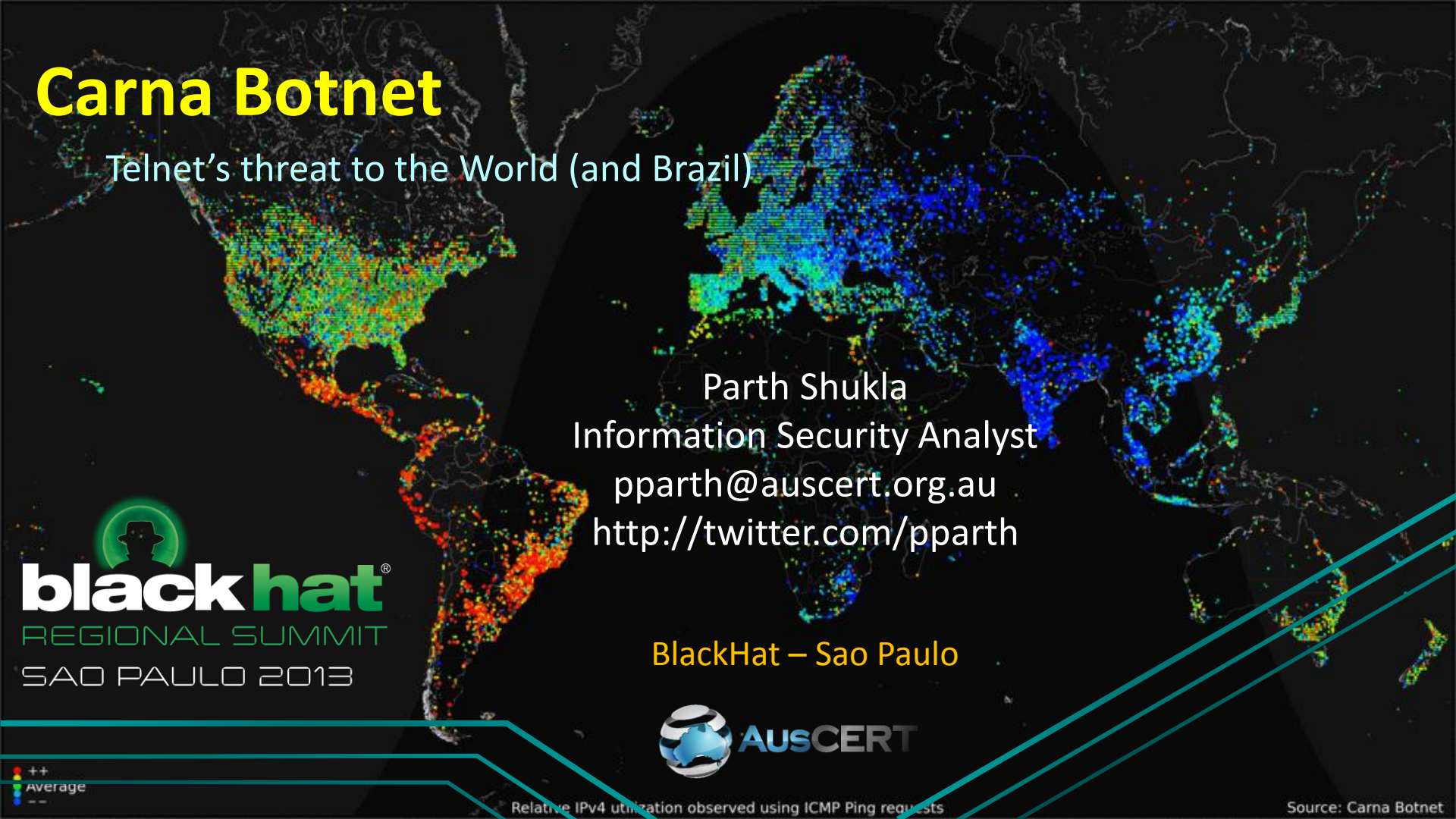
Carna Botnet

Telnet's threat to the World (and Brazil)

Parth Shukla
Information Security Analyst
pparth@auscert.org.au
http://twitter.com/pparth

BlackHat – Sao Paulo



Why should you care about Carna Botnet?

- This research shows that there is a high ratio of easily vulnerable devices Worldwide – which highlights a major security concern.
- Malicious agents can use this to take root control of vulnerable devices
- Can have a serious impact on any economy because root shell access means:
  - Sniff all passing network traffic, and/or
  - Modify passing network traffic, and/or
  - Shutdown/reboot devices at will, and/or
  - Use it to relay illegal traffic (such as child porn) and/or
  - Perform cyber attacks on other countries or companies and innocent countries or companies might be blamed

Format of this presentation

- This presentation will split over two sessions with a break in between:
  - 10:30am to 11:10am
  - 11:30am to 12:10pm
- Take a break in the middle and restart at 11:30am for Part 2
  - If I go too quickly then we will take a break at 11:10am as scheduled
  - Otherwise if people are happy then I can continue until around 11:20am and we can take only a 10 minute break?
- If we run out of time – we can continue after 12:10.
- Please fill out Session Evaluation at the end to provide feedback.

Introduction

What is the Carna Botnet?

- Millions of devices that were compromised for use in conducting the “Internet Census 2012” by an anonymous researcher

- Wait, what is Internet Census 2012?

What is the Internet Census 2012?

- Complete Scan of the allocated IPv4 ranges of the Internet
- Results & paper released Mid-March by an anonymous researcher
- http://internetcensus2012.bitbucket.org/paper.html
- Results contain 9 TB of logs (pure text!)
- Publicly available for download through a torrent as 568 GB of highly compressed (ZPAQ) files
- I just finished my thesis on this public data of the Internet Census
  - My thesis project: http://bit.ly/census-thesis

What is in the 9 TB of data?

- ICMP Ping (52 billion records) - 1.8 TB
- Reverse DNS (10.5 billion records) - 366 GB
- Service Probes (175 billion records; 4000 billion requests) - 5.5 TB
- Hostprobes (19.5 billion records) - 771 GB
- Syncscans (71 billion ports scanned) - 435 GB
- TCP IP Fingerprint (80 million records) - 50 GB
- IP ID Sequence (75 million records) - 2.7 GB
- Traceroutes (68 million records) - 18 GB

How is this feasible?

- Maximum of 4,294,967,296 IPv4 Addresses
  - Only 3,706,650,624 are allocated
- Using 1 device to scan and perform a *comprehensive* scan of 1 IP address per second, it would take:
  - 3.7 billion seconds ≈ 117.5 Years (11.7yrs if 10 IP/s or 1.17yrs if 100 IP/s)
- But with 420,000 devices it would only take 2.6 hours!
- In well under 24 hours you can easily collect all the data you need for all allocated IPv4 ranges!
- For problems of logistics and how the researcher handled collection of the data, refer to the Internet Census 2012 paper

What is the Carna Botnet?

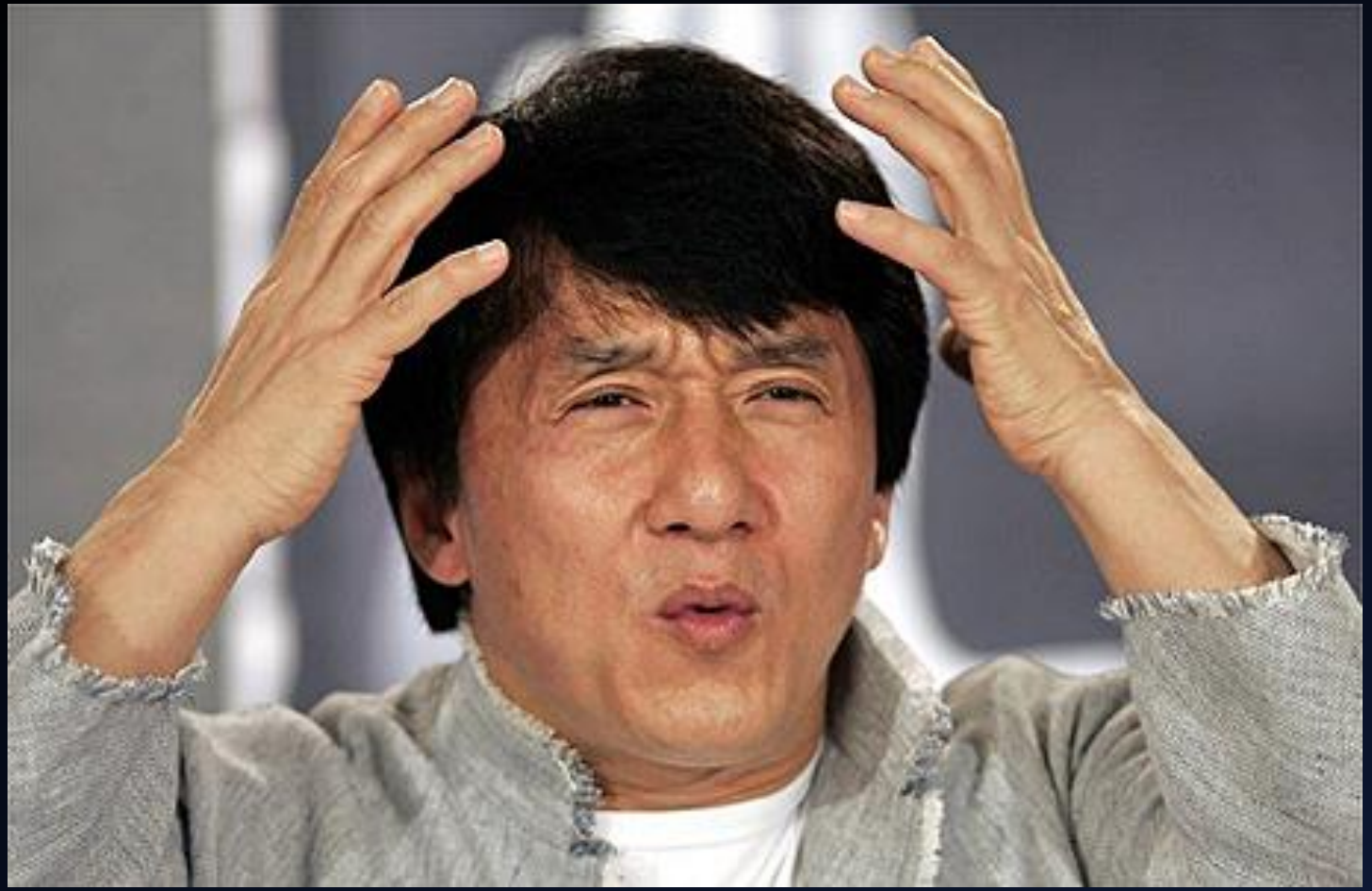
- Millions of devices that were compromised for use in conducting the “Internet Census 2012” by an anonymous researcher
- 70% of these devices were either too small, didn’t run Linux or were somehow limited (e.g. no “ifconfig”)
  - Traceroutes of **some** of these devices are part of the public torrent
- ≈1.3 million of these were not limited and had “ifconfig” on them so they could be identified
  - 420 Thousand of these met minimum CPU/RAM requirements of the researcher and were therefore used to perform Internet Census 2012
  - A list of these devices has never been released or published
  - I am the only researcher/organisation who has this full list (not counting the original anonymous researcher)

The obvious

- Just stating the obvious so it's in people's mind when thinking of this data and the rest of the presentation.
- Why has this researcher remained Anonymous?
- No permission was asked to compromise devices.
- What was done would most definitely be considered illegal in many countries. That's why I'm presenting on the data and not him or her.
- And **NO**, I am **not** the original researcher.
  - I don't want to steal his/her credit

How to be part of the Carna Botnet?

- A device must be directly reachable from the Internet
- Have telnet listening on default port 23 (with no firewall)
- Allow login using one of the default manufacturer credentials
  - E.g. admin:admin, admin:password, root:password etc



by Parth Shukla on 2013-11-27 @ BlackHat - Sao Paulo

How to be part of the Carna Botnet?

- A device must be directly reachable from the Internet
- Have Telnet listening on default port 23 (with no firewall)
- Allow login using one of the default manufacturer credentials
  - E.g. admin:admin, admin:password, root:password etc
- Not just make 1 mistake but 3 mistakes to be part of this botnet!
- To be part of the ≈1.3 million analysed here, also needed ‘ifconfig’
  - To be part of the subnet of ≈420k actually used for the Internet Census, needed ability to upload custom binary and must meet some minimum RAM and CPU requirements. According to the anonymous researcher, this was to avoid interfering with industrial controls or mission critical hardware

This presentation

- About the ≈1.3 million identifiable compromised devices
  - This data obtained directly from the anonymous researcher
  - Used for analysis for the rest of the presentation
  - **NOT** publicly available! Exclusively given to me.
- Particular focus on devices located in **Brazil** and **South America** in the data
  - However we will look at some other countries and continents.
- Clarify:
  - Internet Census = 9 TB data (public torrent) – my thesis
  - Carna Botnet = Devices compromised **to conduct** Internet Census (private) – my research in this presentation
- Side note: the Carna botnet is unusual because it's not created by phishing, exploiting a coding error or social engineering!

Why so many compromised devices?

- ≈1.3 million! WHY?!
- Are there really that many ‘stupid’ people?
- We will come back to this later

- Let’s develop some foundation and context first by looking at what data I was given and what it contains

Story time

The story begins

- Small story to get started on how I got this exclusive data.
- It hasn't even been one full year since I started working in the IT Security Industry.
  - So, it is really a story of luck in my opinion.
- Anonymous research sent email to Full Disclosure mailing list on 18th of March.
- I first emailed him on 28th of March.
- He replied on 30th of March.
- Last communication from him/her on 27th of May.

My first email

- I sent my first email on 28th of March (10 days after original researcher's email on Full Disclosure).

- "I was hoping you can provide AusCERT with the list of devices and their information (IP addresses, date/time and any other relevant information) for all the unprotected devices you discovered that are located in Australia."

- Didn't really expect a reply. I sent it because I thought "why not?"

First reply email

- “Hi and thanks for your email. Congratulations, you are the first CERT/Security research who contacted me and asked if I had something more.” – 30th of March 2013.
- S/he then sent 12 MB of encrypted ZPAQ files which would decompress to 882 MB.
- More importantly, it contained ALL compromised devices not just the ones located in Australia.
  - Nice surprise =)

Last email

- Dated 27th of May 2013.
- I asked: “I'm curious to know if I'm still the only one with the data?”
- Reply: “No, no other emails, you are still the only one.”
- S/he offline since. I have sent a couple of emails since then but no reply. I guess this is not too surprising!

Infected Devices Data

by Parth Shukla on 2013-11-27 @ BlackHat - Sao Paulo

Information for each compromised device

1. **MAC address -** the last byte replaced by an ascending number
2. **Manufacturer** - derived from MAC address
3. **RAM** - in kilobytes - parsed from /proc/meminfo
4. **Uname** - output of uname -a
5. **CPU Info** - output of /proc/cpuinfo
6. **IPs** - list of all IPs associated with this device. Last byte of each IP was zeroed by researcher. Accuracy of each IP to within a C class.
   - **Country Code** - two letter country code for each of the IPs. Correct at the time the device was compromised.

How was this data supplied?

- 3 ZPAQ files totalling 12 MB which uncompressed to 3 text files which totalled about ~900 MB.
- Each line contained information on one device
  - All information for each device was separated by tabs.
  - Eg.: [MAC]\t[Manufacturer] \t[RAM] \t[Uname] \t[CPU Info] \t[IP1];[Country1] \t[IP2];[Country2] etc..
- Creates a small problem because number of fields per line cannot be predicted since even the IP addresses are separated by tabs

Priorities

- Originally I only cared about the IP addresses because we can use that to find these devices!
- Ideal situation would be to notify all device owners
  - But not possible because we only have C class of the IP address of the compromised device.
  - NO timestamps!
  - An entire C class might not necessary belong to one company or the same company as it belonged to at the time of the compromise
- Next best: analyse the data in detail so to make it easier for people to find these devices themselves in their networks and homes.

Re arranging data to create conformity

- First needed to create uniform columns.
- Need to be reliably able to say for example that the 4th column is always a list of IP addresses for example.
- Through some linux/PHP magic, I replaced the separator for the fields with a $ instead of a tab.
  - I left the IP address list separator as tabs.
  - I chose $ because some searching revealed that I couldn't use other characters such as commas etc. because some of the data actually contained those characters
  - $ was the character that didn't exist anywhere in the data.

