

Port scan is not for pussies

Know yourself, know your enemy

Fred RAYNAL

fraynal@quarkslab.com

@fredraynal

Adrien GUINET

aguinet@quarkslab.com

@adriengnt



How did it start?

I want to scan the Internet!!!

- Scan for obscure web forums to gather versions of phpBB, vBulletin and others
- Scan for card sharing servers
- Get carrot juice, a veggie burger and some sleep
- Idea: **scan for everything everywhere**
- Internet Census (2012): well played, f*****g Carna Botnet

Why do we care about network recon?

Motivation

- For attackers: information is as valuable as 0 days
 - Allow to build the attack path
 - Avoid wasting 0 days
 - Find opportunistic targets
- For defenders: learn about yourself
 - *Should* allow to learn about their own attack surface
 - *Should* guide to concentrate defenses where one is the most exposed and sensitive



Roadmap

This talk

- Engineering: how to design an Internet wide scanner
- Targeting: what is a target?
- Applications: what we find on the Internet



Plan

- 1 I need an engineer
 - Overview
 - Defuse mines: why port scan is not for pussies
 - Scalability: I need a medic
 - Optimization: I REALLY need an engineer
 - Another step with libleeloo and nodescan

- 2 Targeting: snipe or mass destruction?

- 3 What can be done / found on the Internet



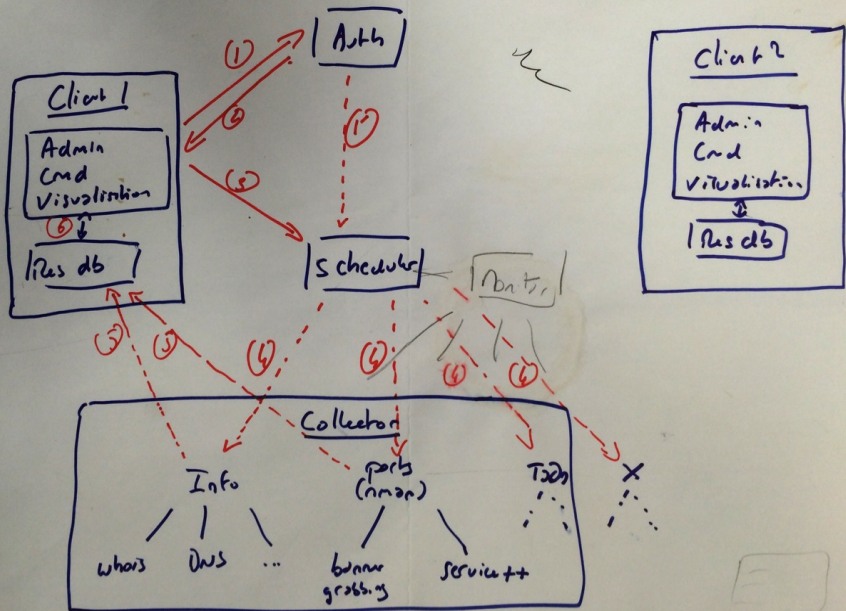
Plan

- 1 I need an engineer
 - Overview
 - Defuse mines: why port scan is not for pussies
 - Scalability: I need a medic
 - Optimization: I REALLY need an engineer
 - Another step with liblibleeloo and nodescan

- 2 Targeting: snipe or mass destruction?

- 3 What can be done / found on the Internet







I LOVE IT WHEN A PLAN COMES TOGETHER.

Plan

- 1 **I need an engineer**
 - Overview
 - Defuse mines: why port scan is not for pussies
 - **Scalability: I need a medic**
 - Optimization: I REALLY need an engineer
 - Another step with liblibleeloo and nodescan

- 2 **Targeting: snipe or mass destruction?**

- 3 **What can be done / found on the Internet**



At first, we had nmap

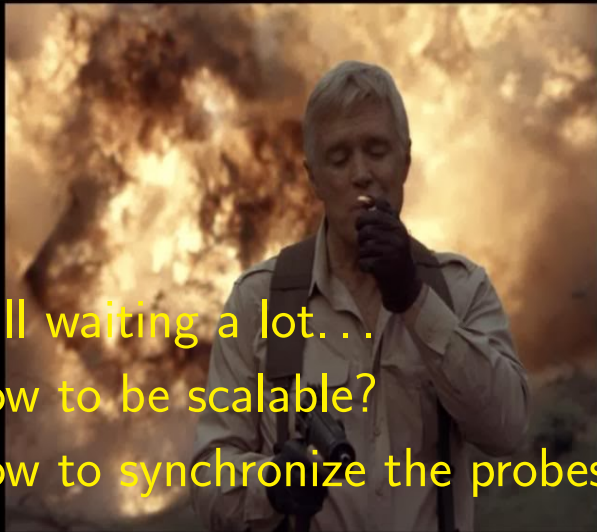
Pros

- Stable and widely used
- Powerful NSE scripts engine
- Correctly fast with good timing options

Cons

- Runs on a single host
- Can not add target on the fly (even with -iL -)





- Still waiting a lot...
- How to be scalable?
- How to synchronize the probes?

