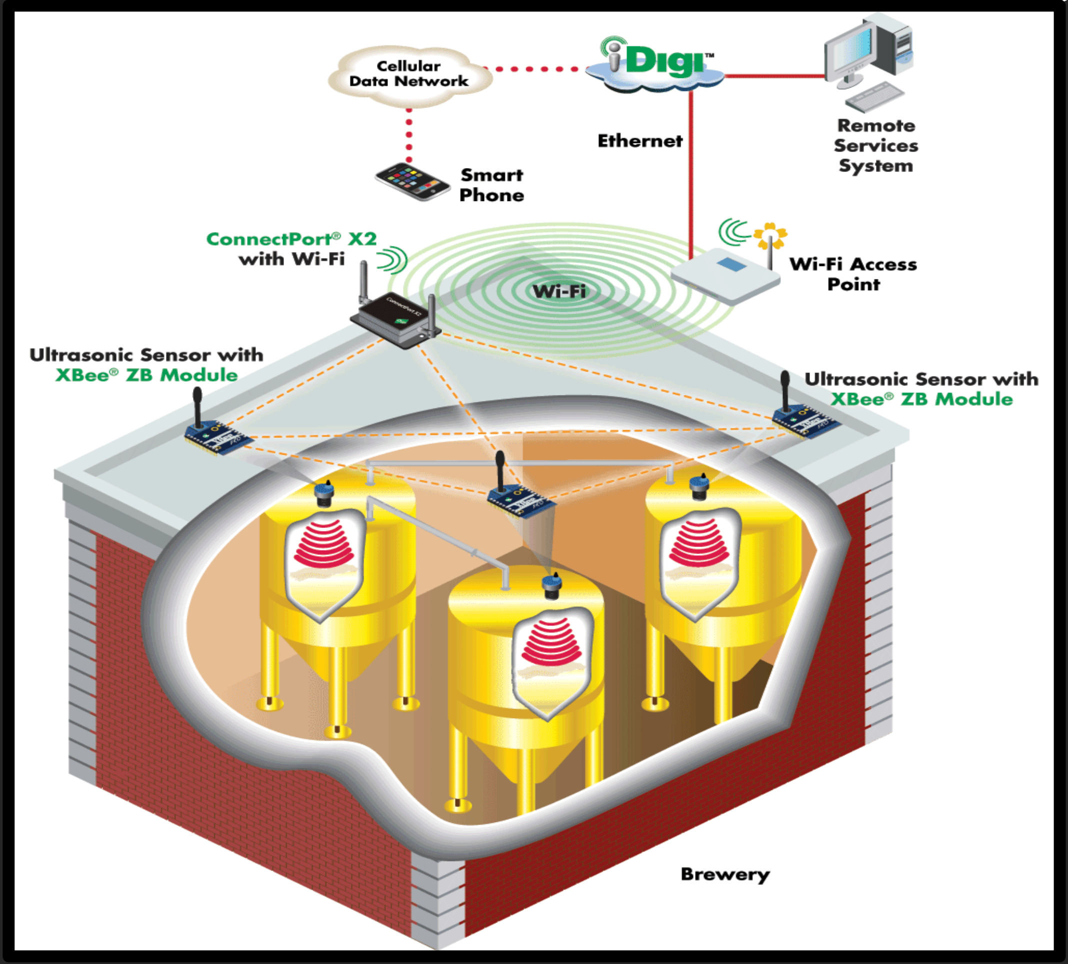
Serial Port Servers: Features

- Remote serial port access
  - Interact with target ports through telnet, SSH, and HTTP
  - TCP socket proxy ports provide direct pass-through
  - Proprietary protocols for virtual COM port drivers
- Serial port monitoring and automation
  - Some products offer basic automated interaction
  - Use expect-style logic, can alert, send commands
  - Stream to remote hosts when criteria are met