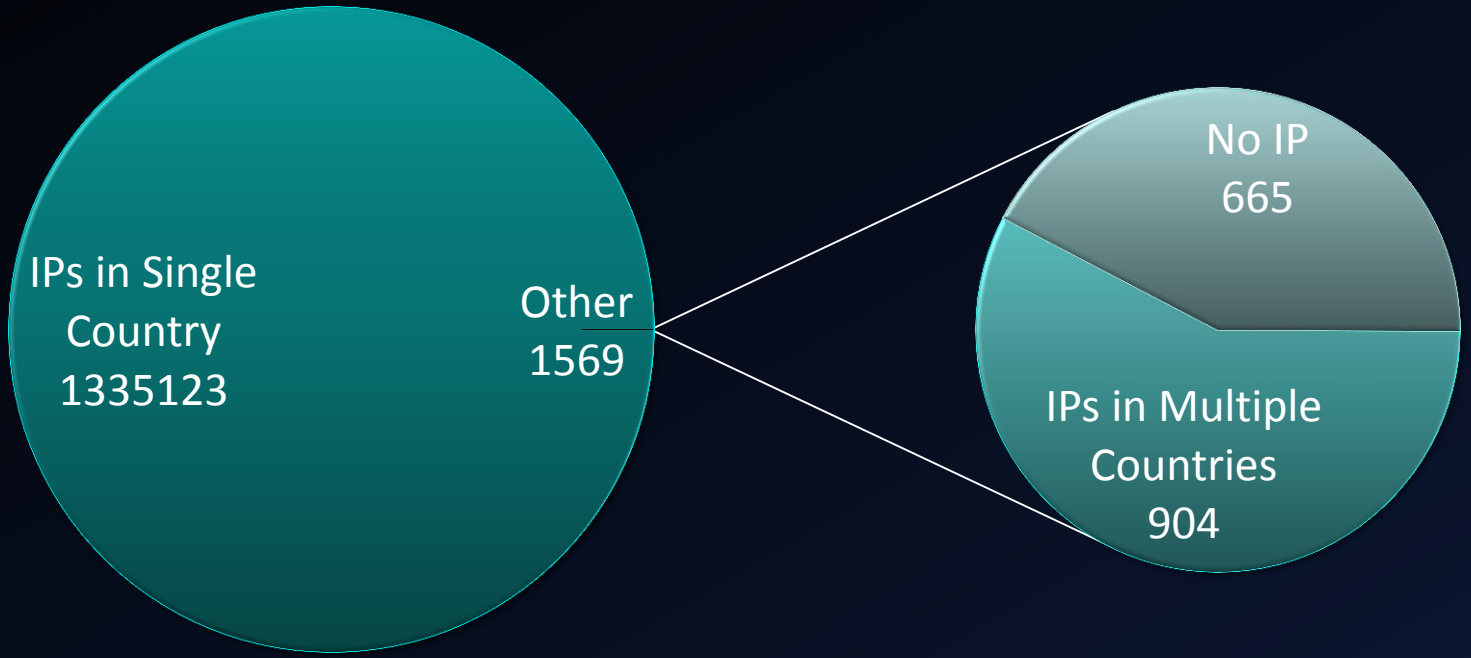
Re-arranged data example

- Re-arranged data looked like this:
- [MAC address]$[Manufacturer]$[RAM]$[uname]$[IP addresses and country code]$[/proc/cpuinfo]
- [IP addresses and country code] = [IP1];[Country1]\t[IP2];[Country2]\t[IP3];[Country3]…
- Uniformity of fields achieved.
- Moved CPU info to the end because it's a huge and messy field. Purely for visual reasons.

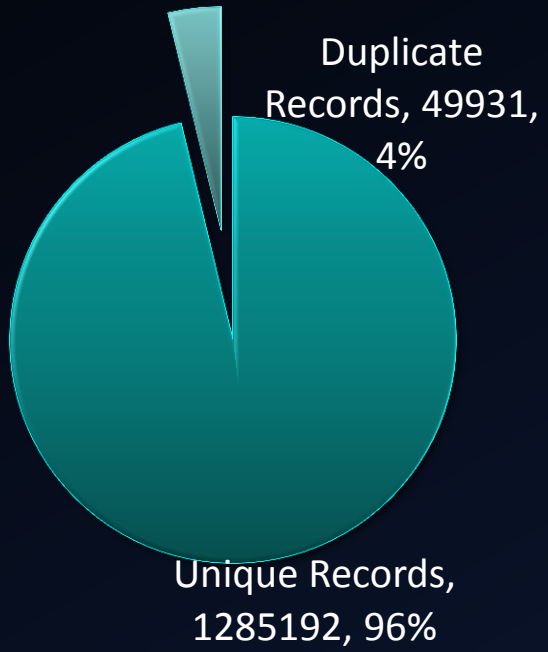
Initial analysis

Total Records:	1,336,692
Records with No IPs:	-665
Records with IPs in more than 1 country:	-904
Records with IPs in a single country:	1,335,123

Distribution of Records based on IPs



Further refining



In the 1,285,192 compromised devices there were

- 200 unique country codes
- 2,098 unique device manufacturers
- 3,880 different RAM sizes
- 10,875 unique unames
- 35,997 unique CPUs
- 787,665 unique C class IP ranges
- 1,264,223 unique MAC addresses

Primitive tools for a primitive analysis

- I relied completely on the ‘cut’ utility from Linux for analysis for my first presentation at the AusCERT 2013 conference in May.
- Unique Manufacturer by Count:
  - cut -d $ -f 2 carna_botnet.tsv | sort | uniq -c | sort -n -r > unique-manufacturer
- Very manual work. For every country or continent or manufacturer I want to analyse I have to split the caran_botnet.tsv into a smaller file with ‘egrep’ and then run the ‘cut’ utility on it.
- OK for one presentation. But when I started getting accepted at other conferences, I needed to change my plan.

MariaDB

- What better to run unexpected querying then on a relational database?
- I wrote a PHP script to insert the data into two tables in MariaDB (MySQL fork).
- This would allow me to create arbitrary indexes on the data and search and run queries on it easily.

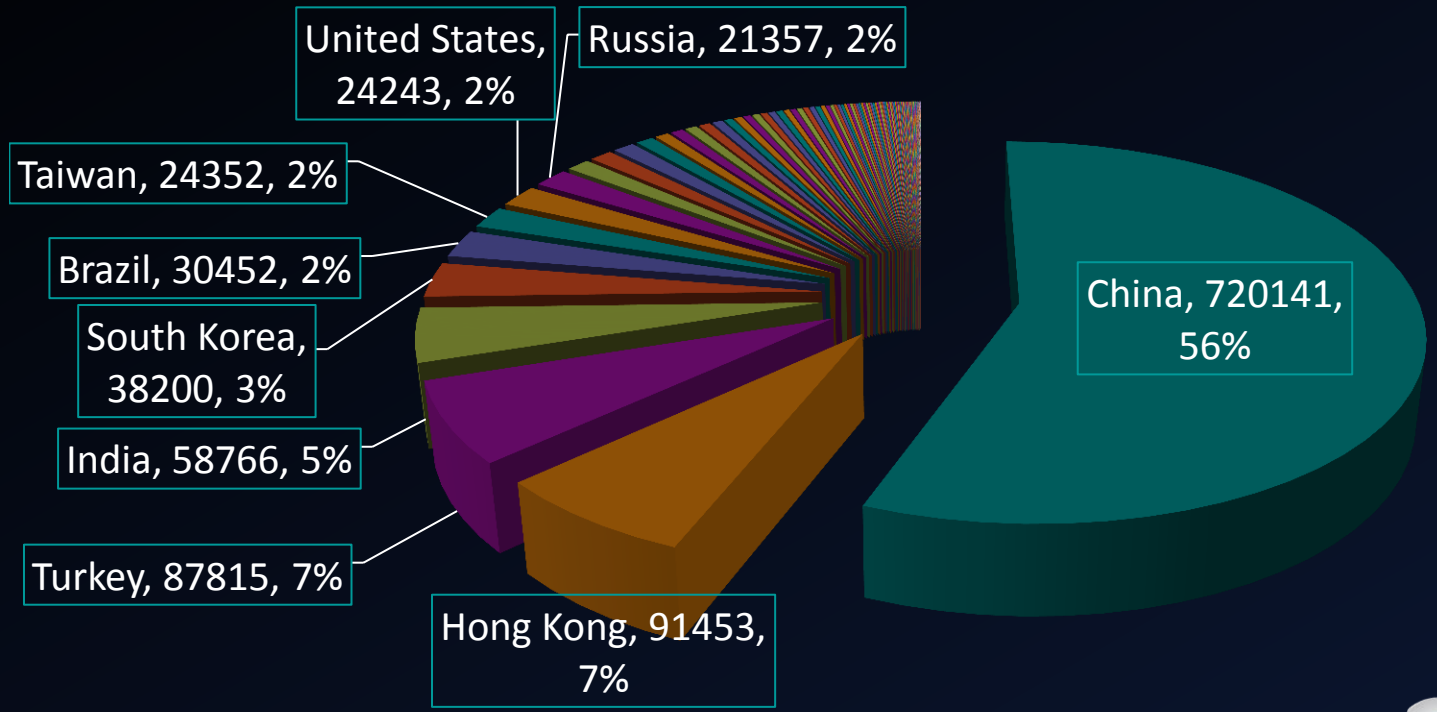
- Let’s take a quick look at the PHP file I used for inserting. Nothing complicated. Also see the final table.

Rest of presentation

- Hopefully you have some foundation of the data now.
- We will focus on the statistics that I created using this.
- 1.3 million is bad, yes, but what does it really mean? Analysis necessary to truly understand the data.
- We need to go beyond the obvious to reveal the hidden.
- I will refer to the MariaDB database through out the presentation as necessary.
- If we still have time at the end of the presentation we will run some live queries so you can see how easy it becomes to create statistics.

Countries

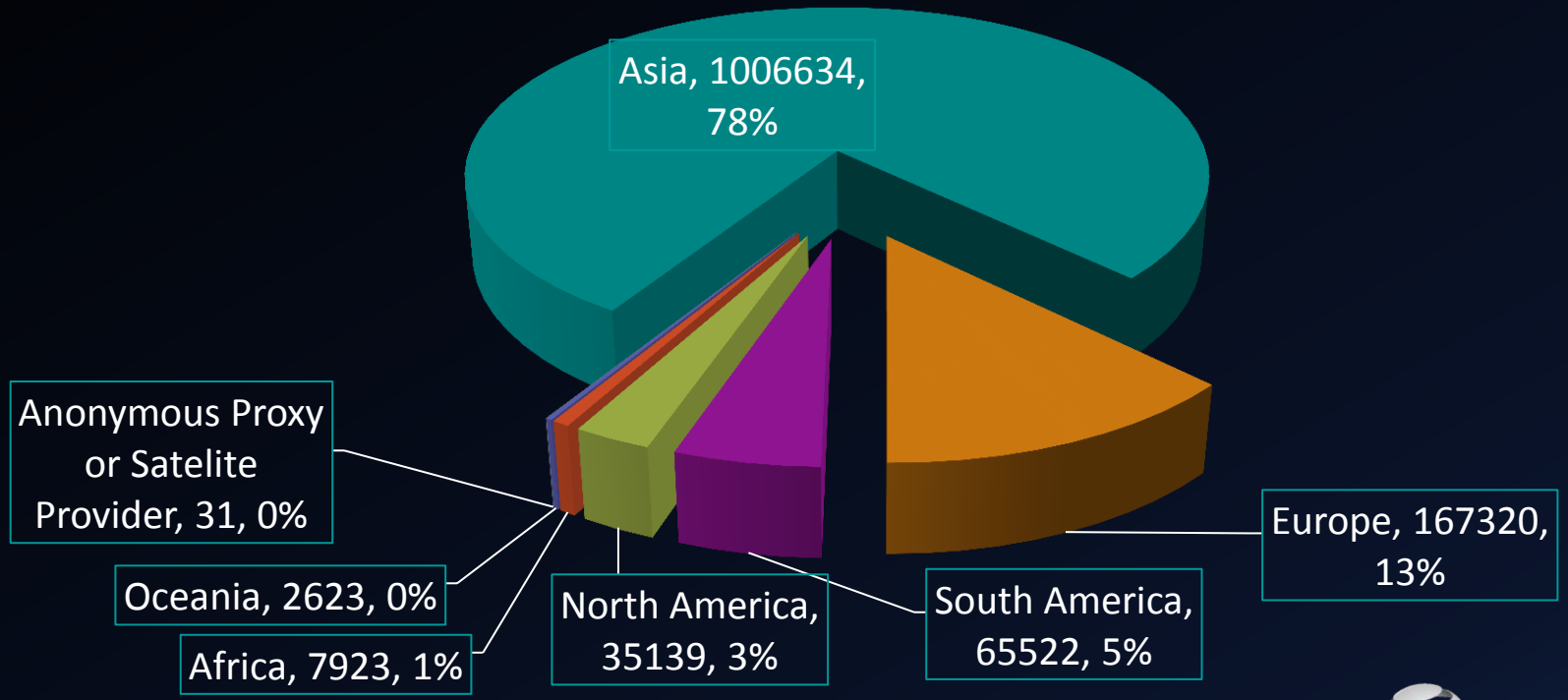
Compromised Device Distribution by Countries



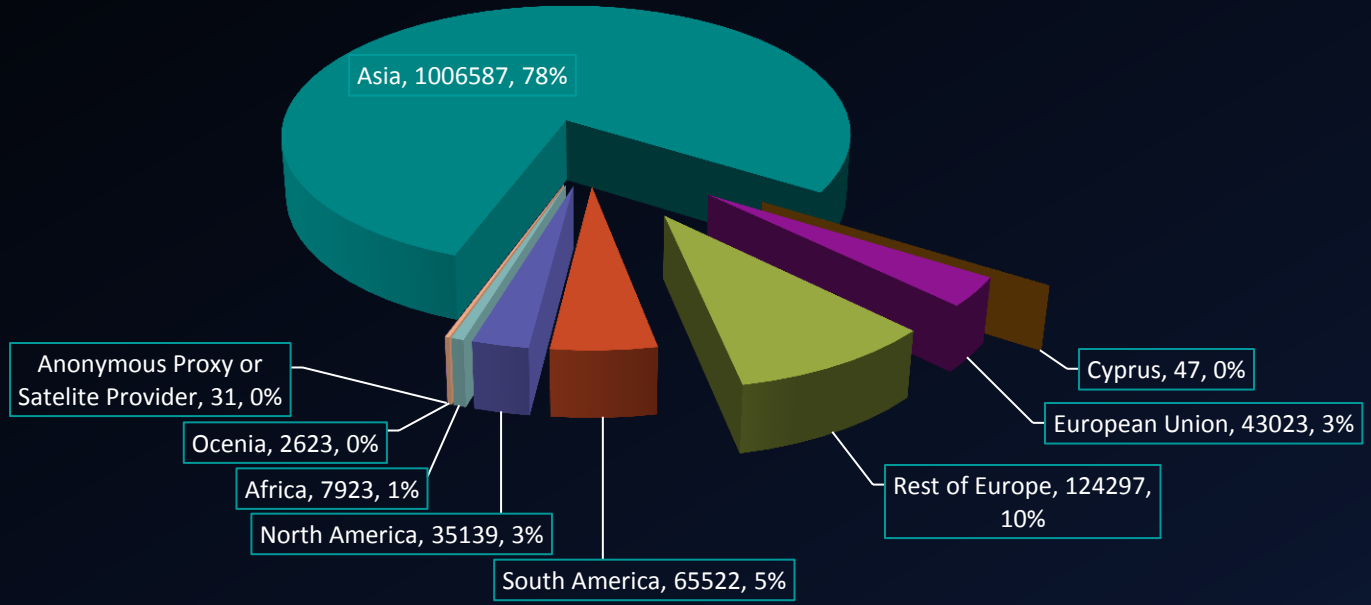
Devices by Continents

- We can also create graphs on which continent is most prevalent.
- I used dev.maxmind.com/geoip/codes/country_continent to decide which country belonged to which geographical continent.
- Created a table in MariaDB to represent this information.
- Let's take a quick look at 'country', 'continent' and 'european_union' table.
- Creating statistics with this new information wouldn't be very fun/easy to do through command line.

Compromised Device Distribution by Continents



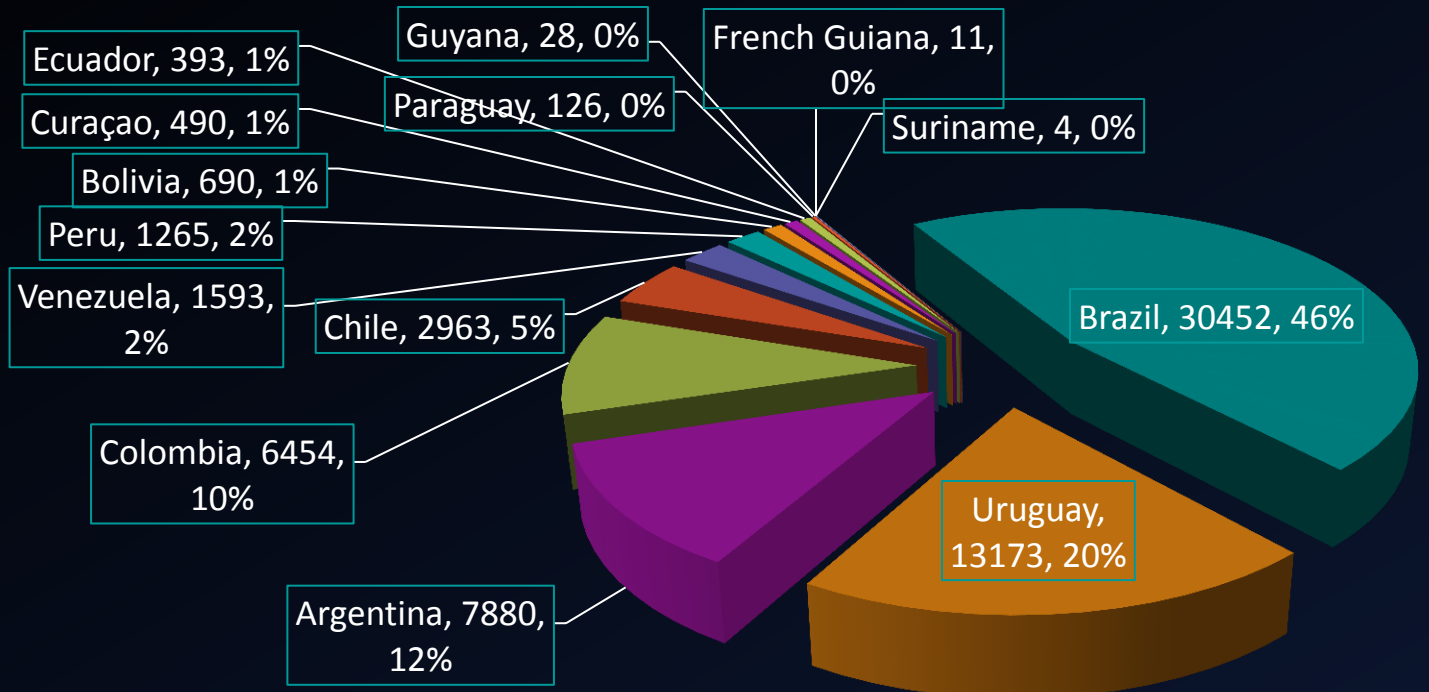
Compromised Device Distribution by Continents – including European Union



- Cyprus is part of continental Asia and European Union, hence the separate slice.