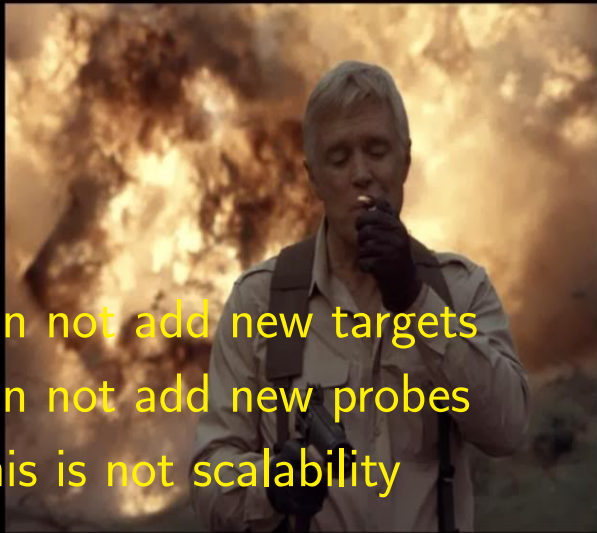
I LOVE IT WHEN A PLAN COMES TOGETHER.

Becoming scalable: a first try

Examples with 3 probes

- Divide the target set in 3
- Give each host a third of the target space
- Collect the results from the probes





- Can not add new targets
 - Can not add new probes
- ⇒ This is not scalability

I LOVE IT WHEN A PLAN COMES TOGETHER.

Becoming scalable: the plan B

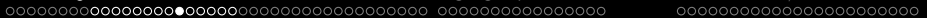
Being scalable

- Divide the target set in fixed-size randomized blocks of IPs/blocks
- Create a queue of tasks to perform
- Send them to your probes on-demand

Scalability 101: what we need

- A message passing protocol (rabbitmq, mpi, ...) to give orders and get back the results
- A scanner (nmap for now)
- Something to keep track of what's been done





Becoming scalable: the plan B (what we need)

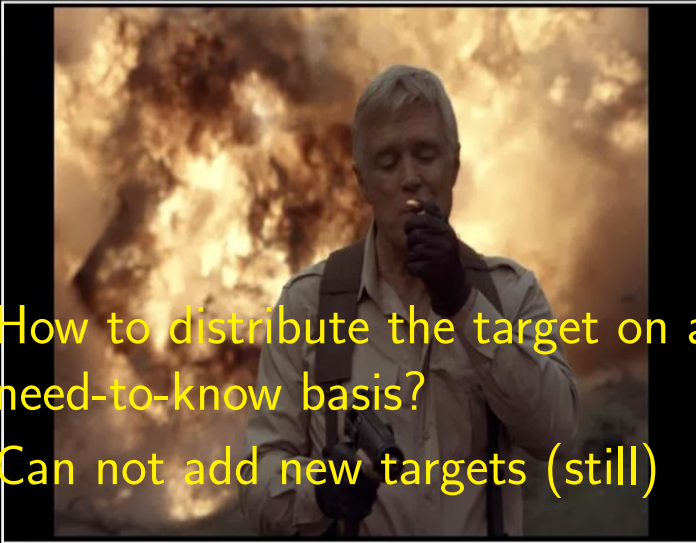
Another piece of cake

- A library that randomize the target set
- AMQP for the task management and tracking

Extra-bonus

- Probes are on a need-to-know basics
- New probes can be added on the fly, they just grab new tasks
- Probes can get away without ACKing a task, it will be performed by a new one





- How to distribute the target on a need-to-know basis?
- Can not add new targets (still)

I LOVE IT WHEN A PLAN COMES TOGETHER.

Splitting the targets

What is a target?

- A target is a union / exclusion of intervals of IP addresses

Naive algorithm

- Create a list of all unique IP addresses
- Randomize the set to avoid consecutive scanning (thus complains)





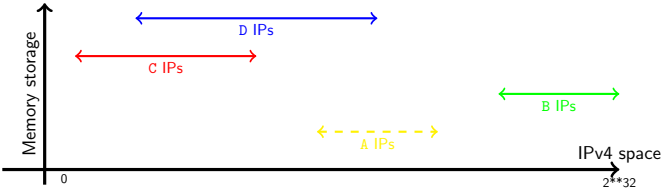
2^{32} IP * 4 bytes = 16Gb in RAM :(

I LOVE IT WHEN A PLAN COMES TOGETHER.