Use Case: Oil and Gas Monitoring

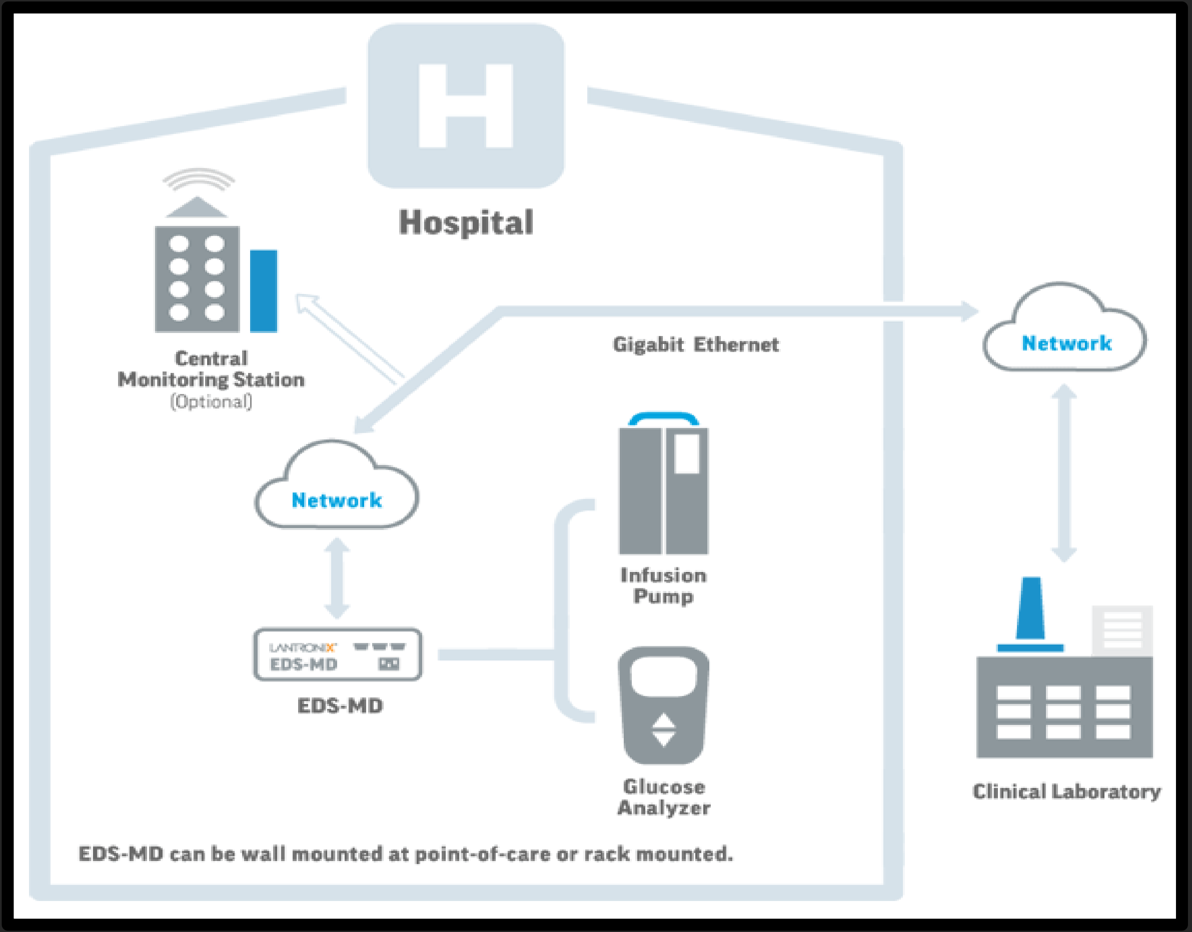


Use Cases: Brewery Tank Monitoring


Cellular Data Network
Digi
Remote Services System
Ethernet
Smart Phone
ConnectPort® X2 with Wi-Fi
Wi-Fi
Wi-Fi Access Point
Ultrasonic Sensor with XBee® ZB Module
Ultrasonic Sensor with XBee® ZB Module
Brewery

RAPID7

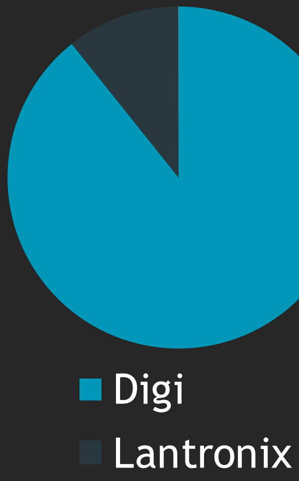
Use Case: Medical Device Monitoring

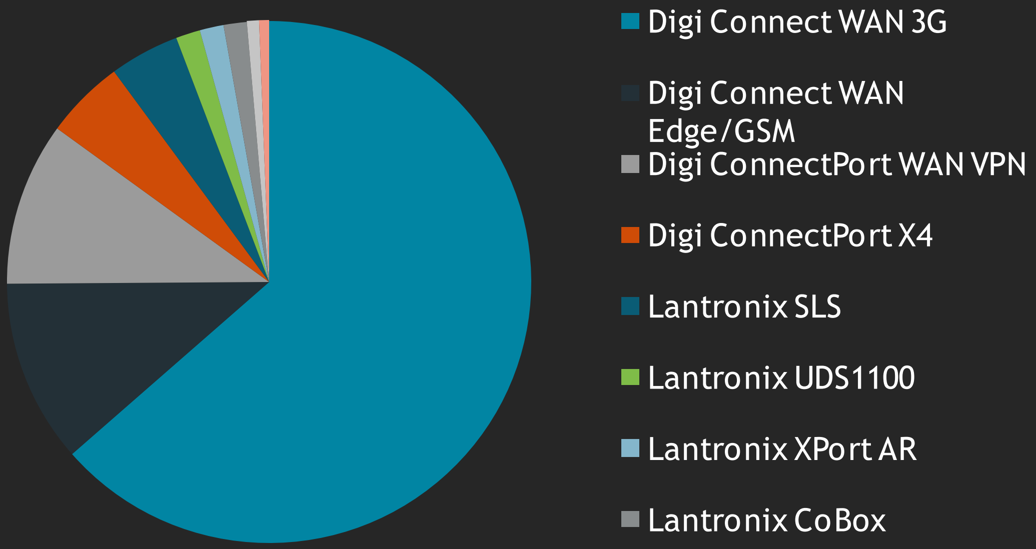


Serial Port Device Exposure: SNMP

SNMP “public” System Description

- Over 114,000 Digi and Lantronix devices expose SNMP
- Over 95,000 Digi devices connected via GPRS, EDGE, & 3G





Serial Port Access Authentication

Access Methods:

- ☺ Authenticated encrypted TCP multiplex ports
- ☺ Authenticated, encrypted ssh or web consoles
- 😐 Authenticated, clear-text telnet or web consoles
- 😐 Authenticated clear-text TCP multiplex ports
- ☹ Unauthenticated clear-text TCP multiplex ports
- ☹ Unauthenticated TCP pass-through ports
- ☹ Unauthenticated encrypted TCP multiplexed ports
- ☹ Unauthenticated UDP mapped ports