AusCERT

Compromised Device Distribution for South America



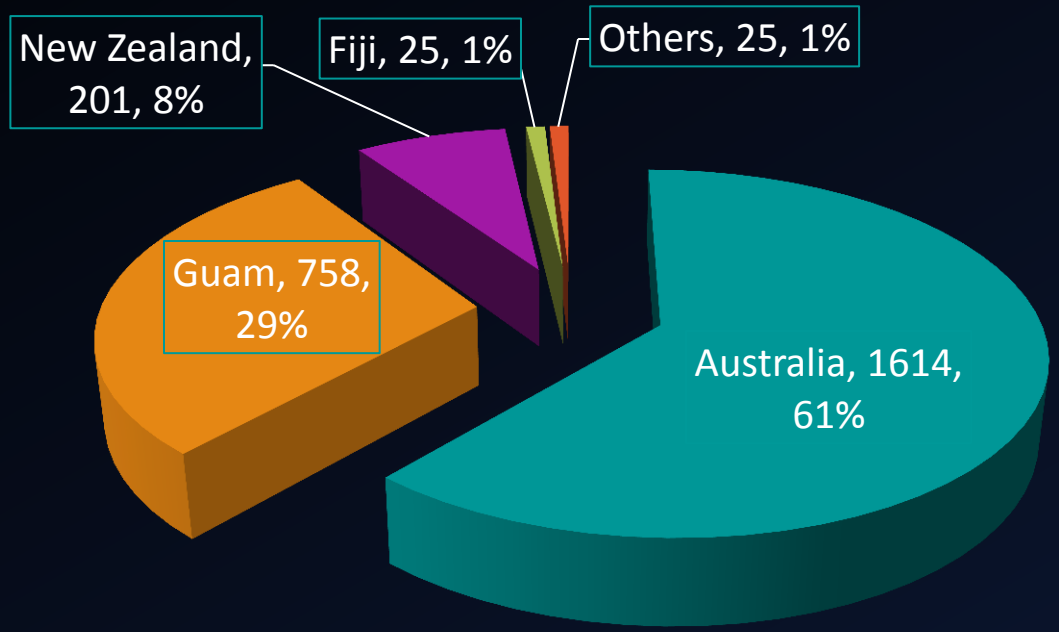
AusCERT

Where is Brazil?

- 30,452 devices located in Brazil in the data
- 6th Largest Slice (2.36%) Worldwide
- Largest Slice (46.47%) in South America

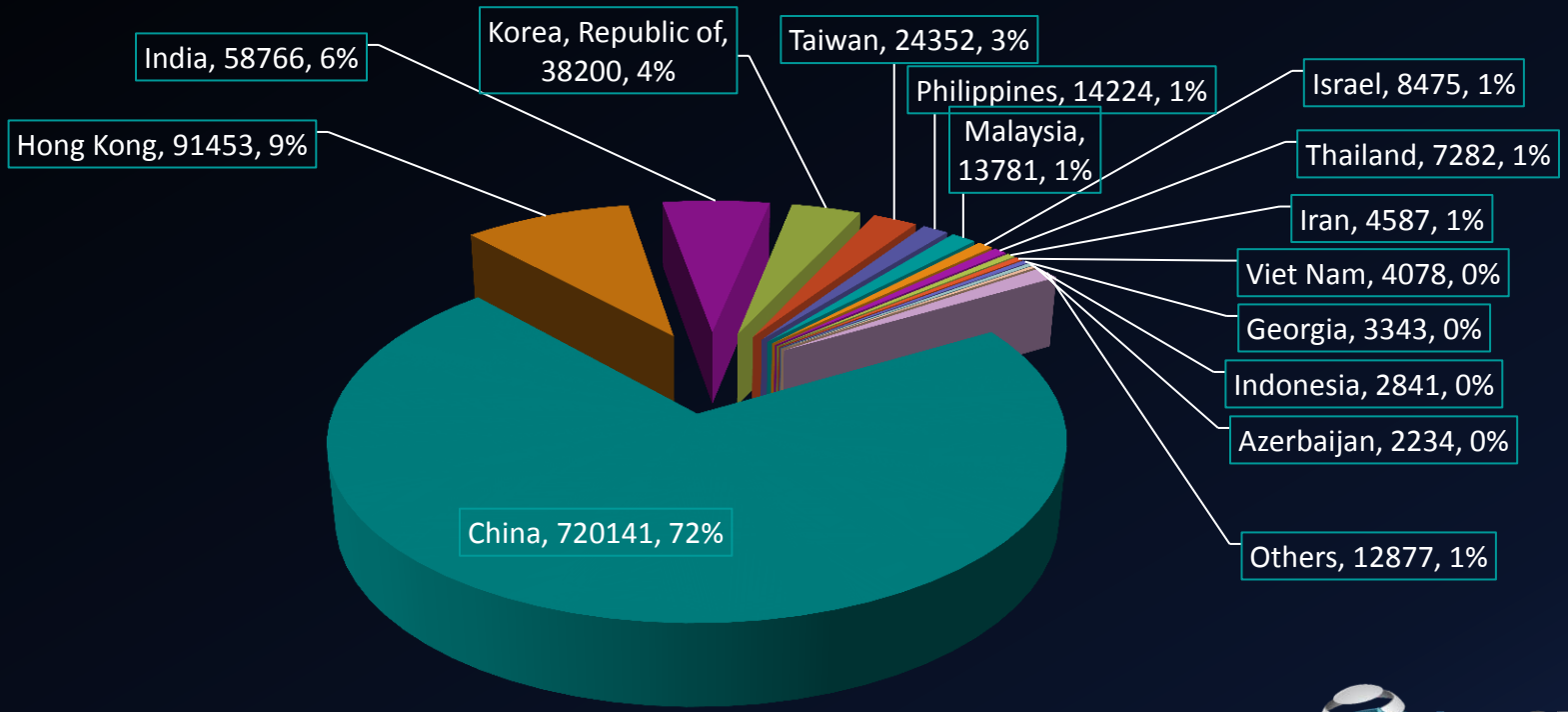
- Not very good. Scary enough for Brazil!
- For more of a scare, we will look at how easy it would be to find one of these devices in Brazil and other countries in the world at the end of the presentation.

Compromised Device Distribution for Oceania



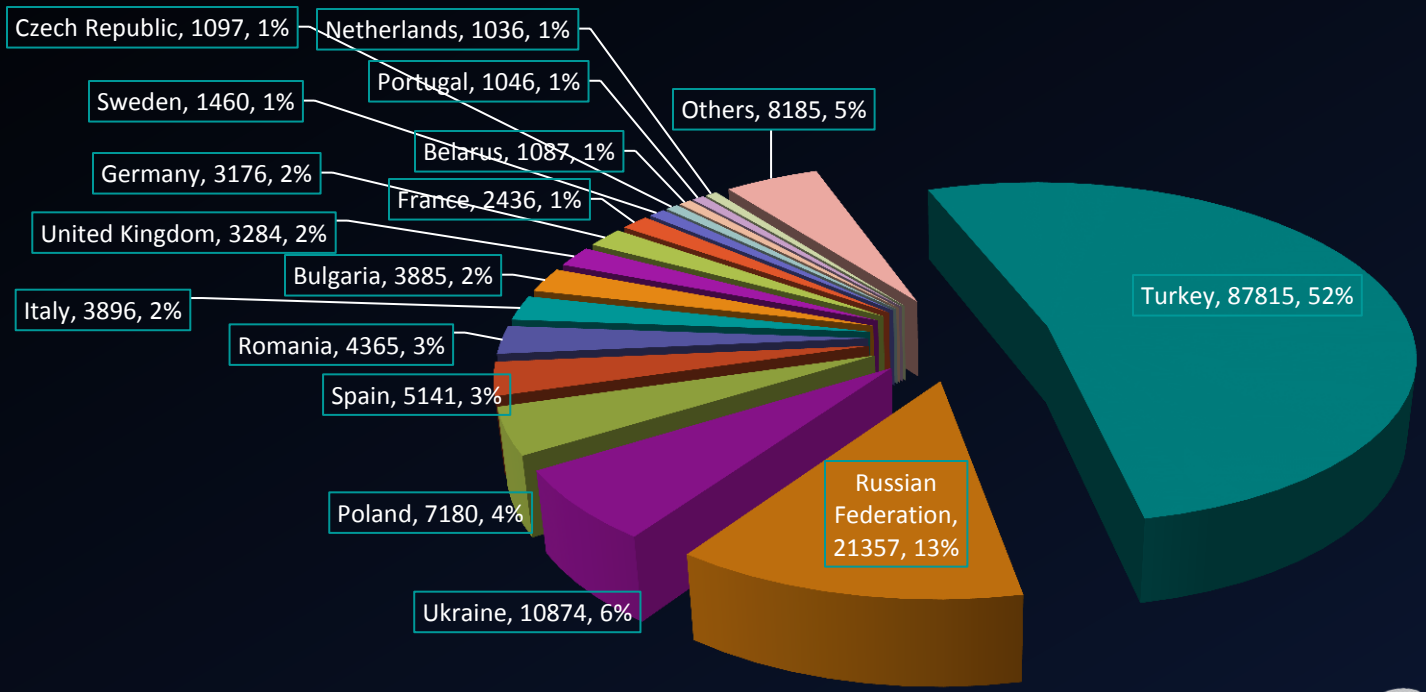
AusCERT

Compromised Device Distribution for Asia



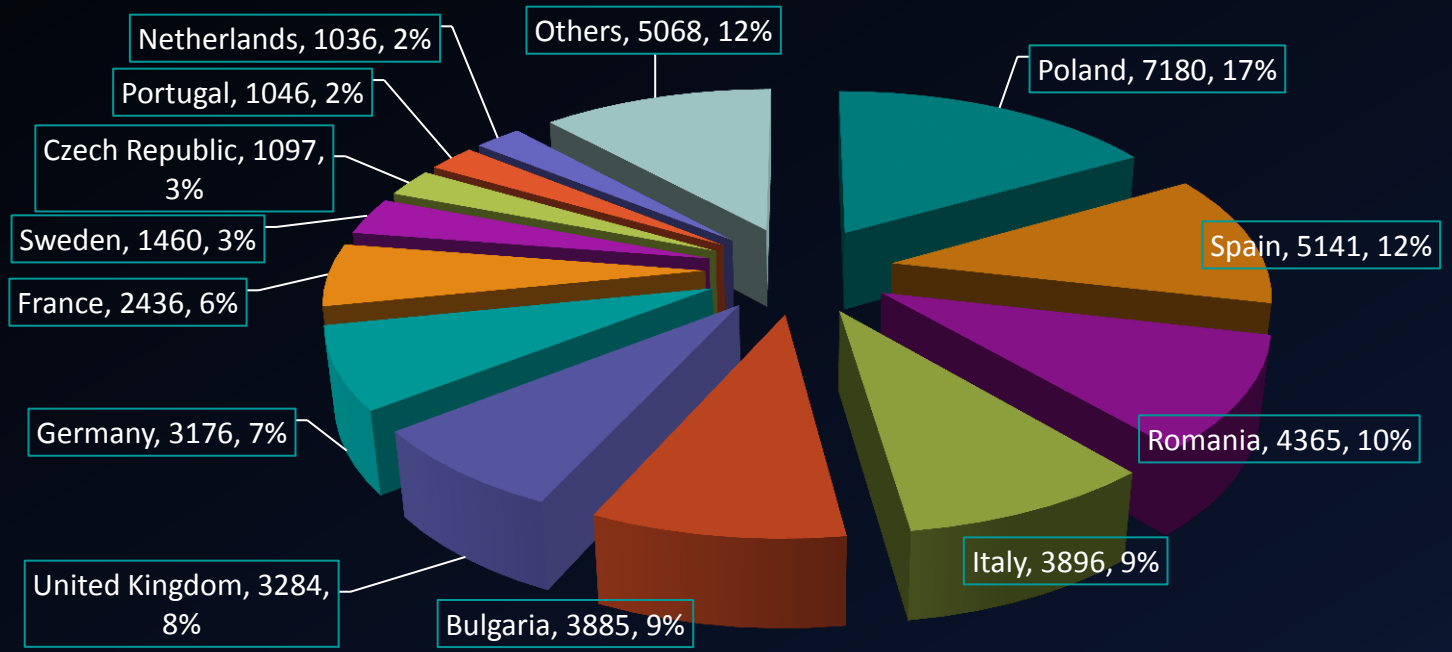
AusCERT

Compromised Device Distribution for Europe



AusCERT

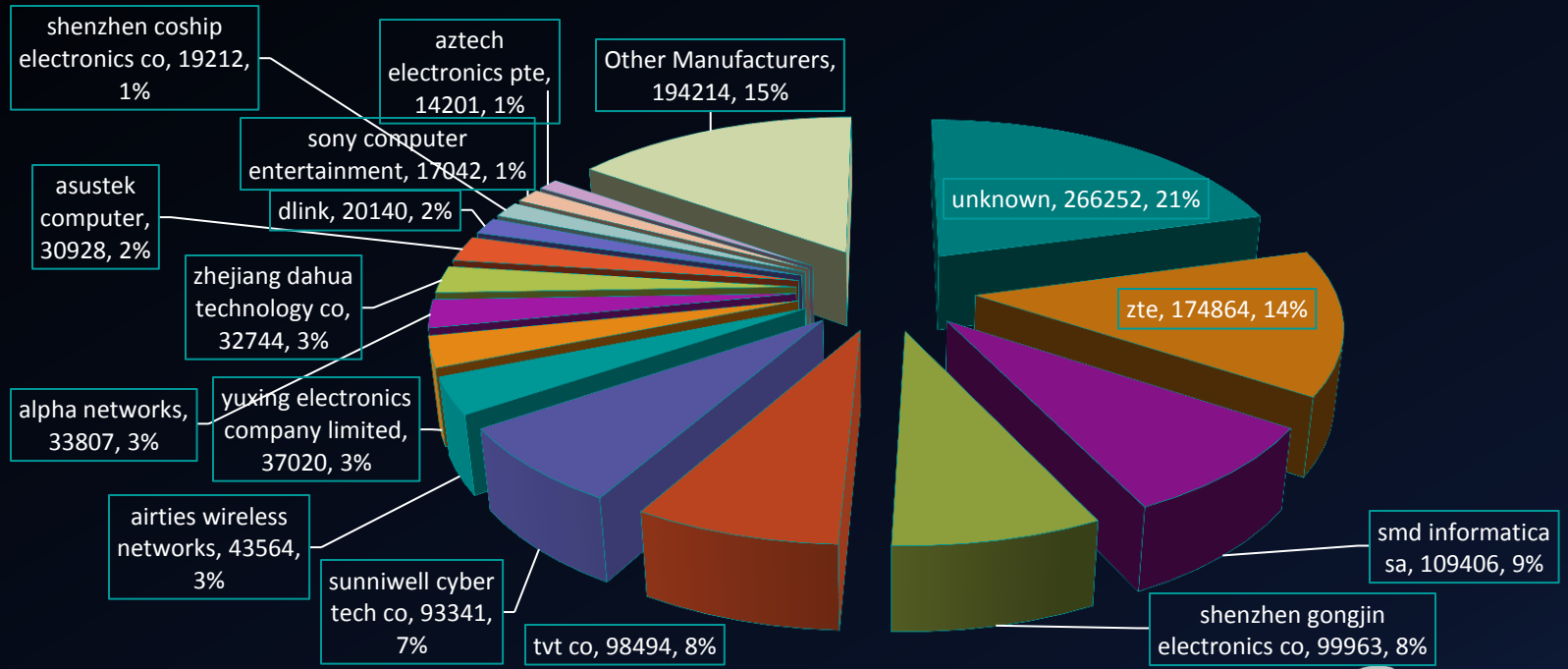
Compromised Device Distribution for European Union