Splitting the targets with a PRNG

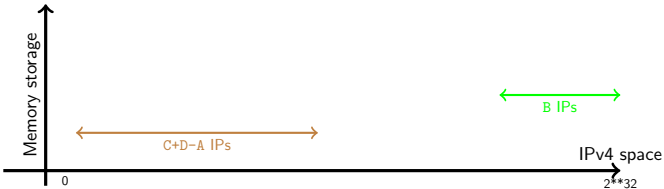
Step 1: initial configuration

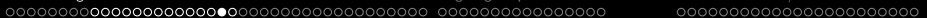
- Wanted ranges are the full lines
- Excluded ranges are the dashed lines



Splitting the targets with a PRNG

Step 2: sorting and merging intervals

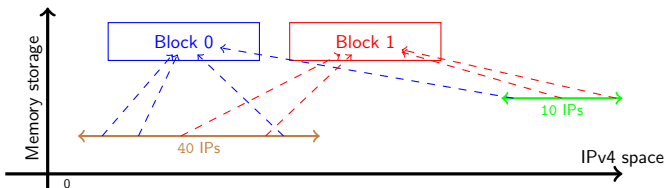




Splitting the targets with a PRNG

Randomization

- There are N ($=C+D+B-A$) IPs among R ($=2$) distinct ranges
- Compute a random permutation of $[0..N]$
- For each integer i of this permutation, grab the IP at the i -th index
- Create blocks of G ($=4$ for instance) randomly chosen IPs and send them to the probes
- An example: $[30, 10, 5, 42, 20, 28, 48, 49, \dots]$





This is where we are now thanks to
distribution of the scan

⇒ Time for optimization!

Plan

- 1 I need an engineer
 - Overview
 - Defuse mines: why port scan is not for pussies
 - Scalability: I need a medic
 - Optimization: I REALLY need an engineer
 - Another step with libleeloo and nodescan
- 2 Targeting: snipe or mass destruction?
- 3 What can be done / found on the Internet



Optimization: upgrade the scanner

zmap

- Asynchronous I/O engine for the packets
- Can share a target on several hosts
- Can not add probes dynamically
- Can not add targets on the fly
- Scripting is a pain
- Requires a Telco for a maximum efficiency

masscan

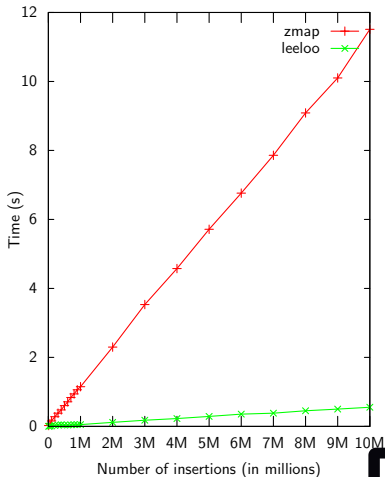
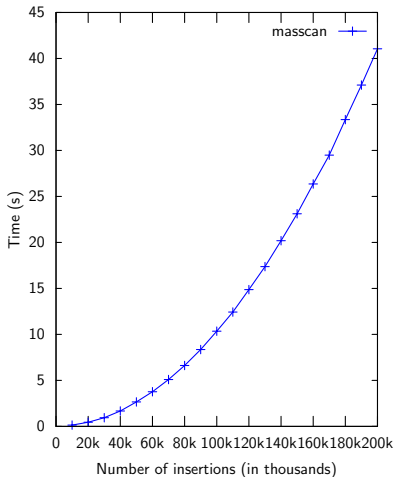
- Asynchronous I/O engine for the packets
- Can share a target on several hosts
- Can not add probes dynamically
- Can not add targets on the fly
- Scripting is a pain++
- Requires a Telco for a maximum efficiency





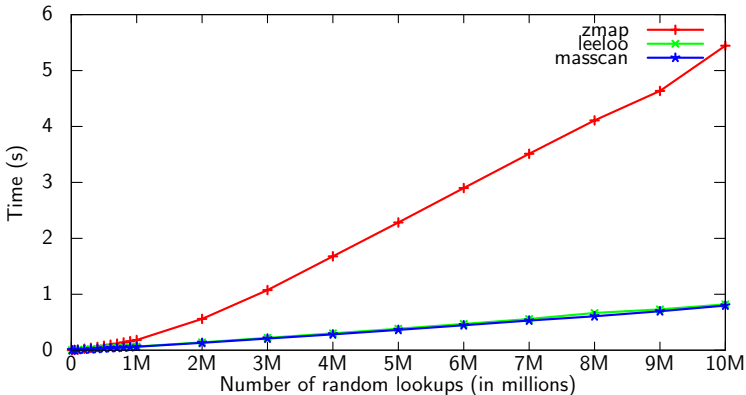
IP intervals management: intervals add performances