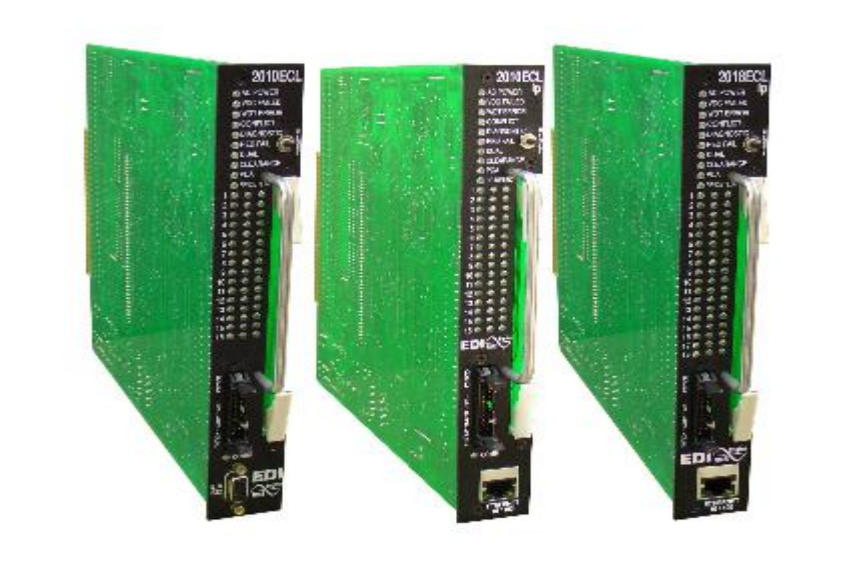
Serial Port Access Authentication

Guess which are most common?

- ☺ Authenticated encrypted TCP multiplex ports
- ☺ Authenticated, encrypted ssh or web consoles
- 😐 Authenticated, clear-text telnet or web consoles
- 😐 Authenticated clear-text TCP multiplex ports
- **☹ Unauthenticated clear-text TCP multiplex ports**
- **☹ Unauthenticated TCP pass-through ports**
- ☹ Unauthenticated encrypted TCP multiplexed ports
- ☹ Unauthenticated UDP mapped ports

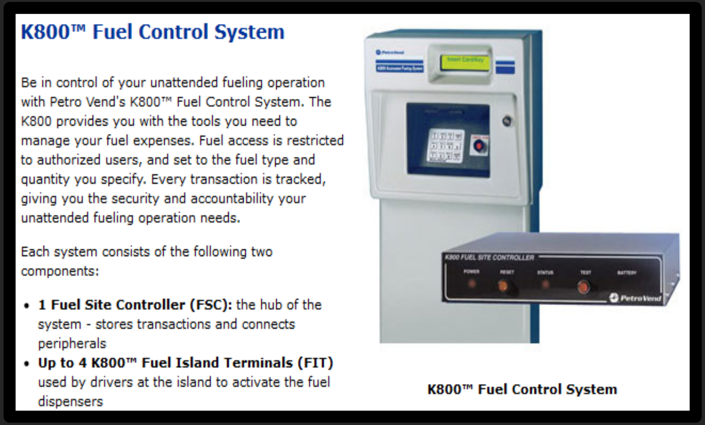
EDI Traffic Signal Monitors

- Based on Digi development kits, exposes ADDP
  - Default password is “dbps” as a result
  - ~40 or so identified in the Internet Census 2012 data



K800 Fuel Control Systems

- Often connected through Digi serial port servers
  - Appears to be a x86 board managed via serial



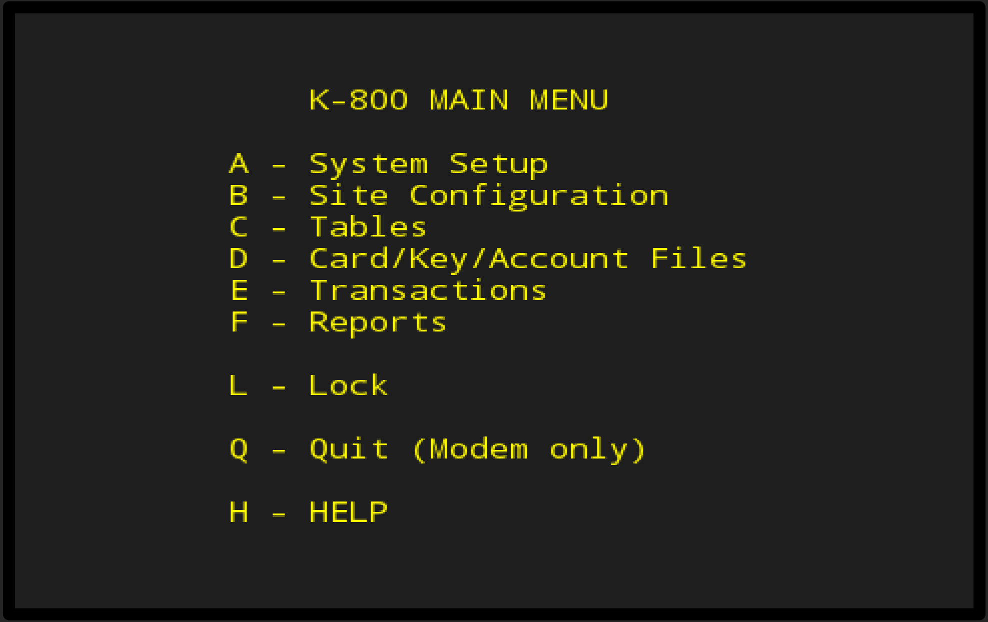
K800™ Fuel Control System

Be in control of your unattended fueling operation with Petro Vend's K800™ Fuel Control System. The K800 provides you with the tools you need to manage your fuel expenses. Fuel access is restricted to authorized users, and set to the fuel type and quantity you specify. Every transaction is tracked, giving you the security and accountability your unattended fueling operation needs.