AusCERT

Manufacturers - Analysis

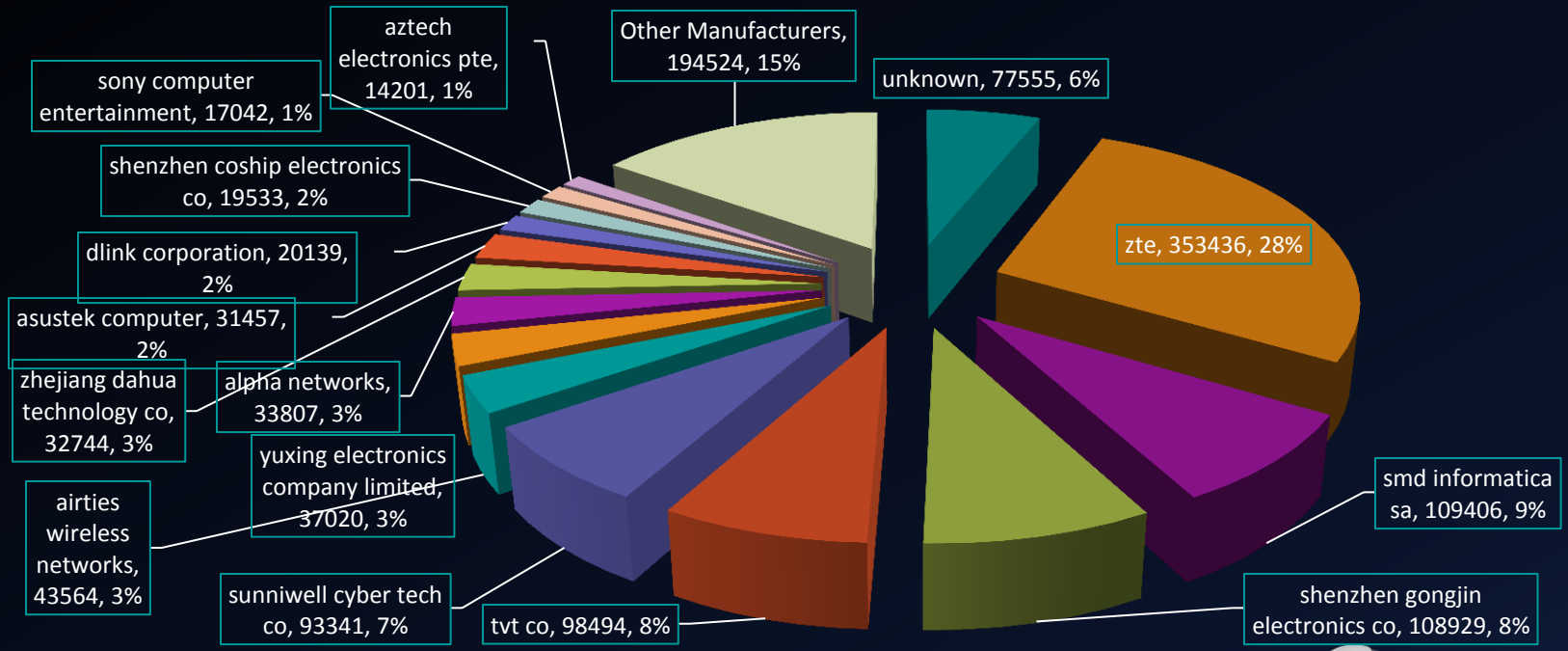
Distribution of Device Manufacturers – Worldwide – Original 'manufacturer' field



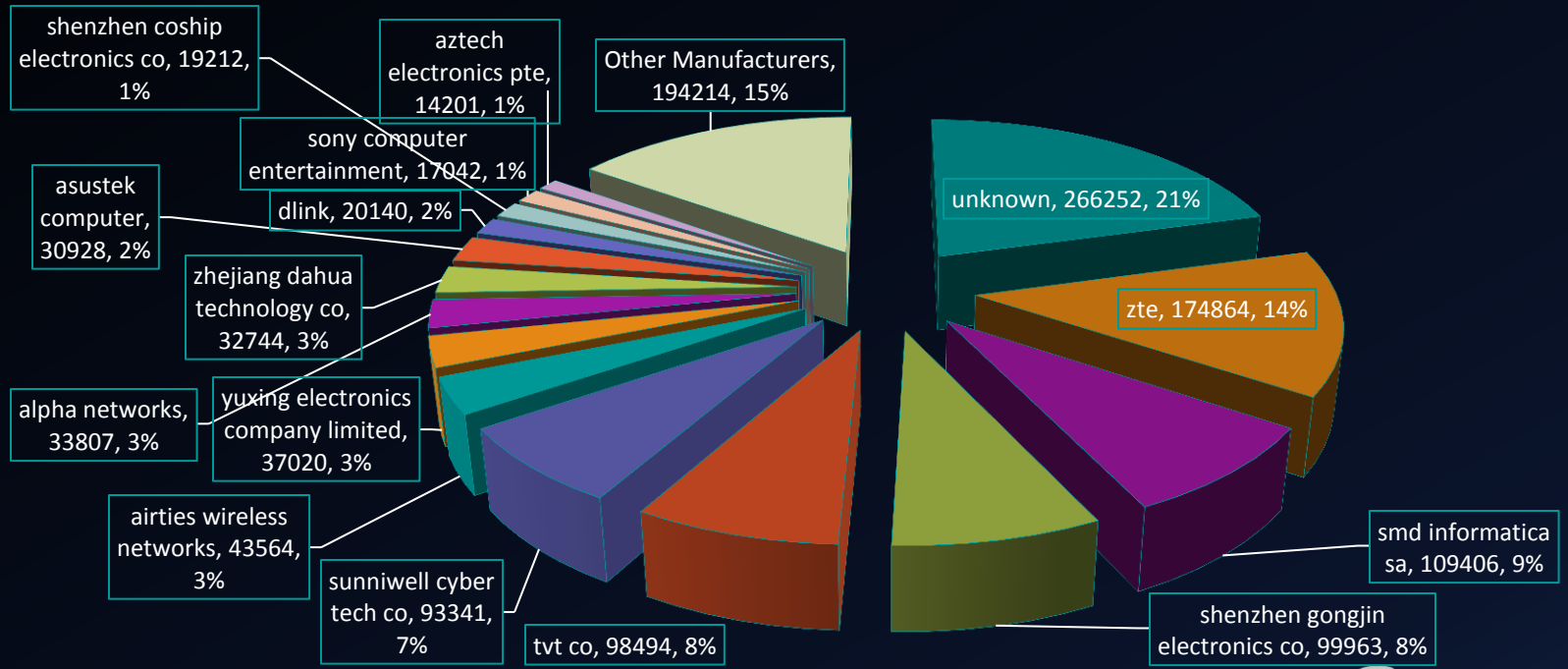
21% ‘unknown’ manufacturer?

- Checked these against the IEEE official list
  - 71% of the 21% unknowns matched to the list!
  - New unknown reduced from 21% to about 6%
- Also checked the other 79% (100-21=79) against IEEE official list
  - Perfect match to already recorded info in the manufacturer field of the data.
  - Reliability/consistency of data confirmed.
  - Hence, ‘manufacturer’ field in data provided incomplete but not inconsistent.
- Could be many reasons why ‘unknown’ was recorded in the data
  - A result of error encountered when deriving manufacturer (e.g. timeout)
  - Using out-dated mac to manufacturer list to derive manufacturers
- The entire field deleted from data and re-derived using IEEE official list

Distribution of Device Manufacturers – Worldwide – Re-derived 'manufacturer' field



Distribution of Device Manufacturers – Worldwide – Original 'manufacturer' field



Still not happy with 6% ‘unknown’

- Provision in the IEEE specification for Locally Administered MACs
  - This range is **not** up for grabs from IEEE
  - Can be used in VMs or when setting manual MACs
- If 7th bit of a MAC address is 1 then it is considered Locally administered
  - i.e. second HEX digit of MAC: 2, 3, 6, 7, A, B, E, F
- Let’s take a look at Universally vs Locally Administered MACs in the Records for the World in the currently 6% ‘unknown’ data.

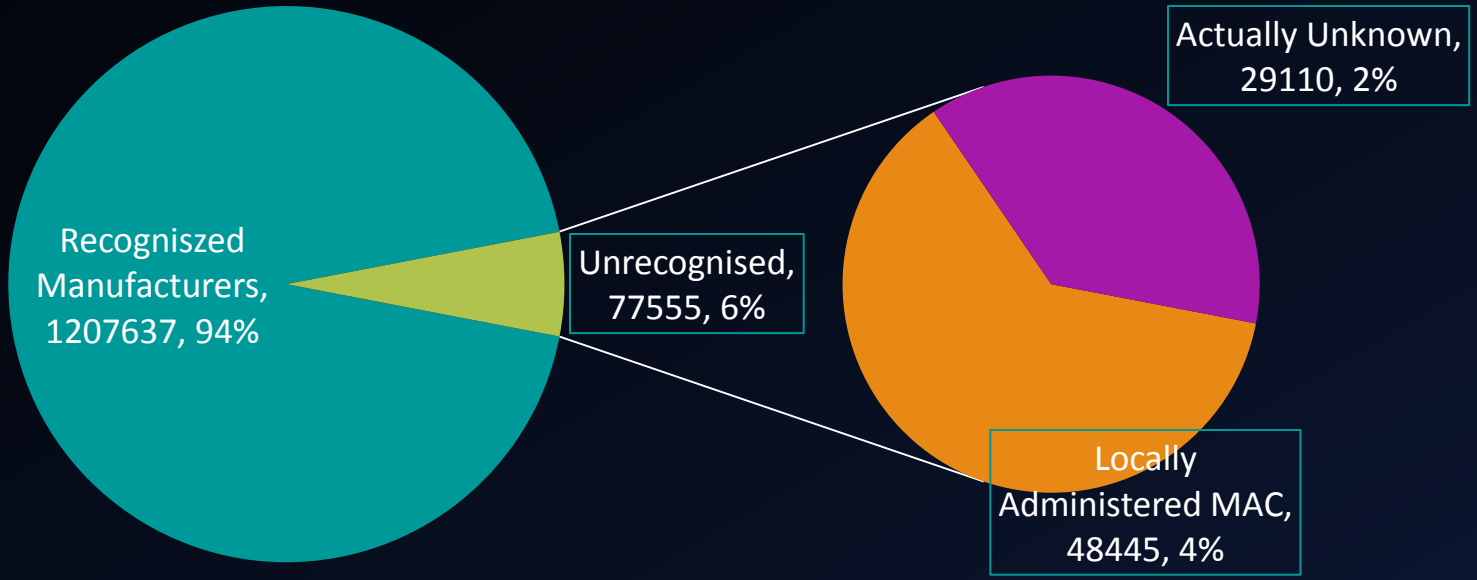
IEEE Local vs Universal sidenote

- Interesting unrelated story from my research
- Some MAC address should be locally administered according to specification but have been assigned to organisations by the IEEE.
- I emailed IEEE 802 committee to ask for what's going on in early August. Last I checked, they were still trying to figure it out.
- I found 13 MAC ranges that should be locally administered but have been allocated by the IEEE.
  - 02:07:01, 02:1C:7C, 02:60:86, 02:60:8C, 02:70:01, 02:70:B0, 02:70:B3, 02:9D:8E, 02:AA:3C, 02:BB:01, 02:C0:8C, 02:CF:1C, 02:E6:D3.

How many ‘Unknown’ vs Locally Administered MACs in the data

2nd Character of MAC	Number of Ocurrances
0	18717
2	47132
3	1
4	1088
6	434
8	9147
9	1
a	394
c	157
e	483
f	1

Unrecognized Manufacturers - Worldwide



Official IEEE List Problems

- List at http://standards.ieee.org/develop/regauth/oui/oui.txt
- Many MACs didn't match – 77555 (6%) devices reported as unkown.
- Manufacturer's don't always use same name to register:
  - 00:19:5B registered to "D-Link Coropration"
  - 00:50:BA registered to "D-LINK"
  - 00:80:C8 Registerd to "D-LINK SYSTEMS, INC."
  - 14:D6:4D registered to "D-Link International"
- They can also be inconsistent by a little bit:
  - FC-8B-97 registered to "Shenzhen Gongjin Electronics Co.,Ltd"
  - F4-3E-61 registered to "Shenzhen Gongjin Electronics Co., Ltd"
  - 00-07-26 registered to "Shenzhen Gongjin Electronics Co., Ltd."

Official IEEE List Problems

- Therefore to create Manufacturer graphs from the Official IEEE list, the following was done to make automatic matching easier:
  - All company names were converted to lower case
  - All special characters (.,_# etc) were removed.
- Hope that most company names would match together for statistics.
  - If they didn't match, then same company would appear twice in the list with slightly different names instead of being summed together into one entry.
  - Not a big problem of concern.
- The Wireshark list gets around this by having a unique manufacturer name that can be easily matched up.

Wireshark MAC to Manufacturer List

- Located at http://anonsvn.wireshark.org/wireshark/trunk/manuf
- Contains IEEE official list and other sources
- Snippet:
  - 00:19:5B listed as “D-Link” (description: “D-Link Corporation”)
  - 00:50:BA listed as to “D-Link” (no description)
  - 00:80:C8 listed as to “D-Link” (description: D-LINK SYSTEMS, INC.)
  - 14:D6:4D listed as to “D-LinkIn” (description: D-Link International)
  - FC:8B:97, F4:3E:61 and 00:07:26 listed as “Shenzhen” (description: Shenzhen Gongjin Electronics Co., Ltd)
- Contains MACs from manufacturers that are known to violate the mac address assignment scheme and use MACs that should be Locally Assigned.
  - These will not be included in the IEEE official list for obvious reasons
  - Linux has this in **/etc/udev/rules.d/70-persistent-net.rules**

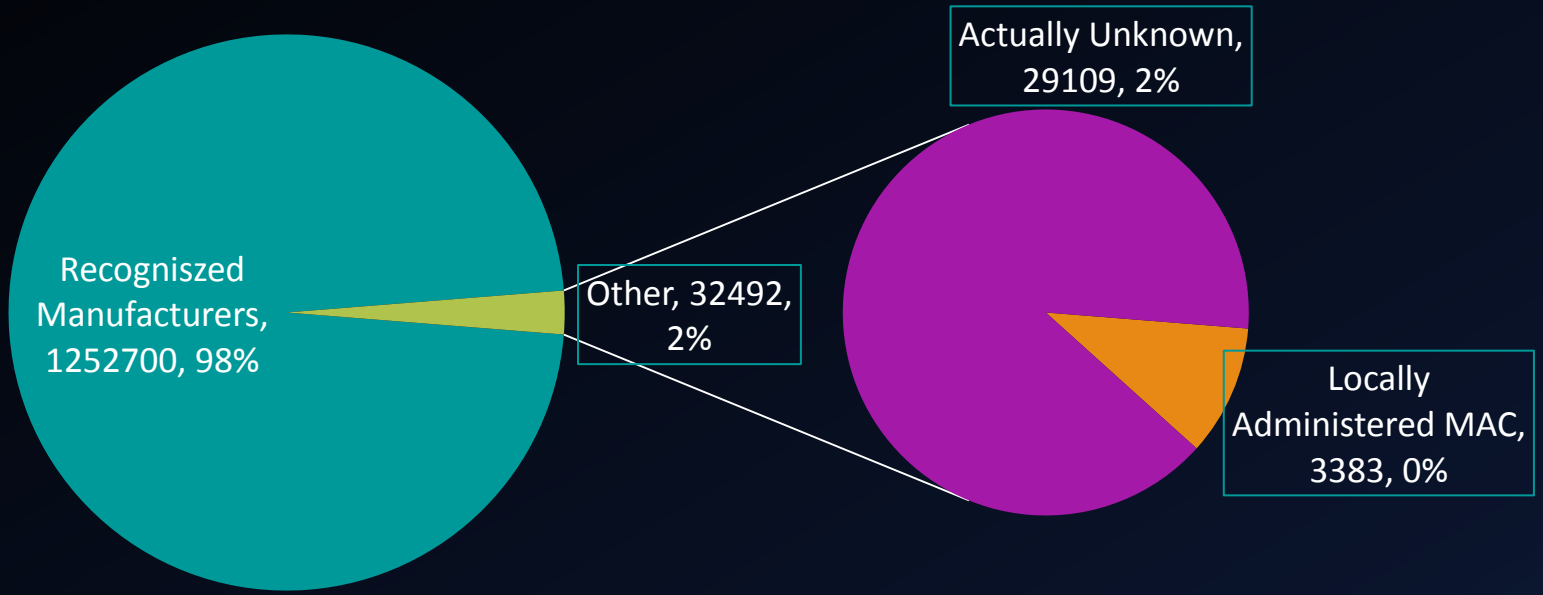
Matching the 77,555 “Unrecognised” against Wireshark’s List

- 3 matches:

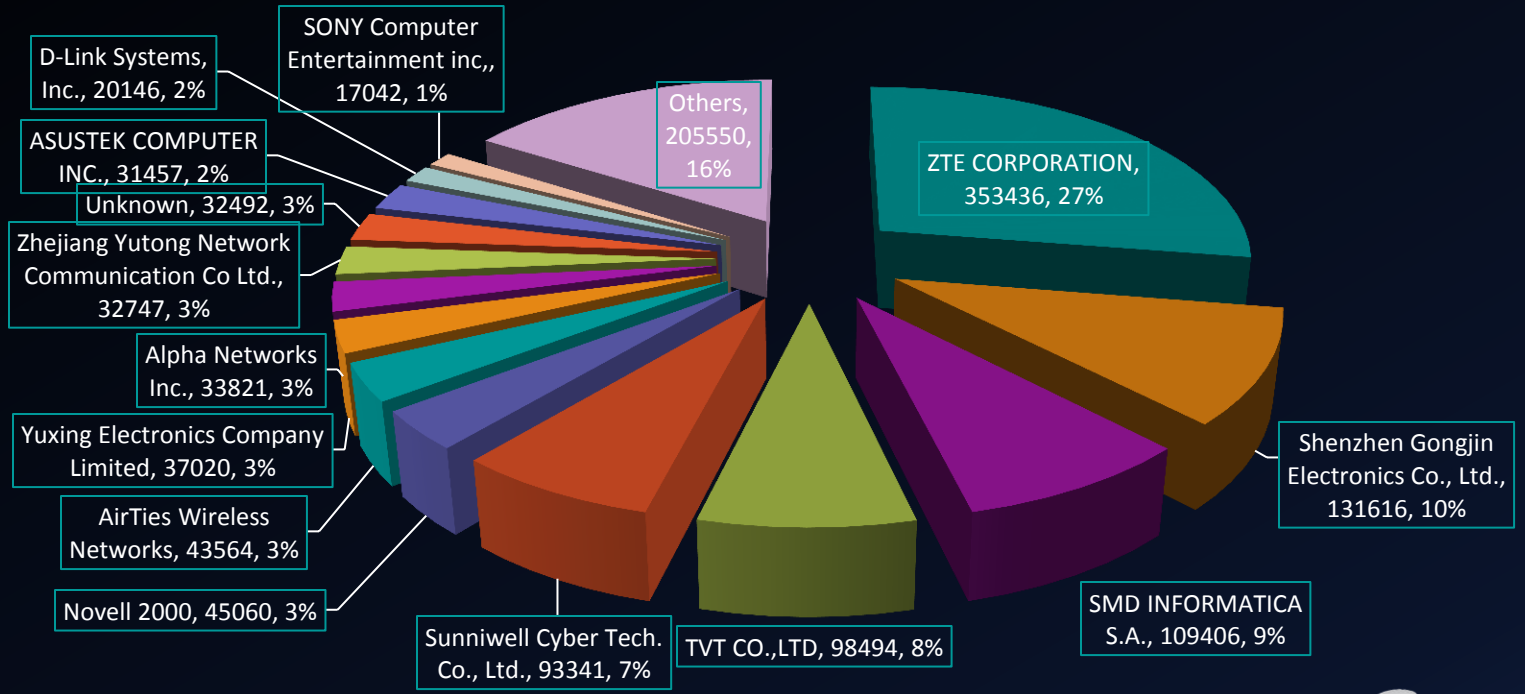
MAC	Manufacturer	Description	Devices
00:55:00	Xerox		1
52:54:00	RealtekU	Realtek (UpTech? also reported)	2
52:54:4c	Novell20	Novell 2000	45060

- Yellow are MACs from manufactures that should be Locally Administered MACs but are known to violate the IEEE rules.
- Unknown Reduced by 45063.
- Updated ‘manufacturer’ field in the data with new findings above.