Benchmarks done on a Core i7-3520M





IP intervals management: random lookup performances



Lookup performances

- **libleeloo** and **masscan** can provide about **12,204,000** random lookups/second



IP intervals management

zmap

- **Model:** intervals stored as a tree (lower memory usage), only support CIDR ranges
- **Add:** logarithmic complexity since the tree is balanced
- **Lookup:** complexity depending on the height of the tree

masscan

libleelo



IP intervals management

zmap

masscan

- **Model:** list of intervals stored as pairs of uint32 in an array
- **Add:** exponential complexity since checking the new one is not already in a former one
- **Lookup:** logarithmic by using a pre-computed cache (non configurable size)

libleelo



IP intervals management

zmap

masscan

libleelo

- **Model:** same as masscan
- **Add:** just add the new intervals in the array, aggregate once at the end
- **Lookup:** logarithmic, also using a cache of configurable size (user-defined memory/performance trade-off)



The scanner of our dreams

What we dream of?

- SYN engine as efficient as masscan
- Scripting as easy as nmap
- Can run as a daemon to stream targets as they come

Patching nmap / zmap / masscan? You have said patching?

- Need to change core components, not maintainable on a long run
- Can not support properties for IPs
- Can not support complex scan actions at layer 7



Plan

- 1 I need an engineer
 - Overview
 - Defuse mines: why port scan is not for pussies
 - Scalability: I need a medic
 - Optimization: I REALLY need an engineer
 - Another step with libleeloo and nodescan

- 2 Targeting: snipe or mass destruction?

- 3 What can be done / found on the Internet



libleeloo: intervals and *properties*

libleeloo

- A **C++ library** with Python bindings
- Manage intervals of IPs as seen previously
- Support *properties*
- Available at <https://github.com/quarkslab/libleeloo>

Properties?

- Specific information for some IPs or ranges
- Custom TCP/UDP ports, specific credentials to test, ...





Example: using multiple properties to IPs intervals

```
1 import pyleeloo
2 ranges = pyleeloo.ip_list_intervals_with_properties()
3
4 # The organisation's range
5 ranges.add("192.42.0.0/16")
6
7 # SSH servers
8 ranges.add_property("192.42.4.0/24", [22, 2222])
9 # VPN servers
10 ranges.add_property("192.42.4.10-20", [1194])
11
12 ranges.aggregate()
13
14 def merge_ports(portsA, portsB):
15     portsA.extend(portsB)
16     ranges.aggregate_properties(merge_ports)
17
18 print(ranges.property_of("192.42.66.0"))
19 >>> None
20
21 print(ranges.property_of("192.42.4.1"))
22 >>> [22, 2222]
23
24 print(ranges.property_of("192.42.4.15"))
25 >>> [22, 2222, 1194]
```



Nodescan

A L7 asynchronous engine

- A **C++ library** with Python bindings to build a custom L7 scanner
- L7 Python scripting à la nodejs with callback definitions
- Support scan pause and resume
- Allow complex actions like in SSL, SSH, SIP, . . .
- Built on asynchronous UNIX sockets (for now)
- Beta on <https://github.com/quarkslab/nodescan>

Scanning L7 with nodescan: architecture

Targets definitions

Level 4: IP/port

Level 7 processing

User defined processing callbacks (plugins)

Reinject new targets

Scan engine

Check targets availability, call user-defined callbacks and process timeouts



Scanning L7 with nodescan by example

By specifying a list of (IP, port) pairs

```
1 targets = pynodescan.SimpleTargetSet()  
2 targets.add_target("37.187.47.70", tcp_port(80));  
3 targets.add_target("173.194.40.134", tcp_port(22));
```



Scanning L7 with nodescan: architecture

Targets definitions

Level 4: IP/port

Level 7 processing

User defined processing callbacks (plugins)

Reinject new targets

Scan engine

Check targets availability, call user-defined callbacks and process timeouts



Scanning L7 with nodescan by example

Simple LVL4 connection to build a HTTP scrapper

```
1 def send_payload(target, lvl4sm, hsm):
2     # Send GET /
3     target.send("GET_/_HTTP/1.0\n\n")
4     # Trigger on newlines
5     lvl4sm.set_char_data_trigger('\n', on_newline)
6
7     # returns True to go on with this target
8     return True
9
10 def on_newline(target, lvl4sm, hsm, buf):
11     with open("res/%d" % target.ipv4(), "ab") as f:
12         f.write(buf.tobytes())
13     return True
14
15 engine.set_lvl4_connected_callback(send_payload)
```



Scanning L7 with nodedscan by example

Getting to level 7...