Each system consists of the following two components:

- **1 Fuel Site Controller (FSC):** the hub of the system - stores transactions and connects peripherals
- **Up to 4 K800™ Fuel Island Terminals (FIT)** used by drivers at the island to activate the fuel dispensers

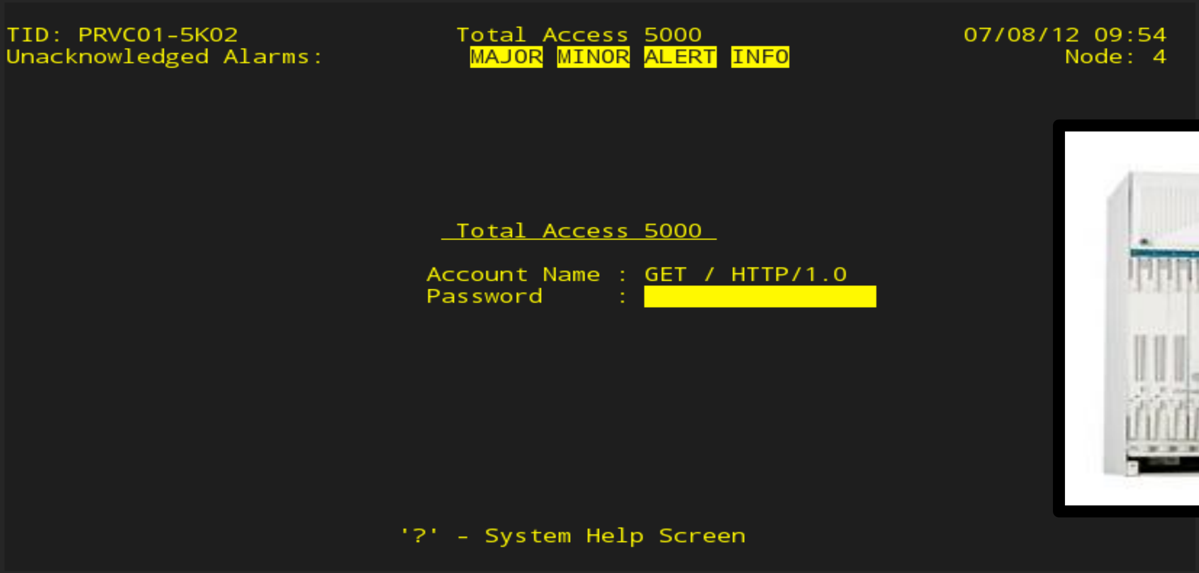
K800™ Fuel Control System



```
          K-800 MAIN MENU

      A - System Setup
      B - Site Configuration
      C - Tables
      D - Card/Key/Account Files
      E - Transactions
      F - Reports

      L - Lock

      Q - Quit (Modem only)

      H - HELP
```

Adtran IPTV Headend Systems

- Actually required authentication…
- Except when left logged-in

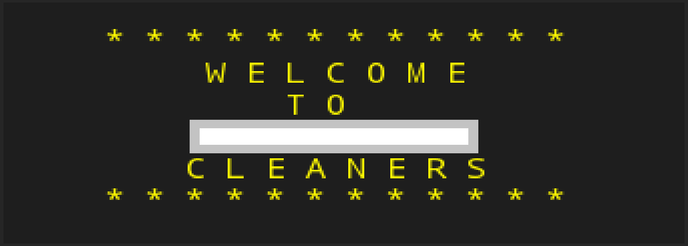


```
TID: PRVC01-5K02                Total Access 5000                  07/08/12 09:54
Unacknowledged Alarms:           MAJOR MINOR ALERT INFO                   Node: 4




                                Total Access 5000

                              Account Name  : GET / HTTP/1.0
                              Password      :




                            '?' - System Help Screen
```