Unrecognized Manufacturers – Updated



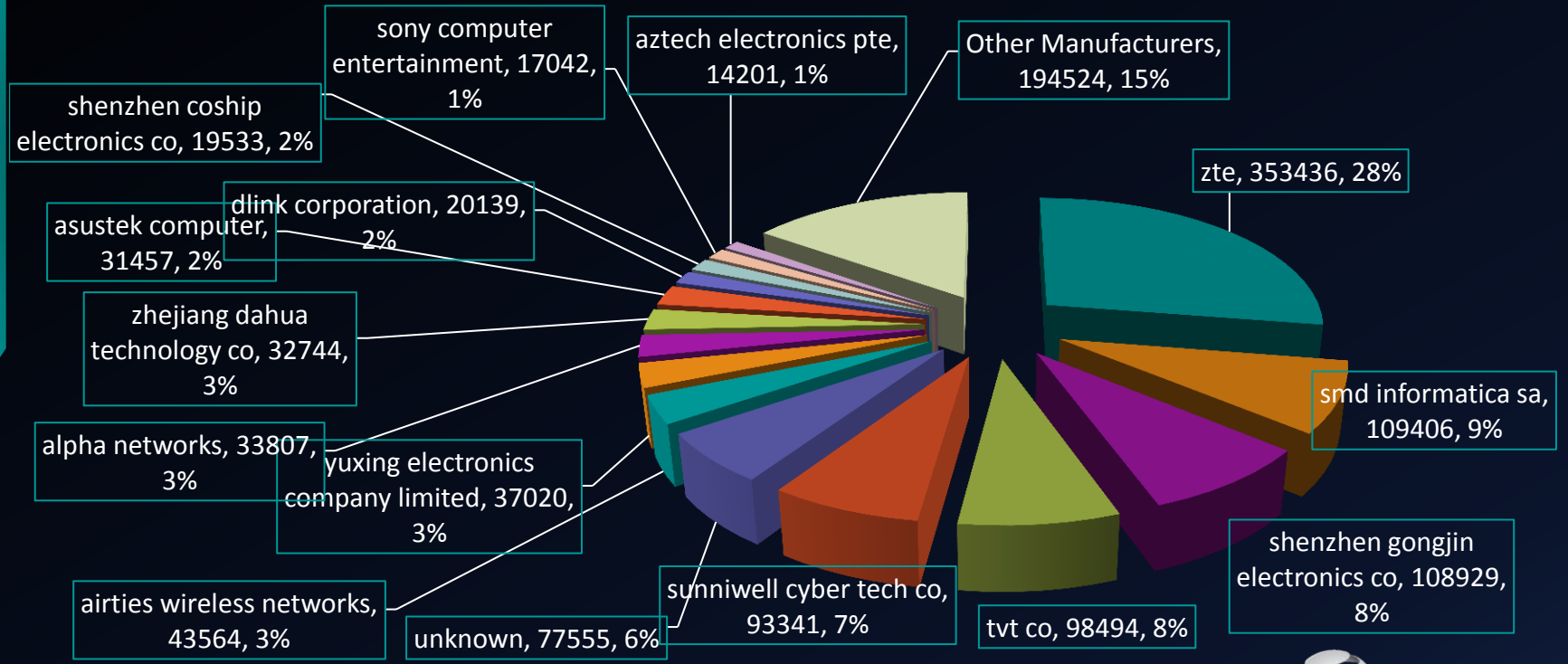
Distribution of Device Manufacturers – Worldwide – Using Wireshark's list



by Parth Shukla on 2013-11-27 @ BlackHat - Sao Paulo

AusCERT

Distribution of Device Manufacturers – Worldwide – re-derived Using Official IEEE list



Story of the ‘Unknowns’

- Original Unknown: 266252 (21%)
- Unknown after Re-driving against IEEE List: 77555 (6%)
- Unknown After checking against Wireshark List: 32492 (3%)
- Difference between IEEE and Wireshark:
  - 77555 - 32492 = 45063
- Therefore the **only** thing Wireshark has contributed is in identifying 3 extra MACs – 2 of which are Locally Administered. Other than that, nothing new has been added from Wireshark.
  - Except confusion – next slide

Shenzhen Gongjin Electronics Co., Ltd.

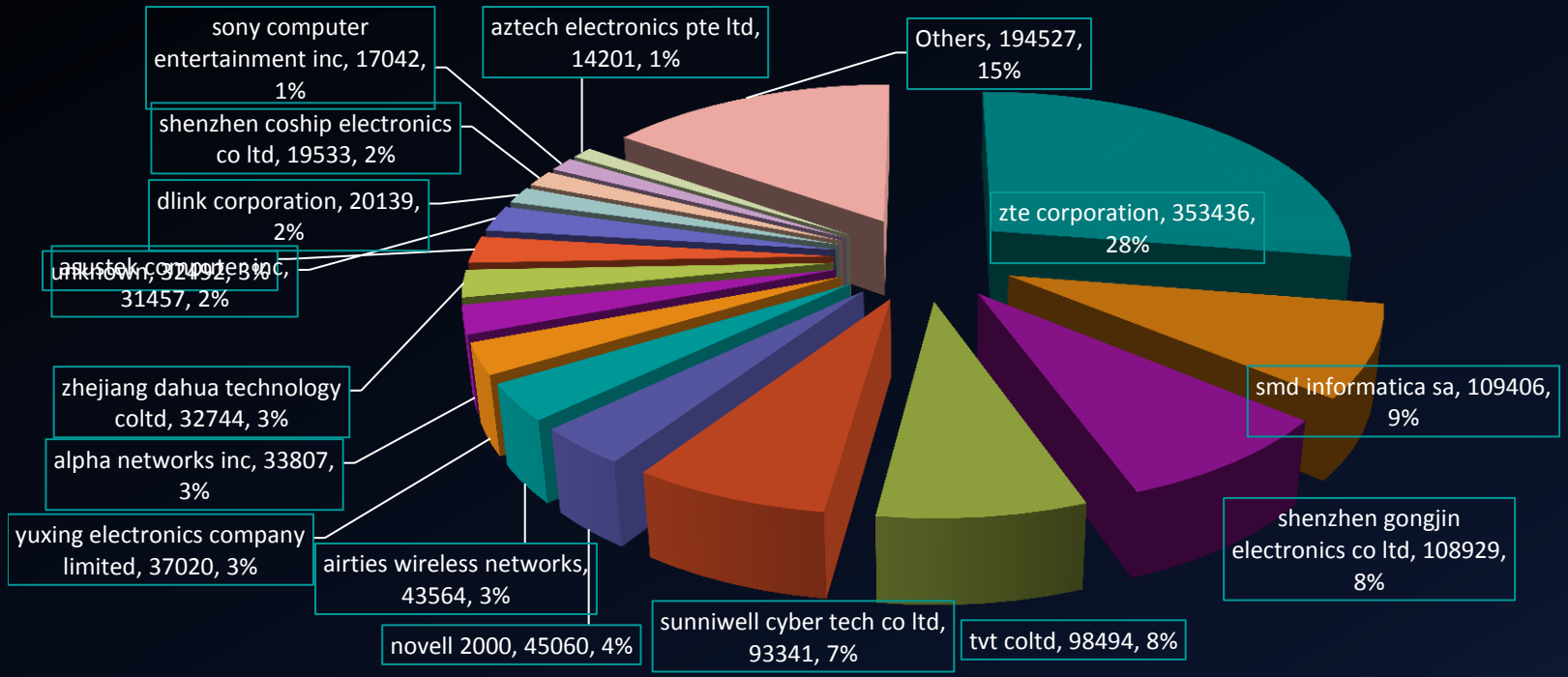
- These are classified as ‘Shenzhen’ in Wireshark’s list:
  - 00:02:09 - Shenzhen SED Information Technology Co., Ltd.
  - 00:0a:eb - Shenzhen Tp-Link Technology Co; Ltd.
  - f0:62:0d - Shenzhen Egreat Tech Corp.,Ltd
- Are these all the same company? 232 companies classified as ‘Shenzhen’!
- ‘Shenzhen’ is a city in China.
  - ‘D-Link International’ and ‘D-Link Corporation’ are the same, but
  - ‘Shenzhen Tp-Link Technology” and “Shenzhen Gongjin Electronics’ are NOT!
- We won’t use Wireshark because it combines companies into one that shouldn’t be combined. Not very reliable.

Reasoning

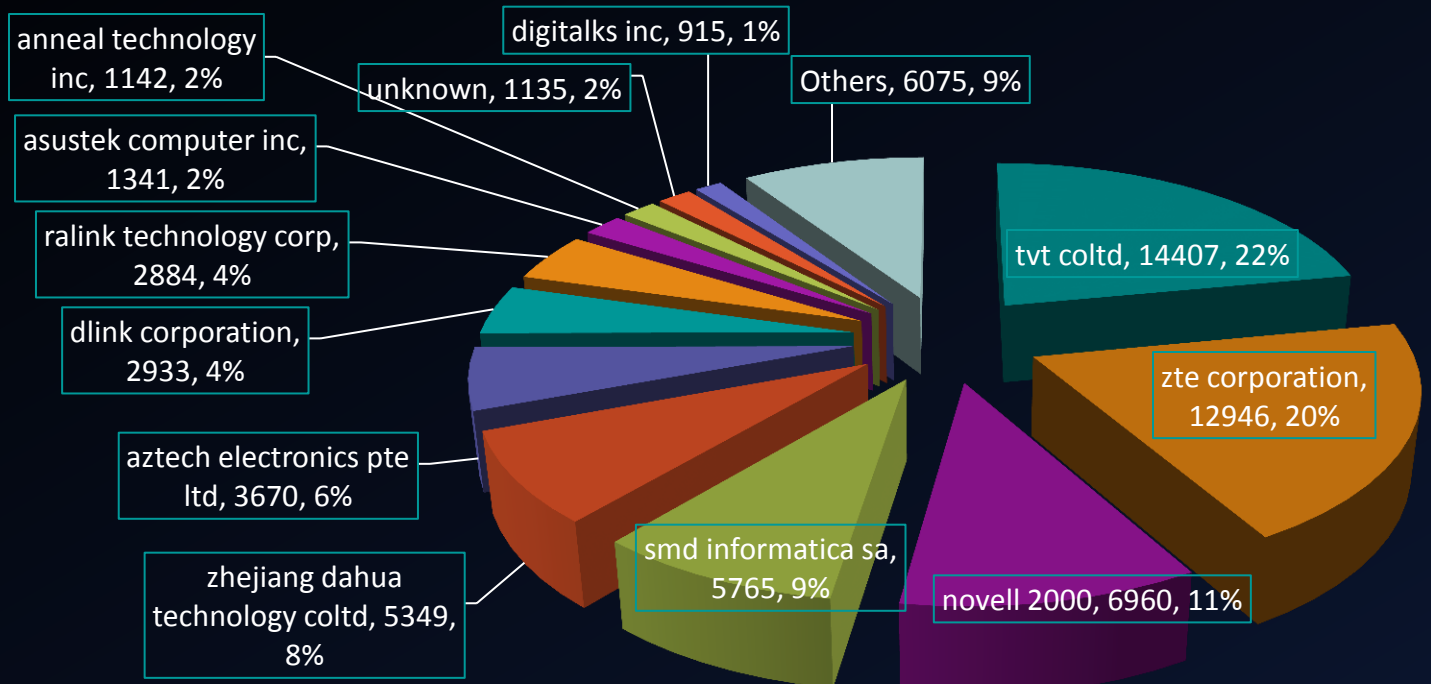
- When pointing an accusing finger at a company it is good idea to have verifiable, solid and reliable information.
  - Can't say 'Shenzhen Gongjing' has 131616 compromised devices
  - The method used to combine the IEEE list (lower case and remove special characters) can miss combining certain companies.
    - It's better to under-estimate how prevalent a company is.
    - Over estimating can create loss of reputation/discarding of results.
- Since the manufacturer field in the data has been updated with the new results, no need to look at Wireshark again.
  - For the new findings, manufacturer was changed from "unknown" to the correct one ("novell 2000", or "realtek (uptech? also reported)" or "xerox corporation")

Manufacturers - Statistics

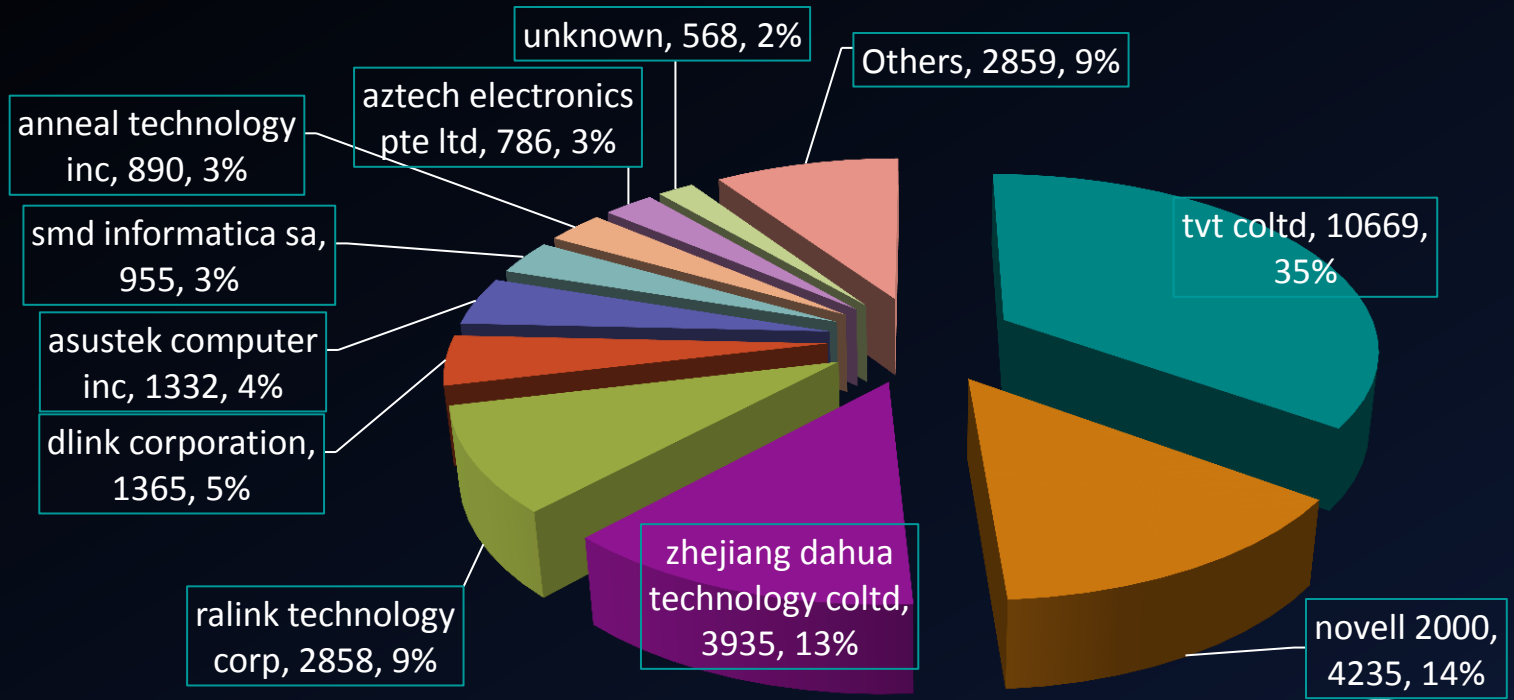
Distribution of Device Manufacturers – Worldwide - Final



Distribution of Device Manufacturers – South America



Distribution of Device Manufacturers - Brazil

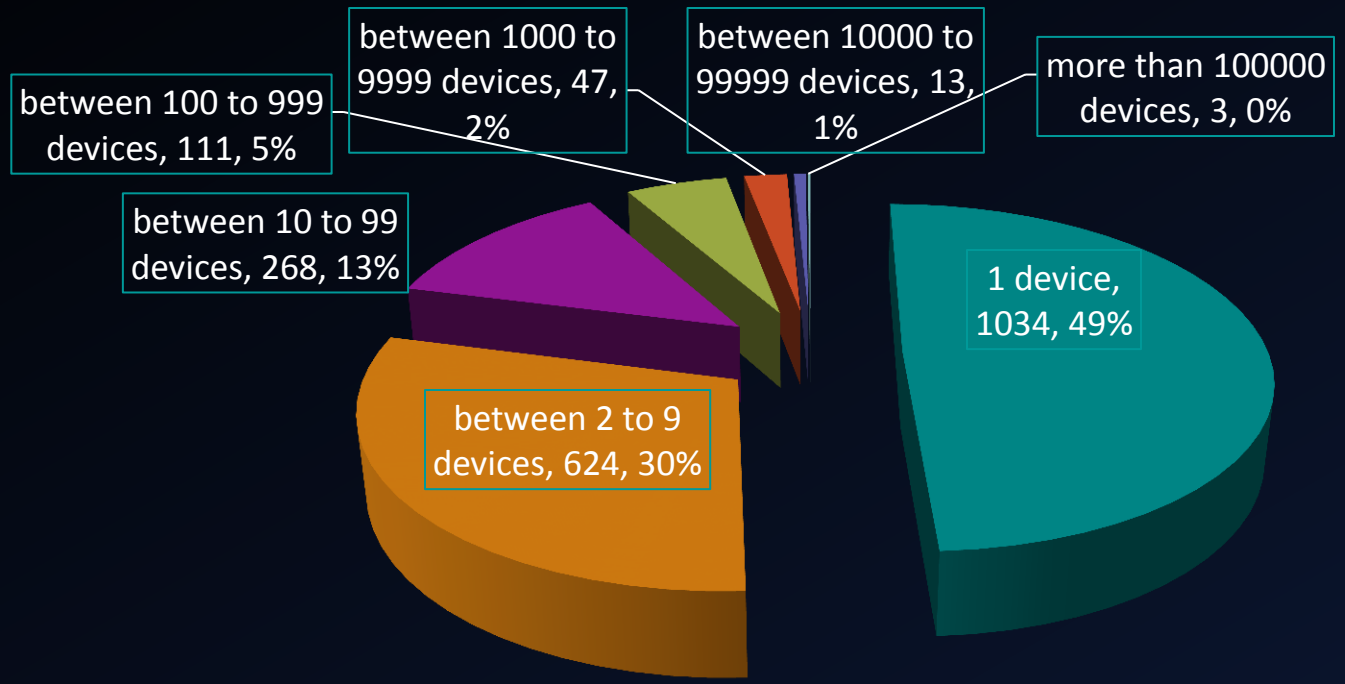


Why so many devices?