- Classes that wrap level 7 protocols
- Provides specific callbacks: on_content, on_certificate, ...
- User just defines what to do on each event
- Currently supports HTTP, SSH and SSL public key/certificate grabbing and SIP headers



Scanning L7 with nodescan by example

Same with HTTP wrapper

```

1  def write_header(target, key, value):
2      with open("res/%d" % target.ipv4(), "wb+") as f:
3          f.write("%s: %s\n", (key, value))
4
5  def write_content(target, code, content):
6      with open("res/%d" % target.ipv4(), "wb+") as f:
7          f.write(content.tobytes())
8
9  HTTPGrabber =
10     pynodescan.protocols.HTTPMethod("GET", "/", {"User-agent": "
        pony_1.0"})
11     .on_header(write_header)
12     .on_content(write_content)
13     .on_error(lambda target, err: print((target, err), file=sys
        .stderr))
14 )
15 engine.set_lvl4_connected_callback(HTTPGrabber)

```



Nodescan: you have just seen the scripting

Targets definitions

Level 4: IP/port

Reinject new targets

Level 7 processing

User defined processing callbacks (plugins)

Scan engine

Check targets availability, call user-defined callbacks and process timeouts



Engineering conclusion

- Scanning large sets of IPs is not only about sending raw SYN packets
- Especially if you want to do that dynamically (adding targets or probes)
- Especially if you want to collect data at layer 7 and react accordingly



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
 - What is a *target*?
 - Targeting subdomain *.gouv.fr
 - Retrieving the reverse whois database
 - Domain scrapping
- 3 What can be done / found on the Internet



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
 - What is a *target*?
 - Targeting subdomain *.gouv.fr
 - Retrieving the reverse whois database
 - Domain scrapping
- 3 What can be done / found on the Internet



Targeting a country

Country acquisition

- Based on GeolP
- Outsource the problem of figuring it out
- Misses some DNS names hosted overseas
- Simplify the jurisdictional issues

Country	GeolP	whois	GeolP \cup whois
France	79M	75M	97M
Spain	29M	16M	30M



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
 - What is a *target*?
 - Targeting subdomain *.gouv.fr
 - Retrieving the reverse whois database
 - Domain scrapping
- 3 What can be done / found on the Internet



Targeting *.gouv.fr howto

Algorithm

- Find as much domains ending with *.gouv.fr as possible
- For each domain:
 - Get the corresponding IP
 - Get the whois associated to the IP
 - Consider the netrange the IP belongs to^a

a. Assumes a hosting company might host several IPs related to *.gouv.fr

Problems / subgoals

- #1: get a whois database, which is a pain to parse
- #2: get domains from Google / Bing / other which do not want to be scrapped

Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
 - What is a *target*?
 - Targeting subdomain *.gouv.fr
 - Retrieving the reverse whois database
 - Domain scrapping
- 3 What can be done / found on the Internet



whois issue: build your own reverse whois cache

Accessing whois database

- Formerly available at ipindex.homelinux.net but domain is dead now
- Bulk access to whois data has to be asked for each registrar
 - And you have to send a letter to APNIC (so 2014)

What we just need: reverse whois database

- Goal: for each IP, know to what netblock it belongs to, and who owns this netblock
- Ex.: who owns 42.0.0.0/8, 42.0.0.0/16, 42.0.0.0/24 and any potential subnetwork





WHAT THE FUCK

DID I JUST SEE ?

Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
 - What is a *target*?
 - Targeting subdomain *.gouv.fr
 - Retrieving the reverse whois database
 - Domain scrapping
- 3 What can be done / found on the Internet



filetype:pdf inurl:gouv.fr "ne pas diffuser"



Web

Shopping

Vidéos

Images

Actualités

Plus ▾

Outils de recherche

Environ 1 150 résultats (0,13 secondes)

^[PDF] [Document provisoire, ne pas diffuser - Haut Conseil de l...](#)

www.sante-jeunesse-sports.gouv.fr/IMG/pdf/r_mt_300905_vhb5.pdf ▾

de D ANTONA - [Autres articles](#)

RAPPORT DU GROUPE DE TRAVAIL du Conseil supérieur d'hygiène publique de France. Risque de contamination horizontale au sein de collectivité d' ...

^[PDF] [Guide du bon usage des médias sociaux - Ministère de la ...](#)

www.defense.gouv.fr/guide-medias-sociaux/telecharger.pdf ▾

à **ne pas diffuser**. • Éviter les publications, statuts ou commentaires tels que : « Super ! Plus que 11 jours et 2 heures et vous serez à quai et je pourrai enfin te ...

Getting domains (plan B): using the *cloud*

Wait a second. . .

- We have a scalable architecture
- We have France == 97M IPs (GeoIP + whois)
- We have libeleloo to distribute these 97M IPs over our probes



Getting domains (plan B): using the *cloud*

Wait a second. . .

- We have a scalable architecture
 - We have France == 97M IPs (GeoIP + whois)
 - We have libeleloo to distribute these 97M IPs over our probes
- ⇒ Let's distribute the 97M DNS lookups !!