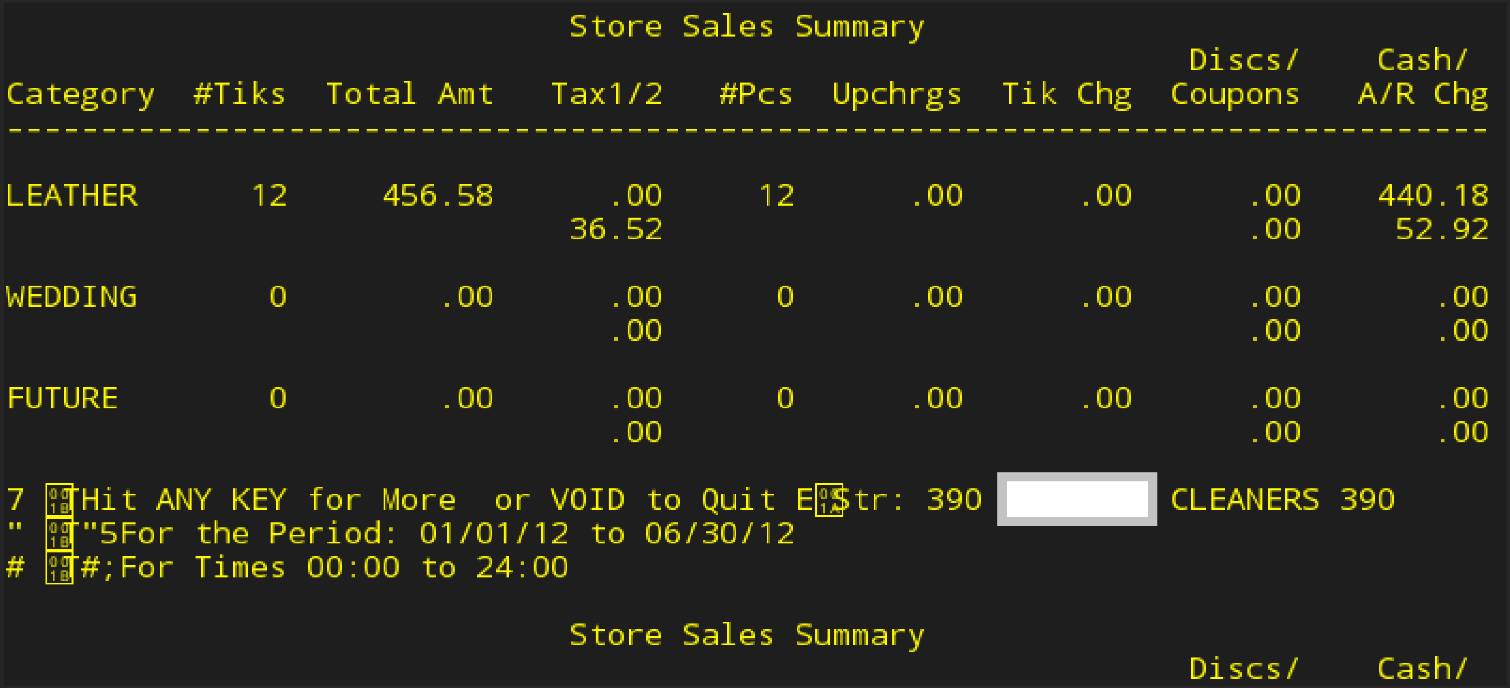
RAPID7

National Dry Cleaner Chains

- Full access to PoS systems
- No authentication



```
 *   *   *   *   *   *   *   *   *   *   *   *
          W  E  L  C  O  M  E
                   T  O
              [redacted]
          C  L  E  A  N  E  R  S
 *   *   *   *   *   *   *   *   *   *   *   *
```



```
                              Store Sales Summary
                                                               Discs/    Cash/
Category  #Tiks  Total Amt   Tax1/2   #Pcs  Upchrgs  Tik Chg  Coupons   A/R Chg
-------------------------------------------------------------------------------

LEATHER      12     456.58      .00     12      .00      .00      .00    440.18
                              36.52                               .00     52.92

WEDDING       0        .00      .00      0      .00      .00      .00       .00
                                .00                               .00       .00

FUTURE        0        .00      .00      0      .00      .00      .00       .00
                                .00                               .00       .00

7 ␛THit ANY KEY for More  or VOID to Quit E␚$tr: 390 [redacted] CLEANERS 390
" ␛T"5For the Period: 01/01/12 to 06/30/12
# ␛T#;For Times 00:00 to 24:00

                              Store Sales Summary
                                                               Discs/    Cash/
```

Please stop, I can’t take it any more…

More recent scanning - state-of-the-art

University of Michigan – Zmap – June 2013

- Research out of Michigan by Zakir Durumeric, Eric Wustrow, and J. Alex Halderman, published at USENIX
- Releasing ZMap, a purpose-built fast internet scanner
- Drastically reducing time needed to reach entire IPv4
  - 45 minutes on 1GbE (EFF SSL Observatory took 3 months)
- Comparative research on weak keys
  - ~50k weak (factorable) RSA keys, 20% decrease from 2011
  - ~3k Debian weak keys, 34% decrease from 2011

Several factors lead to interesting times...

- As shown by ZMap developers, internet-wide measurements technically feasible, far more accurate than a few years ago
  - 3 month duration has high amount of duplicates (DHCP churn)
  - 45 minutes is probably faster than necessary
- Cloud and hosting prices also reduced a lot in recent years, adding to the fact
  - Storage and computing prices now “cheap” instead of “too damn high”
- Instead of one-shot measurements and time-limited projects, continuous regular monitoring is now feasible

Traditional view on scanning

Port-scanners are evil hacking tools

- Send a SYN and wait for complaints, insults, threats
- A port-scan typically used to be the first stage of an attack
- Leads to debate, theoretical issues and technical issues

Is the Public Internet really public?

- Firewalls and authentication features exist
- If I share my documents on the internet, is it a problem if someone downloads them?

Abuse reports

Dear Mr,

Please investigate the incident described in the following partial log,
giving the treatment as your AUP permit, reporting the measures to all
recipients of this message.

In case of non acceptable treatment or reincidence, it will be taken restrictive measures to protect .BR registry.

This email is from the IT Security Team at Utah State University.

This email describes suspicious and/or malicious network activity that appears to be sourced from your network. We have included IP Addresses as well as description, documentation, log snippets, and other useful information about this event.

Please review this information and/or forward to the responsible person.

We have detected abuse from the IP address X. See below for how we obtained your email address in case it is wrong. We would appreciate if you would investigate and take action as appropriate.

** THIS IP ADDRESS IS NULL ROUTED on our entire network, including peering and transit, for a period of time not exceeding 24 hours from the date and time of this email. YOU ARE NOT REQUIRED to reply to this email unless you need more information.