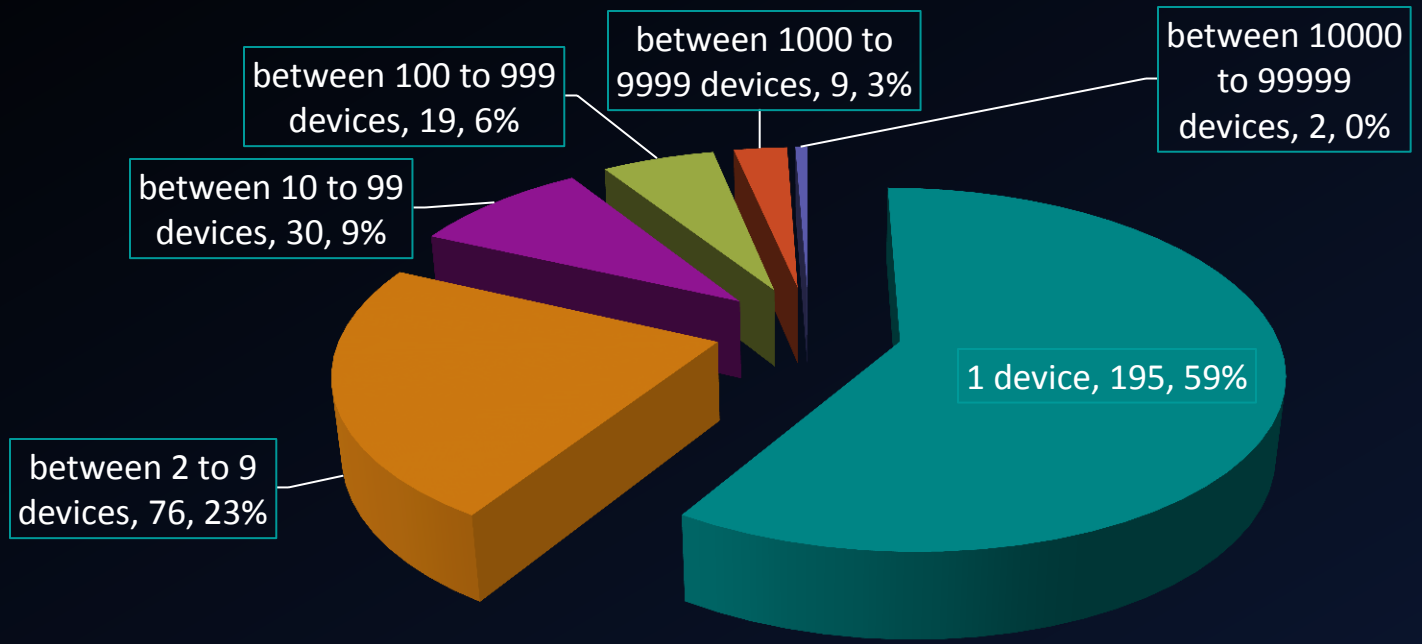
Number of Manufacturers

- Worldwide - 2100 unique device manufacturers
  - Can only see 15 in the Graph – rest in the “Others” category
- South America – 331 unique device manufacturers
  - Can only see 11 in the Graph – rest in the “Others” category
- Brazil – 204 unique device manufacturers
  - Can see 9 in the Graph – rest in the “Others” category
- Above does not count “unknown” manufacturers or “Others” category.
- Why are the rest of the manufacturers so small?
- More importantly why are some SO BIG!

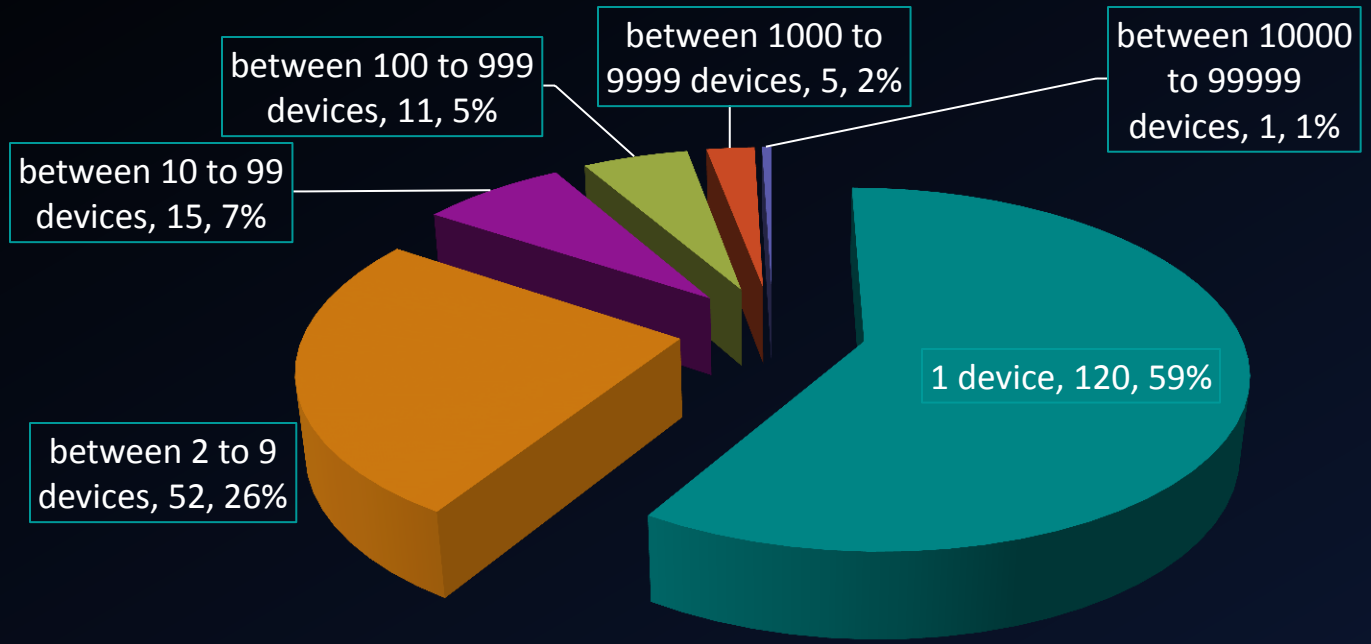
Worldwide - Number of Manufacturers by Number of Records



South America - Number of Manufacturers by Number of Records



Brazil - Number of Manufacturers by Number of Records



So why so many devices?

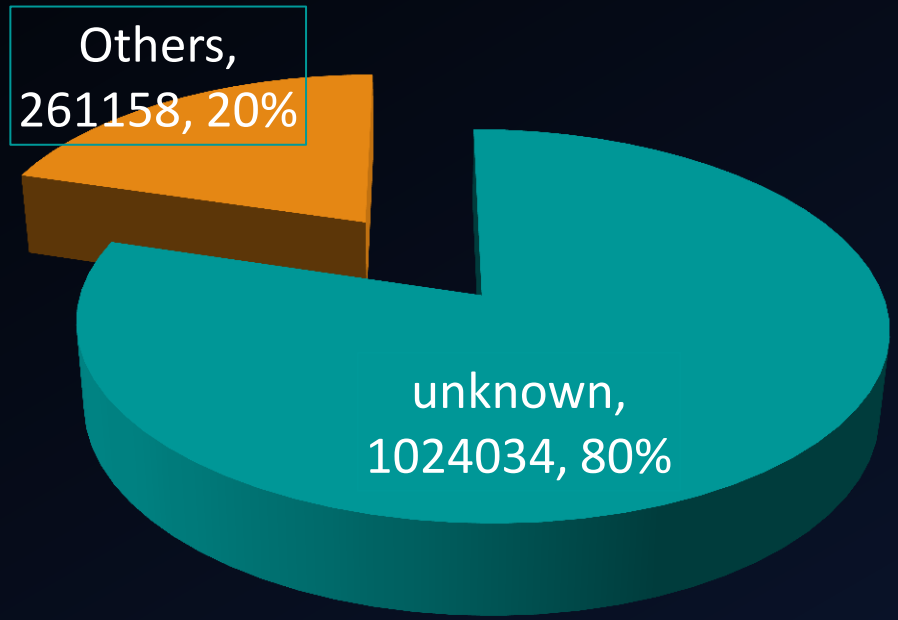
- Given the prominence of certain manufacturers, it seems obvious that most devices in the data are not because of ‘stupid’ people
- Certain devices by certain manufacturers may:
  - not allow the change of default logins for telnet
  - Have a ‘backdoor’ hardcoded with default credentials perhaps to allow for remote diagnostics (ISPs could have requested this!)
  - Lack of documentation that there is even a telnet server running on it!
    - what device wouldn’t you bother looking for an open telnet port?
  - Require devices to have Internet Reachable IP to benefit from full functionality of the product (i.e. remote viewing of CCTV camera)

Progress of presentation

- Data fields analysed so far:
  - Countries
  - Manufacturers (& MAC Addresses)
- Still to have a look:
  - Uname – Very quick look next
  - RAM – Up next
  - CPU Info – Not analysed due to inconsistency of the field
  - IP Addresses – at the end for a scary ending

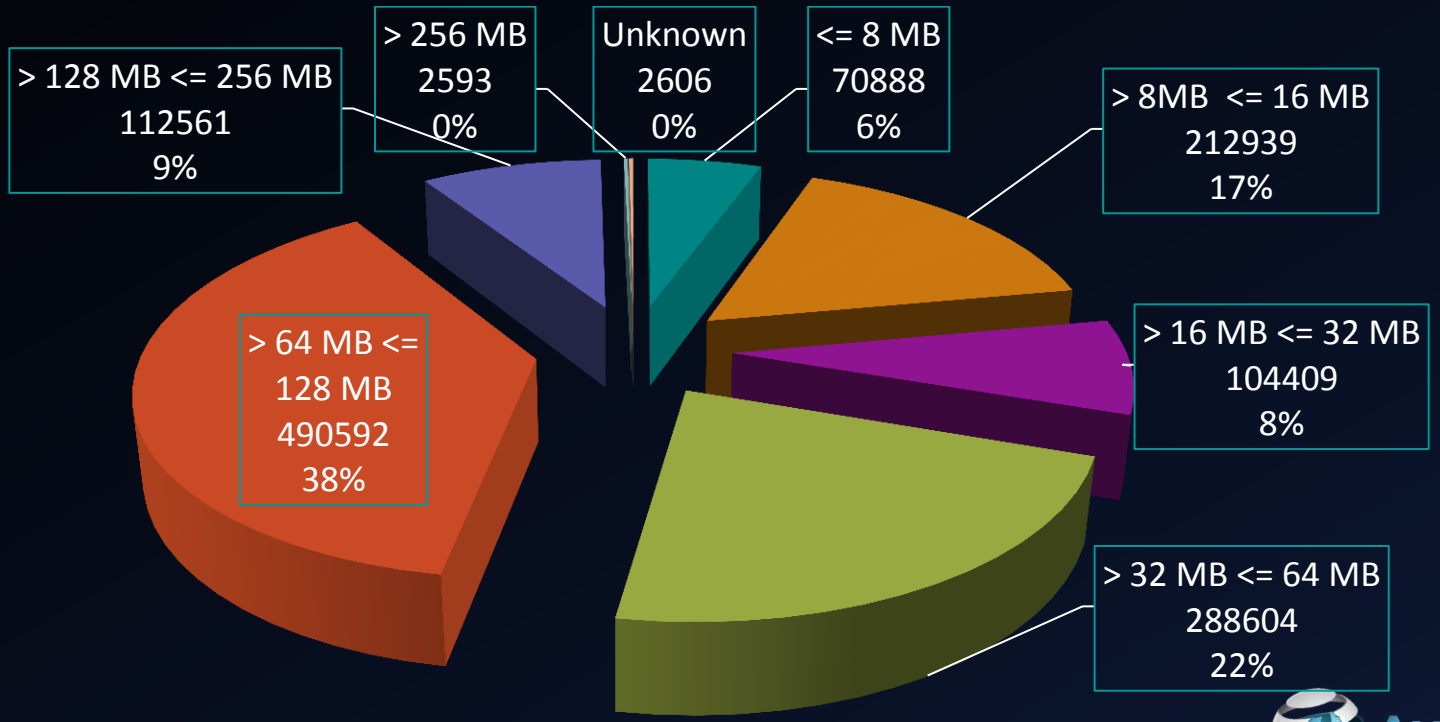
Uname

Worldwide Distribution of ‘uname’



RAM

Worldwide Devices – RAM Distribution

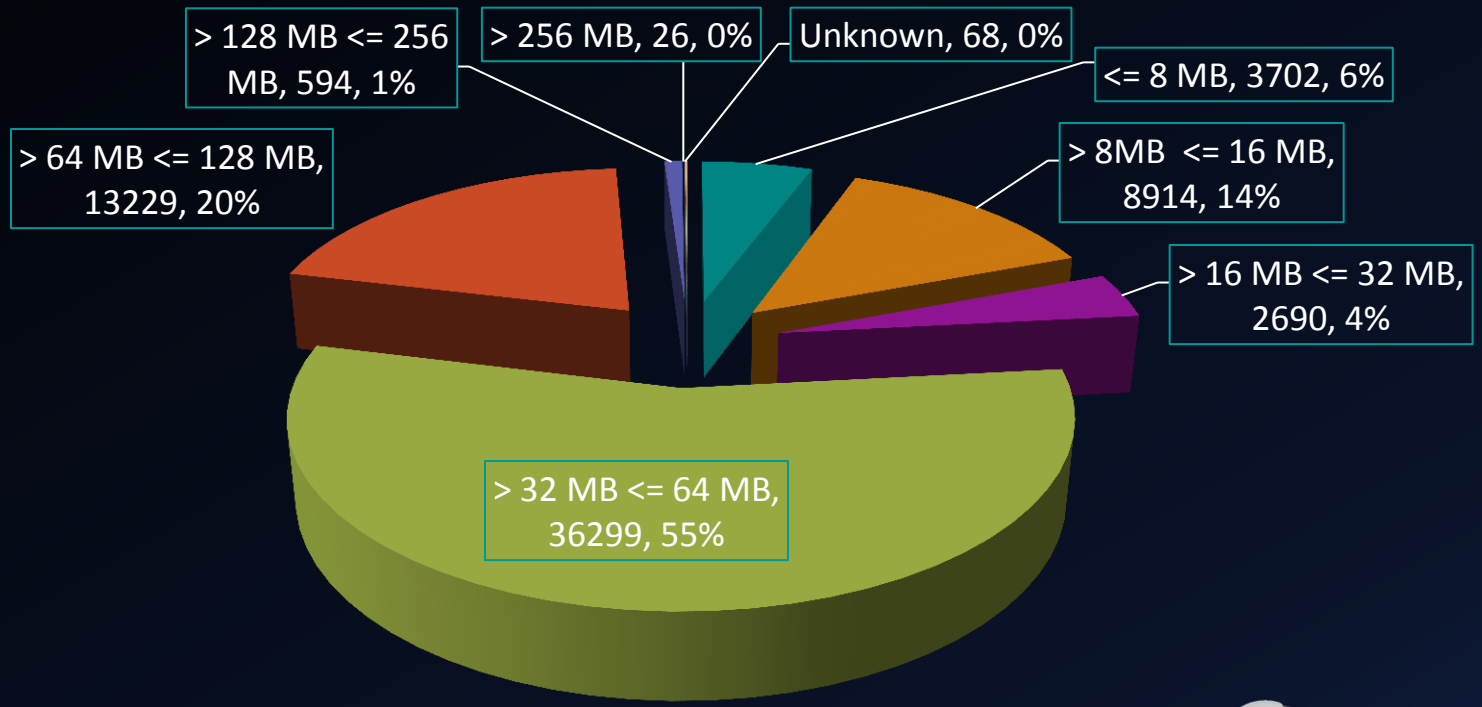


AusCERT

Worldwide Devices - RAM Statistics

Description	Value
Unique RAMs	3,880 different RAM sizes
Lowest RAM	5,488 kilobytes (5.35 MB) – 1 device in Germany
2nd Lowest RAM	5,688 kilobytes (5.55 MB) – 1 device in USA
Highest RAM	4,828,263,435 kilobytes (4.49 TB) – 1 device in China
2nd Highest RAM	1,000,000,000 kilobytes (0.93 TB) – 5 in China, 1 in Ukraine
Most common	11,500 kilobytes (11.2 MB) – 98,947 devices (7.7%)
2nd Most common	124,620 kilobytes (121.7 MB) – 96,543 devices (7.5%)