Results

- Duration: 15h
- Hosts: 5
- Unique domains found: 1342
- Unique IPs: 1295
- Subdomains: 143
- Network size: 7M IPs



Conclusion: targeting *.gouv.fr at cloud age

Finding targets

```

1 def domains2IP( hostnames, patter ):
2     domains = hostnames.grep( pattern ) # 1342 domains
3     targets = []
4     for d in domains:
5         ip = gethostbyaddr( d )
6         targets += net.add( whois.get_range( ip ) )
7     return targets

```



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
- 3 What can be done / found on the Internet
 - Vulnerability research
 - Scanning Spain
 - Diffing networks
 - Usage monitoring



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
- 3 What can be done / found on the Internet
 - Vulnerability research
 - Scanning Spain
 - Diffing networks
 - Usage monitoring



A quick word about heartbleed

- Many scans looking for vulnerable servers. . .
- Most of the focus is on 443 port
- Free advice: people should also look at OpenVPN and some other servers



Looking for a backdoor

I'm gonna owned the Internet

- Backdoor discovered (twice :) by Eloi Vanderbeken on some routers
- Listen on TCP port 32764
- No authentication, simple protocol
- Let's start some recognition...



How to own the Internet

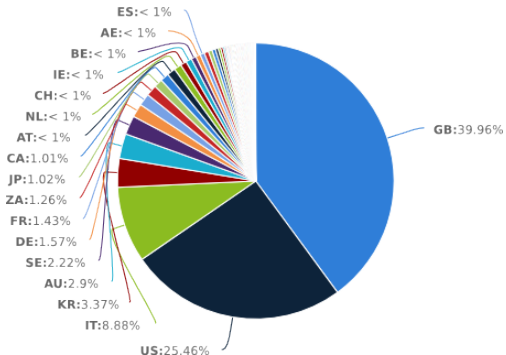
My precious

- Launch masscan on 32764: 30k packets/s
- around 50h later, about 1 million IPs discovered with TCP port 32764 open
- Used nodedscan to verify these hosts: checking for backdoor signature as an answer of an invalid request
- By scanning about 6k IPs/s, a few minutes later, about 6000 devices were found vulnerable



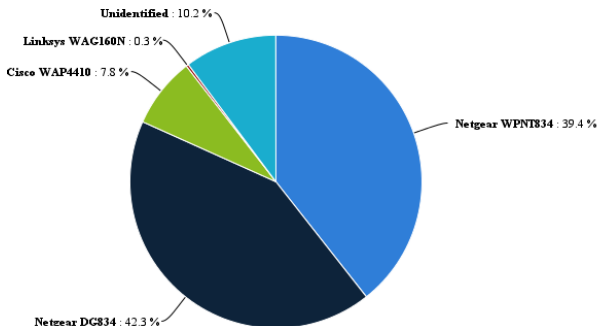
Gathering statistics about the backdoor

- Repartition by country



Gathering statistics about the backdoor

- Repartition by hardware: using the same scanner, a "version" and "sys_desc" field has been grabbed. Manual mapping had to be done (thus the "Unidentified" field).



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
- 3 What can be done / found on the Internet
 - Vulnerability research
 - Scanning Spain
 - Diffing networks
 - Usage monitoring



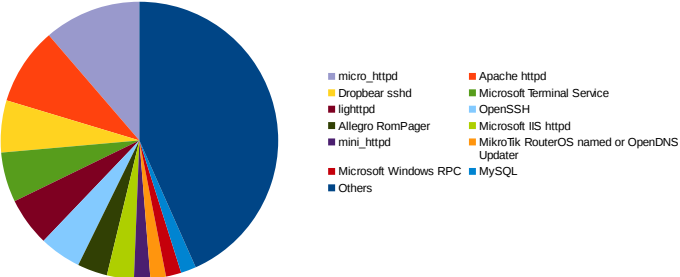
#define Spain

What is Spain?

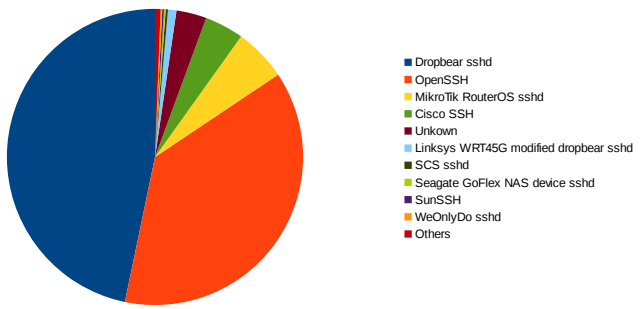
- Country: 30M IPs
- Number of probes: 100
- Number of ports: 30
- Plugins: banners for Telnet & FTP, SSL certificate, SSH key, HTTP (index of, page title, headers, auth), heartbleed, NFS.ls, FTP.ls, MySQL info, hadoop,...
- Scan duration: 25h