You can see more information on this incident by reviewing the data at http://darknet.superb.net/ip/X and log lines are given below. Please ask if you require any further information.

Scanning considerations and best practices

- Scanning leads to abuse complaints (Darknets, traditions)
  - Own IP-space or good relationship with hoster
- High speed scanning has impact on devices on the way and the remote networks
  - Determine the necessary speed for the measurement, clarify and tune thresholds / speeds
  - 45 min scan == 1.4m packets per second (!!!)
- Q/A for the probes following the port-scan (e.g. certificate collector)

ZMap Best Practices

- https://zmap.io/documentation.html#bestpractices
- Coordinate closely with local network administrators to reduce risks and handle inquiries
- Verify that scans will not overwhelm the local network or upstream provider
- Signal the benign nature of the scans in web pages and DNS entries of the source addresses
- Clearly explain the purpose and scope of the scans in all communications
- Provide a simple means of opting out and honor requests promptly
- Conduct scans no larger or more frequent than is necessary for research objectives
- Spread scan traffic over time or source addresses when feasible
- It should go without saying that scan researchers should refrain from exploiting vulnerabilities or accessing protected resources, and should comply with any special legal requirements in their jurisdictions.

Rapid7 Mission

Gain insight into your IT landscape

- Discover assets
- Assess asset security
- Prioritize remediation

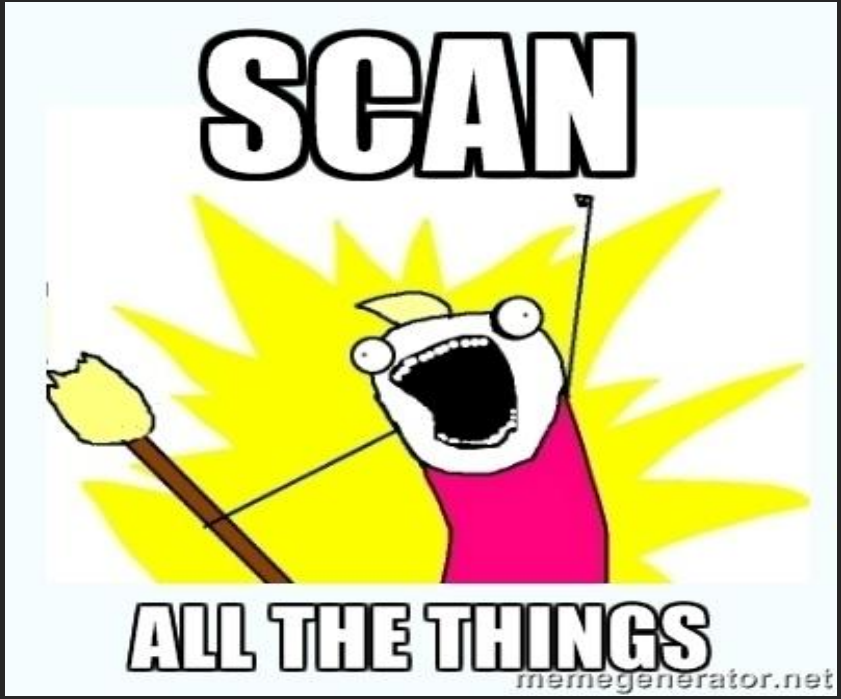
Product features

- Vulnerability scanning modules
- Device fingerprints
- Exploit modules

Finding issues and raising awareness about them is immensely valuable.

Rapid7 Labs starts *Project Sonar*

(announced by HD at Derbycon 2013)



Project Sonar

- Obtain visibility into the state of the Internet by continuously scanning and gathering data
- Compute statistics and trends
- Analyze protocols and devices
- Make data available to the community, as two pairs of eyes are always better than one

Sonar – Data overview

- 443/TCP - SSL Certificates – weekly
- 80/TCP – HTTP GET / (IP vhost) - bi-weekly
- Reverse DNS (PTR records) – bi-weekly

- Currently testing / manually done
  - Compiling name list out of scan data for…
    - HTTP GET / (name vhost)
    - Forward DNS (A records)

Sonar Architecture Details

- More than just SYN scanning
  - Stage 2 collects certificates or does HTTP GET
- Dedicated servers for ZMap / Massscan
- Cloud instances for stage 2 processing of open ports

- “autosonar” manages this process, automates cloud instances
- High-performance tools for stage 2

High-performance stage 2 tools

- Reverse DNS IPv4 - ~3.8 billion DNS lookups
- Feasible?
  - Runtime of 24 hours requires 44k lookups per second
  - Build dnsblast.c
  - Break $cloud DNS infrastructure
  - DDoS $big-ISP authoritative nameserver
  - Apologize, define thresholds, fix process (best practices)

Dealing with “big-enough-data” ™

Sonar – Data sizes and record counts

- 443/TCP - SSL Certificates – weekly
  - ~40M open ports, ~25M SSL certs, ~55GB in < 4 hours
- 80/TCP – HTTP GET / (IP vhost) - bi-weekly
  - ~70M open ports, average ~3.5Kb each, ~220GB in < 10 hours
- Reverse DNS (PTR records) – bi-weekly
  - ~1.1 Billion records, ~50GB in < 24 hours
- HTTP GET / (name vhost)
  - ~ 1.5 TB for ~200M names

- Running since November 2013 (roughly)