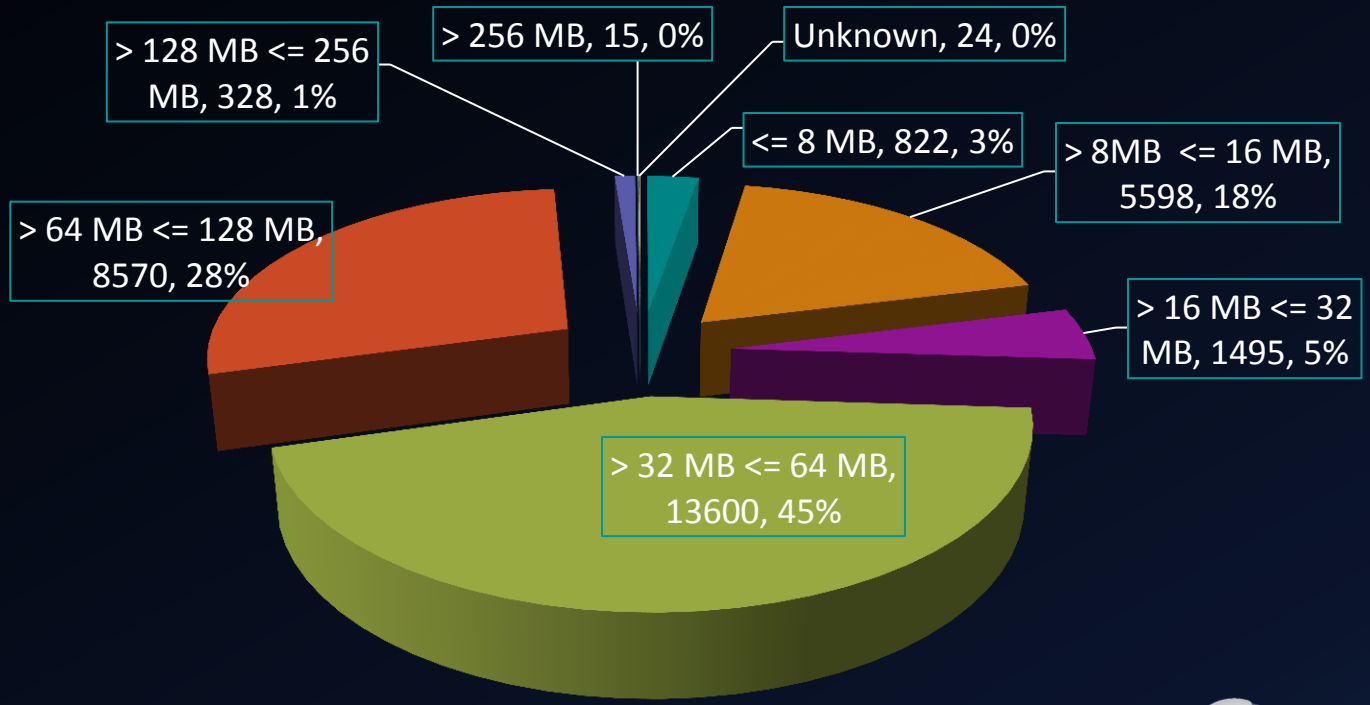
South American Devices – RAM Distribution



South American Devices – RAM Statistics

Description	Value
Unique RAMs	413 different RAM sizes
Lowest RAM	6,072 kilobytes (5.93 MB) – 1 device in Brazil
2nd Lowest RAM	6,268 kilobytes (6.12 MB) – 4 devices in Argentina and 1 device in Bolivia
Highest RAM	2,074,768 kilobytes (1.98 GB) – 1 device in Colombia
2nd Highest RAM	1,810,172 kilobytes (1.73 GB) – 1 device in Brazil
Most common	58,620 kilobytes (57.25 MB) – 12,858 devices (19.6%)
2nd Most common	61,436 kilobytes (59.996 MB) – 6,766 devices (10.3%)

Brazilian Devices – RAM Distribution



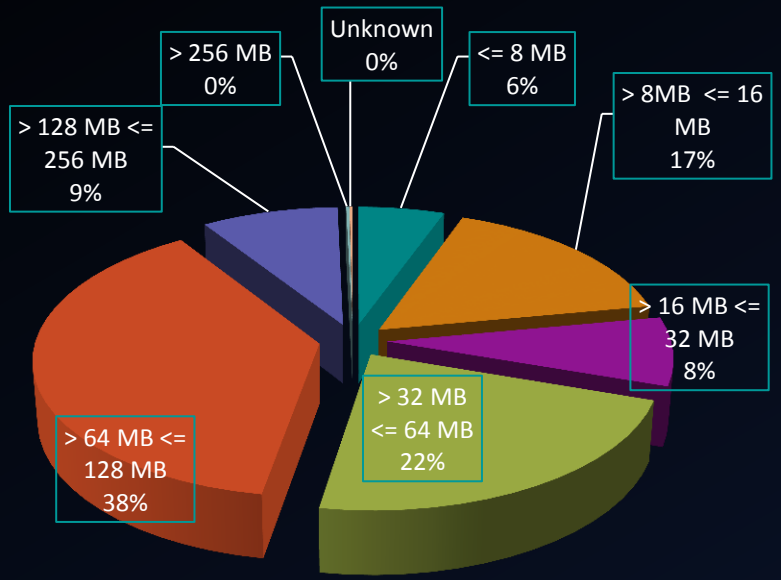
Brazilian Devices – RAM Statistics

Description	Value
Unique RAMs	270 different RAM sizes
Lowest RAM	6,072 kilobytes (5.93 MB) – 1 device
2nd Lowest RAM	6,300 kilobytes (6.15 MB) – 780 devices
Highest RAM	1,810,172 kilobytes (1.73 GB) – 1 device
2nd Highest RAM	1,036,676 kilobytes (0.99 GB) – 1 device
Most common	61,436 kilobytes (59.996 MB) – 4,885 devices (16%) – 4884 of them exact same device made by TVT coltd
2nd Most common	92,736 kilobytes (90.56 MB) – 2,288 devices (7.5%)

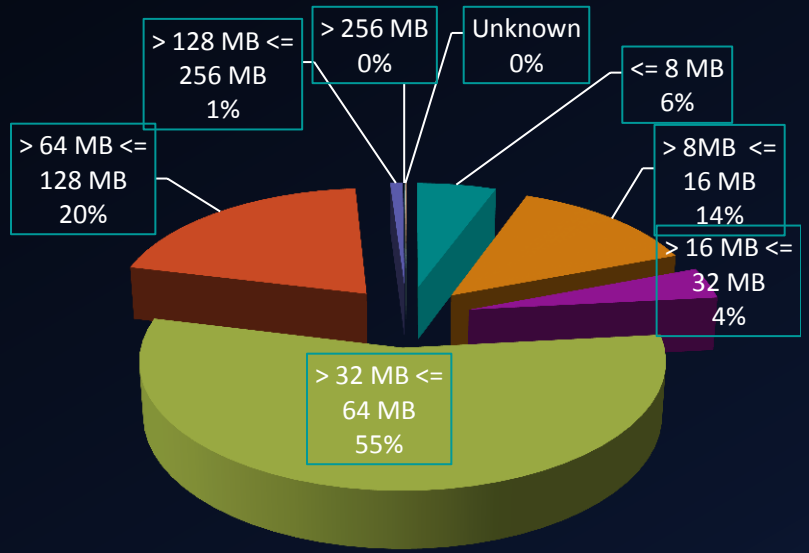
Identifying device by TVT Co. LTD.

- RAM = 61,436 kilobytes
- Manufacturer = TVT Co. LTD.
- MAC Address begins with: 00:18:ae:([21-29] or 2a or 2b)
- Product information on Official website contain **no** RAM in specs: http://en.tvt.net.cn
- Not enough info to identify?!

Worldwide RAM vs South American RAM

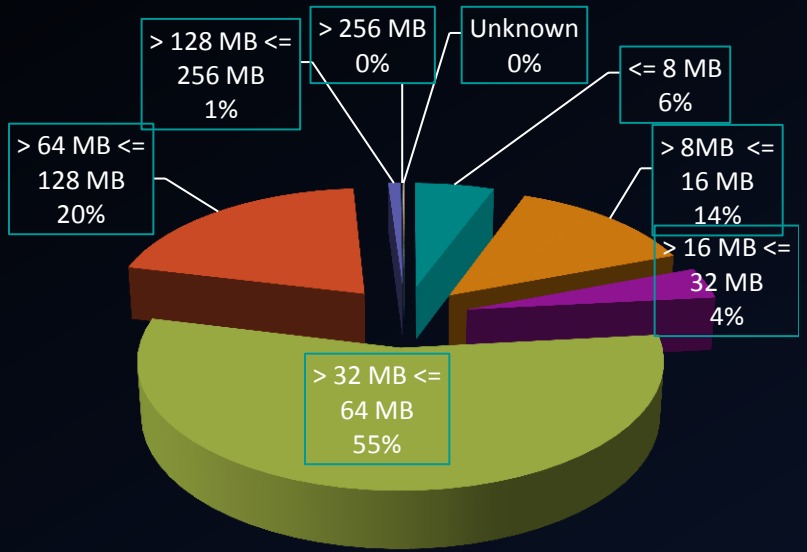


Worldwide

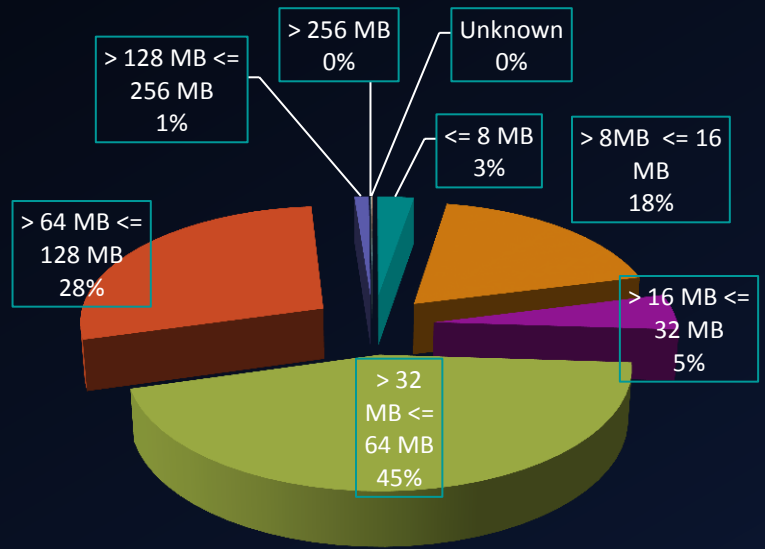


South America

South American RAM vs Brazilian RAM

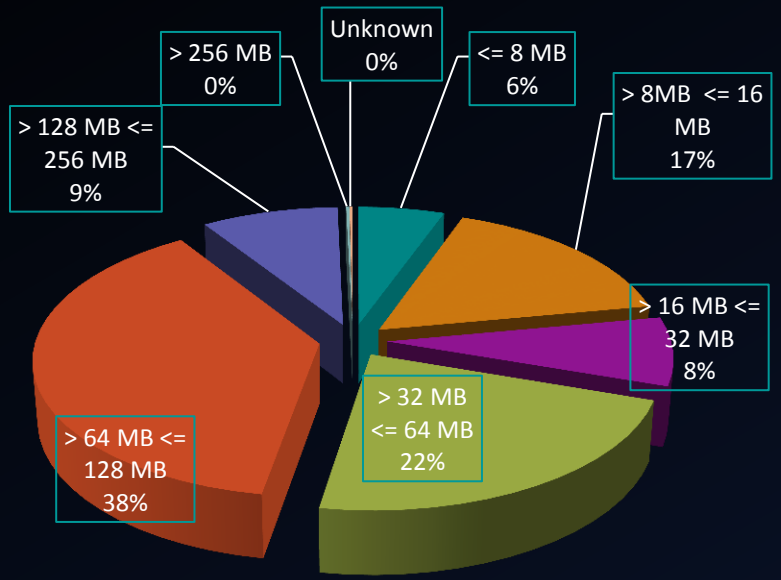


South America

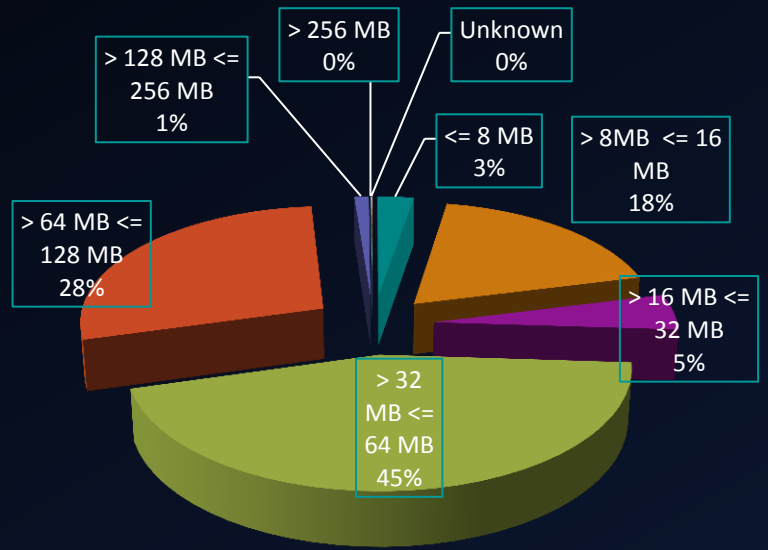


Brazil

Worldwide RAM vs Brazilian RAM



Worldwide



Brazil

Scary Statistics

Scary Stats – How easy to find a device?

- We can calculate how easy it would be for someone to find a vulnerable device with this simple equation:

$$= \frac{No.of\ infected\ devices\ for\ the\ region}{No.of\ Allocated\ C\ class\ IP\ ranges\ for\ the\ region}$$

$$= Infected\ devices\ per\ IP\ range\ for\ the\ region$$

Clarification on Calculations

- IP Allocations have changed over time
  - World/South America or Brazil did not have as many IPs allocated during the initial formation of the Carna Botnet as it does now
- Old IP allocation statistics from RIRs were used in these calculations
  - Allocated IP ranges as of 1st December 2012 were used for calculations to get an accurate idea of infection ratio around the time the Carna Botnet was formed
- Rations are assumed to be a good approximation for now as well
  - assuming that the rate of growth of allocated IPs is directly comparable to rate of growth of vulnerable devices added to the Internet over time

Worldwide - How easy to find a device?

$$\frac{No.of\ infected\ devices\ for\ the\ World}{No.of\ Allocated\ /24\ IP\ ranges\ for\ the\ World} = \frac{1{,}285{,}192}{13{,}587{,}587}$$

- ~0.095 device per /24 IP range
- ~9.46 devices per 100 C class ranges
- Average 1 vulnerable device every ~10.57 subnet
- Average 1 vulnerable device every ~2706 IPs
- Scanning 10 IPs/sec would take ~4 minutes 31 seconds to find a device
- No. of Allocated /24 IP ranges for the world deduced by adding all allocated ranges by each of the Regional Registries as of 1 December 2012

South America - How easy to find a device?

$$\frac{No.\,of\ infected\ devices\ for\ South\ America}{No.\,of\ Allocated\ /24\ IP\ ranges\ for\ South\ America} = \frac{65{,}522}{399{,}232}$$

- ~0.164 device per /24 IP range
- ~1.64 devices per 10 C class ranges
- Average 1 vulnerable device every ~6.09 subnets
- Average 1 vulnerable device every ~1,559.83 IPs
- Scanning 10 IPs/sec would take ~2 minutes 36 seconds to find a device
- No. of Allocated /24 IP ranges for South America deduced by adding all allocated IP ranges for each country in South America as of 1 Dec. 2012

Europe - How easy to find a device?

$$\frac{No.of\ infected\ devices\ for\ Europe}{No.of\ Allocated\ /24\ IP\ ranges\ for\ Europe} = \frac{167{,}320}{2{,}958{,}803.4}$$

- ~0.0565 device per /24 IP range
- ~5.65 devices per 100 C class ranges
- Average 1 vulnerable device every ~17.68 subnets
- Average 1 vulnerable device every ~4,526.98 IPs
- Scanning 10 IPs/sec would take ~7 minutes 32 seconds to find a device
- No. of Allocated /24 IP ranges for Europe deduced by adding all allocated IP ranges for each country in Europe as of 1 December 2012