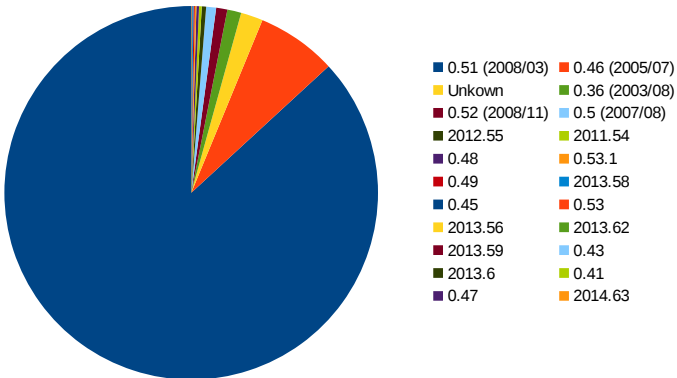
What does Internet.es look like?



Focus on SSH



Digging into dropbear

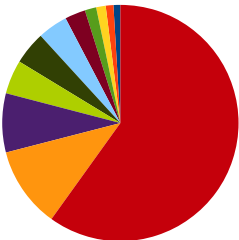


37.152.157.114.txt: OLECOMUNICACION-NET 84.232.91.100.txt: THELLINM-NET
 37.61.251.139.txt: NUBBITEL 84.236.239.171.txt: ADSLSERVICES-ADSL-NET
 62.81.244.73.txt: INFOTEC_TECNOLOGIA_INTEGRAL_Y_TELEC 87.235.106.48.txt: IPCOM-NET
 77.27.81.19.txt: FUNDACIONCULTURALESTRADA-NET 89.140.120.114.txt: INFORMATICA_LIMON



FTP at a glance

- FTP banners: 31959
- `grep -i camera ftps|wc -l` → 216
- `grep -i "DSL router" ftps|wc -l` → 2110



- ProFTPD 1.3.3c Server ready
- FTP Server Ready
- ucftpd(Jul 2 2012-2213:49) FTP server ready
- DiskStation FTP server ready
- FTP server ready 1 active clients of 1 simultaneous clients allowed
- batman FTP server (GNU inetutils 1.3.2) ready
- FTP server ready
- (none) FTP server (GNU inetutils 1.4.1) ready
- FTP Server ready
- DSL Router FTP Server v00.96.114 ready
- MikroTik



MikroTik FTP?

- Actually all FTP banners containing MikroTik are unique

```
LOURDES GARCIA LANDETE FTP server (MikroTik 5.11) ready: 1  
Nodo Formentera 2 V + H FTP server (MikroTik 5.25) ready: 1  
AYTO_SCOLA_MUSICA FTP server (MikroTik 5.25) ready: 1  
Cliente Danubio27 - Francisco Planells FTP server (MikroTik 5.19) ready: 1  
M26002512T FTP server (MikroTik 5.22) ready: 1  
SJVJCostaRd1 FTP server (MikroTik 5.22) ready: 1  
ramon lopez perez FTP server (MikroTik 5.21) ready: 1
```



Long tail of Internet.es (a.k.a. wtf.es)

- 3M Filtrete 3M-50 thermostat: thermostat with WiFi control... on the Internet
- <http://www.radiothermostat.com/filtrete/products/3M-50/>





Long tail of Internet.es (a.k.a. wtf.es)

- merten@home: remote for everything at home



Long tail of Internet.es (a.k.a. wtf.es)

- merten@home: awarded in 2004 and 2006 !!

MERTEN@HOME

SYSTEM DATA AND VERSION DETAILS

The device is currently showing the following equipment and versions

Hardware version: 0001-0101-008
RAM memory: 16
ROM memory: 4

Integrated modem: 1
USB devices: 1

Firmware version: 02.32
Firmware date: 2006-01-24
Interface version: 1.02

[▶ Return to homepage \(login\)](#)



Long tail of Internet.es (a.k.a. wtf.es)

- Moxa NPort 5410: serial to IP converter for PLC, industrial systems, ...