Big-enough-data

- More than 25M records tend to blow up even low-level key-value databases
- When inserting 1.1 billion DNS records, a few thousand records per second are just not enough
    - >100k records per second are necessary to achieve runtime in “hours”
- Given the “research nature” of the data …
    - No budget for database / indexing clusters – working from a handful of servers / VMs
    - hard to determine required indexes / search features in advance
    - lots of analysis and aggregation is done with CSV / JSON flat files and command line tools like sort / uniq / awk / jq
    - I/O throughput is key if doing lots of sequential runs on the data

Building lookup tables

- Processing tool chain in Unix-tradition

  -
    ```
    pigz -dc ${SCANBASE}-linkmap.gz | pv | awk -F, '{ $0=tolower($0); print
    $2","$1","$3 }' | uniq | python $AUTOSONAR/scripts/reverselevel.py | pigz -c
    > ${SCANBASE}-csv-for-leveldb.gz
    ```
- Define search functionality, insert raw data in key-value stores (e.g. LevelDB)
  - Will survive even billions of records and compresses by default (Snappy)
- Extracted fields in SQL (Postgres)
  - Parse certificates and store unique certificates in SQL

Enterprise software for research data?

- Using Ruby/Python/etc is necessary, writing custom C tools too time consuming
  - Often quick and dirty tools chewing on the data
  - JIT interpreters to the rescue (huge thanks to PyPy!)
- Quickly apparent that these scripts/tools pile up
  - Coming up with a data processing framework / pipeline

  -
    ```
    parallel --gnu -j 7 -L 50000 --nice 40 --pipe “ruby
    ${DAP}/bin/dap json + transform data=base64decode + html_links
    data + select vhost ip link element + decode_uri link + json"
    ```

Sonar Lookup databases

- Out of the data we build some interesting lookup indexes
  - Suffix searches on hostnames / domains
    - All DNS records for *.rapid7.com
    - Certificates with *.rapid7.com
  - IP and name history knowledge
    - Which names point to a certain IP?
    - At a certain time, who linked to domain X?

Case studies of internet-scanning profit

Continuous SSL Observatory

- We’re building processing pipelines to have a continuous output of metrics similar to the EFF SSL Observatory
- Still seeing ~50k certificates with weak keys
- High amount of certificates expired
- Revoked certificates still in use (exact numbers still to be released)

Prioritizing vulnerability research

- Looking at HTTP scan data, a widely used router is TD-W8901G
- Buy the device from Amazon and start poking
- ZyXEL ZynOS monolithic kernel
- In this case we failed, but shortly after a researcher published that the “rom-0” backup file could be downloaded without authentication – and that it contains the admin password to the router in plain

Measuring exposure to vulnerabilities

- Using our infrastructure we measured the amount of publicly reachable devices for both disclosed and not-yet-public vulnerabilities
  - Found ~7k routers with SERCOMM backdoor
  - Found 2 Yokogawa ICS controllers in University ranges
- Allows raising awareness, determining threat level
- Notifying exposed institutions

Continuous defacement statistics

- Check scan data for defacement traces

- Top 10 Title Tags containing "Hacked":

1. 187 Hacked by H4x0r HuSsY
2. 70 Hacked By GHoST61
3. 42 Hacked By International Force
4. 33 Hacked By TechnicaL
5. 29 Hacked By TechnicaL
6. 17 hacked by GHoST61
7. 16 Hacked By Kam_06- http://wWw.BlackWorm.Org
8. 12 Hacked by GUARD_FB
9. 10 hacked by brkod ..!
10. 10 HACKED BY 1ND14N CYB3R R4K$H4K

Detecting "odd devices"

- We noticed ~1.5M devices were responding to all TCP connection requests, independent of the port
- Effectively skewing our results, especially for not widely spread services (Sercomm backdoor)
- Built “blacklists” of the respective IPs / Subnets by scanning on random unused ports and correlating
- Current theory is some sort of black-hole behavior in certain firewall / IDS devices

Finding parser bugs

- "Internet data" is stressful for parsers and format libraries
- Found Ruby X509 Certificate parsing bug
  - Occurred on weird certs that had a PEM-encoded certificate in an X509 extension comment
- Currently being worked on upstream
- Potentially useful for SSL man-in-the-middle attacks

› What’s next?

Scan more things

- Extend certificate gathering
  - Increase the SSL certificate dataset coverage
  - Get certs from other SSL and TLS services (SMTP/IMAP/etc)
- Add other widespread services
  - Requires additional specific stage 2 tools

Build reporting / Sonar-web

- In addition to raw datasets, provide public lookup tools
- Publish statistics, trending
- Allow for easier collaboration

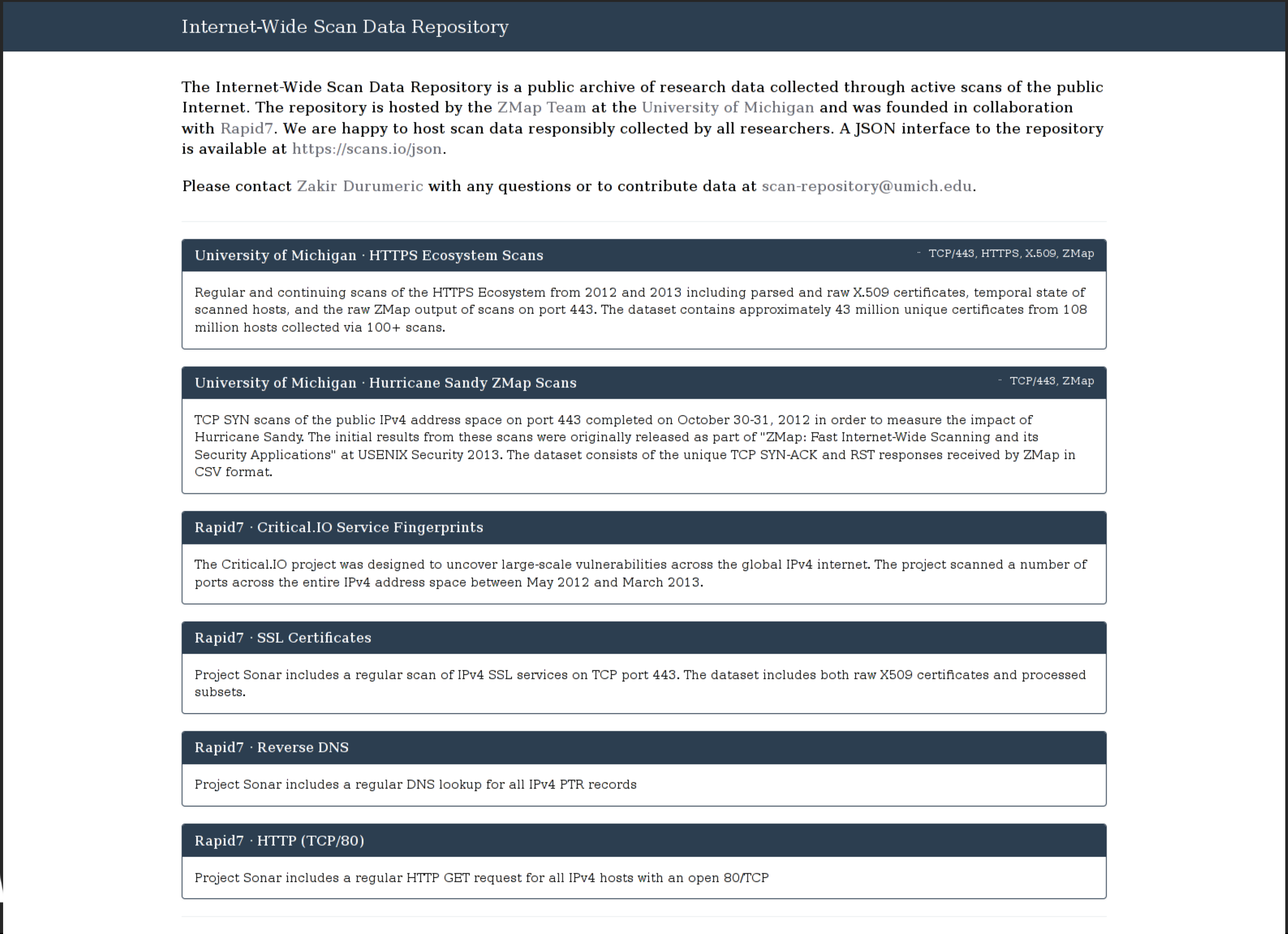
Analyzing results

- Aggregate by industry / network type
  - Specific service distribution at cloud providers?
  - Particular devices in certain industries?
  - Consumer vs enterprise endpoints
- Find more misconfigurations / vulnerabilities and get them reported, fixed - raise awareness
- Monitor decline of vulnerabilities, report on "security progress"

Collaboration is highly important

- Make data available to the Security community
  - Collaboration with University of Michigan
  - Raw Scan data published at http://scans.io/
- Historical upload (critical.io, Michigan data)
- Almost-real-time upload of raw scan output

http://scans.io/



Internet-Wide Scan Data Repository

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.

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.

Rapid7 · Critical.IO Service Fingerprints

The Critical.IO project was designed to uncover large-scale vulnerabilities across the global IPv4 internet. The project scanned a number of ports across the entire IPv4 address space between May 2012 and March 2013.

Rapid7 · SSL Certificates

Project Sonar includes a regular scan of IPv4 SSL services on TCP port 443. The dataset includes both raw X509 certificates and processed subsets.

Rapid7 · Reverse DNS

Project Sonar includes a regular DNS lookup for all IPv4 PTR records

Rapid7 · HTTP (TCP/80)

Project Sonar includes a regular HTTP GET request for all IPv4 hosts with an open 80/TCP

› Wrapping up

The Internet is broken.

- Widespread bugs, vulnerabilities, misconfigurations
- Weak credentials
- Lost and forgotten devices, embedded hardware piling up without update possibilities
- We’re not improving the overall “state of security”

Moving forward

- Can’t stress enough the importance of awareness and visibility
- Internet scanning is a powerful tool that can do a lot of good for the community
  - We need to offset some of the traditional views and work together to define thresholds and best practices
- Collaboration is essential for data collection and analysis

How Internet Scanning helps

- Identify and quantify widespread vulnerabilites
- Measure improvements on disclosed flaws continuously
- Identify risks (amplification attacks, misconfigurations) before they are misused
- Drive research based on data, build awareness around public networks
- Hold vendors and ISPs accountable

Make sure to also check out

- ZMap at http://zmap.io/
- J. Alex Halderman on "*Fast Internet-wide Scanning and its Security Applications*" at 30C3 (Germany)
- HD Moore's keynote "*Scanning Darkly*" at Derbycon 2013
- http://sonar.labs.rapid7.com/

- Internet is broken!
- Internet scanning saves lives!
- Sharing is caring!

RAPID7

Thanks!

Mark Schloesser

mark_schloesser@rapid7.com

@repmovsb