Long tail of Internet.es (a.k.a. wtf.es)

- Cameras of course: Axis



Plan

- ① I need an engineer
- ② Targeting: snipe or mass destruction?
- ③ What can be done / found on the Internet
 - Vulnerability research
 - Scanning Spain
 - Diffing networks
 - Usage monitoring





Monitoring == diffing

IVY Targets Collectors Scans Reports Reports comparison

REPORTS DIFFERENCES

Base report

< 1/1 >

brp

Name	Creation Date	Action
top 100	Dec 12, 2013	Select
top 100	Dec 5, 2013	Select
top 100	Nov 21, 2013	Select
top 100	Nov 7, 2013	Select
top 100	Oct 31, 2013	Select
top 100	Oct 24, 2013	Select
top 100	Oct 7, 2013	Select
top 100	Sep 30, 2013	Select
top 100	Sep 27, 2013	Select
top 100	Sep 27, 2013	Select

Compare against

< 1/1 >

Search by name

Name	Creation Date	Action
top 100	Dec 12, 2013	Select
op 100	Dec 5, 2013	Select
op 100	Nov 21, 2013	Base report
op 100	Nov 7, 2013	Select
op 100	Oct 31, 2013	Select
op 100	Oct 24, 2013	Select
op 100	Oct 7, 2013	Select
op 100	Sep 30, 2013	Select
op 100	Sep 27, 2013	Select



Plan

- 1 I need an engineer
- 2 Targeting: snipe or mass destruction?
- 3 What can be done / found on the Internet
 - Vulnerability research
 - Scanning Spain
 - Diffing networks
 - Usage monitoring



PayTV Internet Sharing

CCcam

- One host (master) shares a card with several clients
- When one client receives an encrypted payload, it is sent to the master
- The master deciphers the payload, sends it back to the client
- Very lucrative business

Usage statistics

- Scan a few ports, the usual ones where CCcam is running
- Connect to the server to get plenty of information





Piracy monitoring

DASHBOARD

Summary

DETAILS

- [View by ip](#)
- [View by port](#)
- [View by plugin](#)

PLUGINS

cccam-info

GRAPHICS

[View history](#)

CCCAM FINAL

KEY NUMBERS

Scanned targets : 65

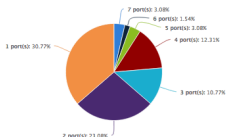
Opened targets : 65 (100%)

Total opened ports : 133

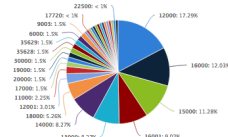
Max/Avg opened ports : 7 / 2

DISTRIBUTION BY PORT NUMBER, COUNTRY AND TLD

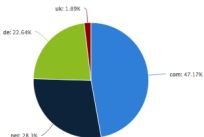
Targets repartition by opened ports
Graph built the 21/01/2014



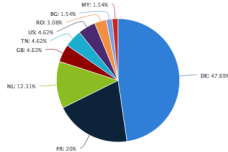
Targets repartition by ports
Graph built the 21/01/2014



Targets repartition by TLD
Graph built the 21/01/2014



Targets repartition by country
Graph built the 21/01/2014





Piracy monitoring

IVY Targets Collectors Scans **Reports** Reports comparison

Back

Summary

PLUGINS DATA

CCCam informations

CCCam Authentication

IP HISTORY

Browse history

IP: [REDACTED]

Scanned at : Nov 6, 2013 - 6:41:01 PM

HISTORY

NUMBER OF OPENED PORTS

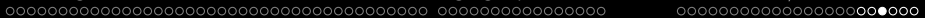
REVERSE DNS

Update

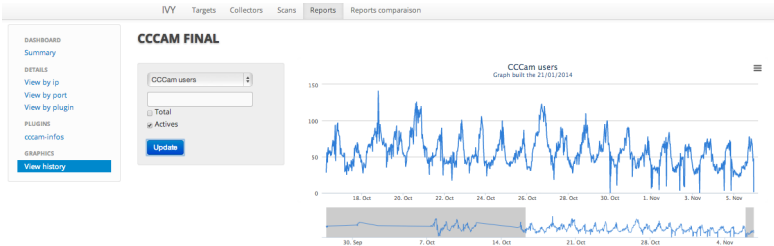
EVOLUTION

GRAPH BUILT THE 21/01/2014



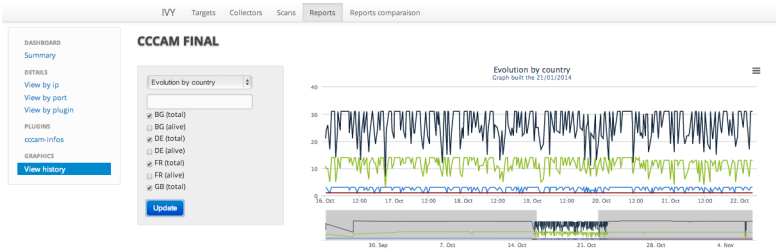


Piracy monitoring





Piracy monitoring



Conclusion

What massive port scan is good for?

- Security is not about patching anymore
 - Try to prevent the attack (ID,PS, exploit mitigation, AV, ...)
 - Assume the attack will succeed anyway :(
- ⇒ Need to know what / where your assets are
- To elaborate your defensive strategy
 - To elaborate your recovery plan



Questions?

Challenge accepted:klapspaan



www.quarkslab.com

contact@quarkslab.com | [@quarkslab.com](https://twitter.com/quarkslab)