Oceania - How easy to find a device?

$$\frac{No.\,of\ infected\ devices\ for\ Oceania}{No.\,of\ Allocated\ /24\ IP\ ranges\ for\ Oceania} = \frac{2{,}623}{216{,}405}$$

- ~0.012 device per /24 IP range
- ~1.21 devices per 100 C class ranges
- Average 1 vulnerable device every ~82.5 subnets
- Average 1 vulnerable device every ~21,120 IPs
- Scanning 10 IPs/sec would take ~35 minutes 12 seconds to find a device
- No. of Allocated /24 IP ranges for Oceania deduced by adding all allocated IP ranges for each country in Oceania as of 1 December 2012

Asia - How easy to find a device?

$$\frac{No.\,of\ infected\ devices\ for\ Asia}{No.\,of\ Allocated\ /24\ IP\ ranges\ for\ Asia} = \frac{1{,}006{,}634}{3{,}260{,}028}$$

- ~0.309 device per /24 IP range
- ~3.09 devices per 10 C class ranges
- Average 1 vulnerable device every ~3.24 subnets
- Average 1 vulnerable device every ~829 IPs
- Scanning 10 IPs/sec would take ~1 minute 23 seconds to find a device
- No. of Allocated /24 IP ranges for Asia deduced by adding all allocated IP ranges for each country in Asia as of 1 December 2012

All Continents

Continent Name	Infected Devices	Allocated IPs	Allocated C Classes	Infected devices per C class	1 device per how many C classes?	1 device per how many Ips?	Seconds to Find a device
Asia	1,006,634	834,567,136	3,260,027.88	0.3088	3.24	829.07	82.91
South America	65,522	102,203,392	399,232.00	0.1641	6.09	1,559.83	155.98
Europe	167,320	757,453,664	2,958,803.38	0.0565	17.68	4,526.98	452.7
Africa	7,923	51,335,680	200,530.00	0.0395	25.31	6,479.33	647.93
Oceania	2,623	55,399,680	216,405.00	0.0121	82.5	21,121.11	2,112.11
North America	35,139	1,677,457,280	6,552,567.50	0.0054	186.48	47,739.53	4,773.95
Anonymous Proxy and Satelite Provider	31	-	-	-	-	-	-

Brazil - How easy to find a device?

$$\frac{No.\,of\ infected\ devices\ for\ Brazil}{No.\,of\ Allocated\ /24\ IP\ ranges\ for\ Brazil} = \frac{30{,}452}{214{,}278}$$

- ~0.14 device per /24 IP range
- ~1.42 devices per 10 C class ranges
- Average 1 vulnerable device every ~7.035 subnets
- Average 1 vulnerable device every ~1,801 IPs
- Scanning 10 IPs/sec would take ~3 minutes to find a device
- No. of Allocated /24 IP ranges for Brazil deduced by adding all allocated IP ranges for Brazil as of 1 December 2012.
- Infection Ration Rank of 42 out of 197.

Australia - How easy to find a device?

$$\frac{No.\,of\ infected\ devices\ for\ Australia}{No.\,of\ Allocated\ /24\ IP\ ranges\ for\ Autralia} = \frac{1{,}614}{186{,}620}$$

- ~0.0086 device per /24 IP range
- ~0.86 devices per 100 C class ranges
- Average 1 vulnerable device every ~115.6 subnets
- Average 1 vulnerable device every ~29,600 IPs
- Scanning 10 IPs/sec would take ~49.3 minutes to find a device
- No. of Allocated /24 IP ranges for Australia deduced by adding all allocated IP ranges for Australia from APNIC as of 1 December 2012.
- Infection Ration Rank of 175 out of 197.

India - How easy to find a device?

$$\frac{No.of\ infected\ devices\ for\ India}{No.of\ Allocated\ /24\ IP\ ranges\ for\ India} = \frac{58{,}766}{135{,}984}$$

- ~0.4322 device per /24 IP range
- ~4.32 devices per 10 C class ranges
- Average 1 vulnerable device every ~2.31 subnets
- Average 1 vulnerable device every ~592 IPs
- Scanning 10 IPs/sec would take ~59 seconds to find a device
- No. of Allocated /24 IP ranges for India deduced by adding all allocated IP ranges for India from APNIC as of 1 December 2012
- Infection Ration Rank of 21 out of 197.

China - How easy to find a device?

$$\frac{No.\,of\ infected\ devices\ for\ China}{No.\,of\ Allocated\ /24\ IP\ ranges\ for\ China} = \frac{720{,}141}{1{,}289{,}054}$$

- ~0.5587 device per /24 IP range
- ~5.59 devices per 10 C class ranges
- Average 1 vulnerable device every ~1.79 subnets
- Average 1 vulnerable device every ~458 IPs
- Scanning 10 IPs/sec would take ~45 seconds to find a device
- No. of Allocated /24 IP ranges for China deduced by adding all allocated IP ranges for China from APNIC as of 1 December 2012.
- Infection Ration Rank of 18 out of 197.

Other South American Countries and Their Rank

Countries	Infected Devices	Infection ratio out of 197	Seconds to Finds
Brazil	30452	41	180.14
Uruguay	13173	3	[Surprise]
Argentina	7880	49	205.39
Colombia	6454	38	158.55
Chile	2963	62	266.17
Venezuela	1593	80	366.79
Peru	1265	50	207.8
Bolivia	690	27	88.56
Curaçao	490	1	[Surprise]
Ecuador	393	91	506.53
Paraguay	126	93	526.02
Guyana	28	37	149.94
French Guiana	11	7	[Surprise]
Suriname	4	152	1,945.60

Uruguay - How easy to find a device?

$$\frac{No.of\ infected\ devices\ for\ Uruguay}{No.of\ Allocated\ /24\ IP\ ranges\ for\ Uruguay} = \frac{13{,}173}{4{,}627}$$

- ~2.84 devices per /24 IP range
- Average 1 vulnerable device every ~0.35 subnets
- Average 1 vulnerable device every ~89.92 IPs
- Scanning 10 IPs/sec would take ~9 seconds to find a device
- No. of Allocated /24 IP ranges for Uruguay deduced by adding all allocated IP ranges for Uruguay as of 1 December 2012.
- Infection Ration Rank of 3 out of 197.

Curaçao - How easy to find a device?

$$\frac{No.of\ infected\ devices\ for\ Curaçao}{No.of\ Allocated\ /24\ IP\ ranges\ for\ Curaçao} = \frac{490}{4}$$

- ~122.5 device per /24 IP range
- Average 1 vulnerable device every ~0.01 subnets
- Average 1 vulnerable device every ~2.09 IPs
- Scanning 10 IPs/sec would take ~0.21 seconds to find a device
- No. of Allocated /24 IP ranges for Curaçao deduced by adding all allocated IP ranges for Curaçao as of 1 December 2012.
- Infection Ration Rank of 1 out of 197.
- Most likely inaccurate as it's probably using IP ranges allocated to neighbouring countries.

French Guiana - How easy to find a device?

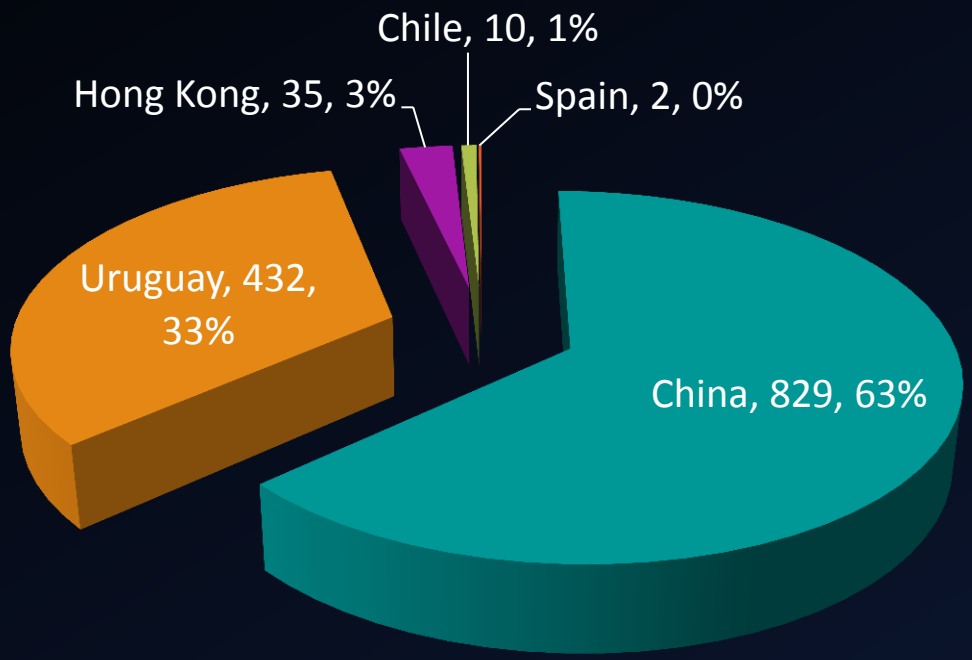
$$\frac{No.of\ infected\ devices\ for\ French\ Guiana}{No.of\ Allocated\ /24\ IP\ ranges\ for\ French\ Guiana} = \frac{11}{8}$$

- ~1.375 device per /24 IP range
- Average 1 vulnerable device every ~0.73 subnets
- Average 1 vulnerable device every ~186.18 IPs
- Scanning 10 IPs/sec would take ~18.62 seconds to find a device
- No. of Allocated /24 IP ranges for French Guiana deduced by adding all allocated IP ranges for French Guiana as of 1 December 2012.
- Infection Ration Rank of 7 out of 197.
- Hard to say if this is accurate or not. Generally if No. of allocated C lasses are too small then the infection ratio is likely to be fairly inaccurate.

Terrified now? It gets worse!

- 1308 IP ranges were found that appear in more than or equal to 260 different records.
  - Same IP range in more than 260 different device records
  - i.e. almost all IPs within these 1308 IP ranges likely to contain vulnerable devices!
  - This is not a ‘guarantee’ as devices compromised at different times
- If you were to find these 1308 IP ranges and “hog” them then you’d have a botnet of: ~327 thousand
- Which countries more prominent in these records?

Countries of IP ranges appearing in more than 259 records



Still worse news!

- So just probe these devices on port 23 to find them?
- NO, because a scan of some of the IP range from the data show almost all ranges had port 23 closed?! Faulty Data?
- Carna Botnet shutdown telnet to close port 23 as soon as it had control of the device and setup iptable rules where possible
  - Primarily to avoid interference from other botnets
  - Temporary only - settings lost on reboot
  - Other botnets won't be so nice
- If telnet is the only shell into the device then hardware reset is the best chance of ensuring a clean device

Scariest News – Light Aidra

- Open source tool ‘lightaidra’ does exactly what Carna botnet did:
  - Auto searches for telnet ports with default credentials
  - Allows you to upload your custom binary that can do anything!
    - For routers it can sniff traffic, modify traffic! Spam the world! Anything!
  - Joins IRC chat room to read latest commands
- Bad guys really don’t need to do much
- Carna detected presence of Aidra (as noted in the paper)
  - So this data might not be ALL vulnerable devices
  - However, Carna was a lot more cross-platform than lightaidra is by default
- Info: https://github.com/eurialo/lightaidra

Hope

A Glimmer of Hope

- Most embedded devices mount their partitions as read-only. tmp and other directories stored in RAM
- For most devices a reboot would lose the malware
  - May leave port 23 closed!
  - Start up scripts (if functionality exists) could've been modified to re-infect
- Almost all the time a hardware reset and/or firmware reflash will resolve the problem
  - If malware authors wanted they could interfere with re-flashing and hardware reset depending on how much control telnet allowed over the device

What have I done?

- Supplied relevant data CERTs from any country with more than 10 thousand compromised devices and some others on request etc.
  - Adds up to about 30 countries in total
- Split Australian data by ISP and provided it to them
- Contacted IEEE to contact the worst affected manufacturers
  - Reached out to manufacturers to work with us; only 1 of 23 has responded
- Presentations and Research Paper!
  - Talking to people and publication of presentations and research paper
  - Presented The Hackers Conference in Delhi, India; APNIC 36 in Xian, China. Will present on the South American data at BlackHat in Sao Paulo next week.

Research Paper

- Data wasn't provided on a silver platter ready for analysis
- Checking for consistency both internally and externally was done
- Manufacturer field was re-derived
- Assumptions had to be made; duplicates were removed
- Efforts to check for accuracy were made
- Contains a complete list of countries, infection ratios and manufacturers
- All of this plus more is covered in detail in the Research paper at: http://bit.ly/carna-paper

Problem needs attention NOW

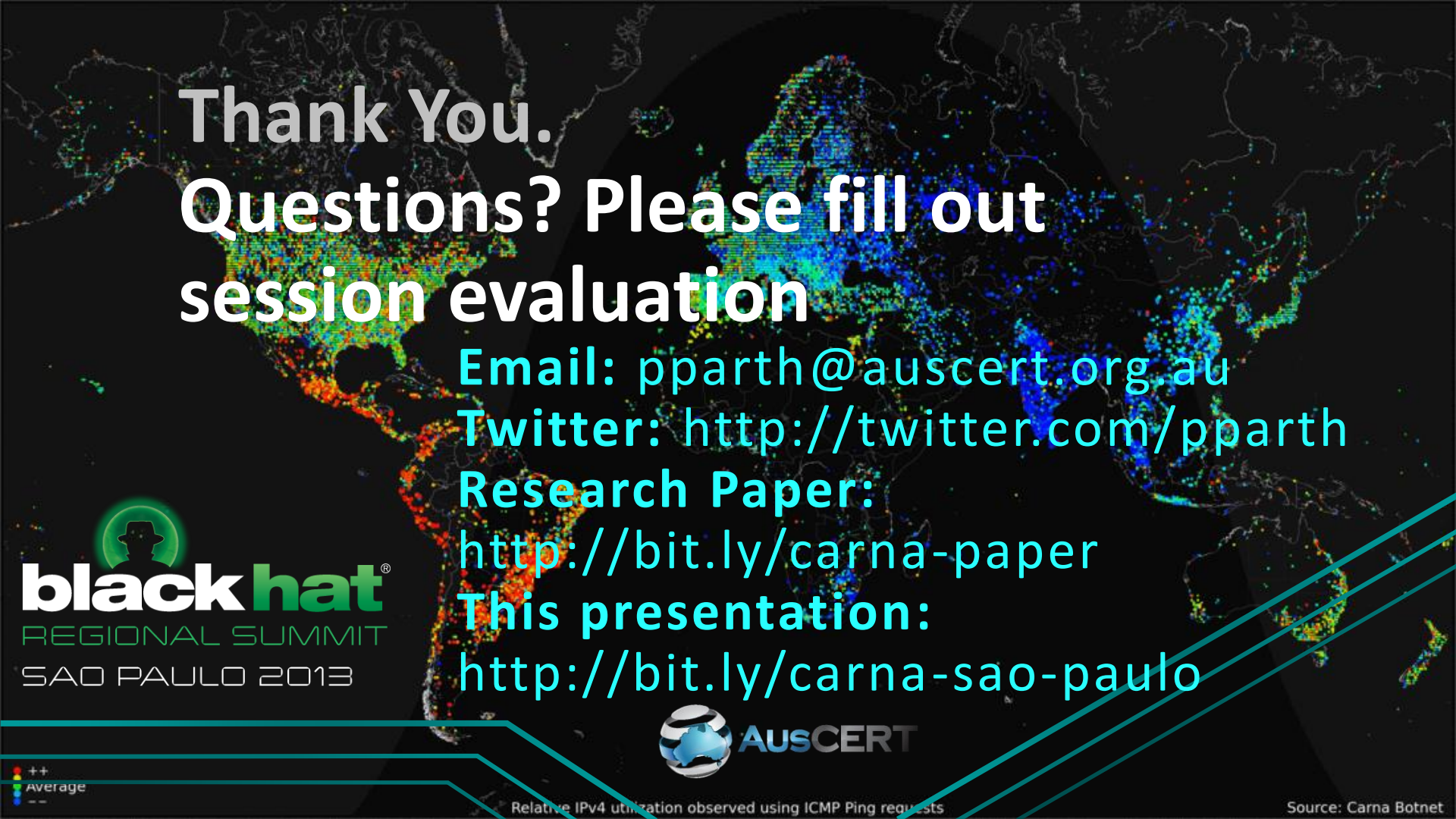
- Devices behind NAT not in the data
  - Researcher did not scan internal network when a router was compromised.
  - So the number of actually vulnerable devices likely to be a lot bigger
- With IPv4 to IPv6 transition happening now, this is the time to make sure such devices are secure by default.
  - Specially since NAT “protection” will not be available on IPv6 and
  - Bad router firewalls might expose even more devices then visible in this data set.

Public Awareness

- Seemingly harmless/small threat is real and big!
- Huge problem for economy like Brazil. Can have massive financial impacts if foreign powers or malicious agents launch a coordinated attack.
- Awareness of problem is the first step
- Participation of diverse range of players from the industry required
- Awareness to be raised with public/manufacturers/ISPs-selling-the-devices and governments on the Carna Botnet problem
- This presentation is one of the many steps required to tackle the problem as a whole
- Please spread the word!

I can’t do it alone. What can you do?

- It’s a long and hard battle because no ‘easy’ or ‘quick’ solutions
  - Read my detailed Research Paper & Re-look at these slides online (links on next slide)
  - Tell people: family, relatives, employer, friends, colleagues
  - Secure your devices! Secure others’ devices
- Do you know anyone in a position who can help?
- Help influence government, companies & manufacturers into ensuring devices used and sold by them are secure by default
- Does your ISP sell a device that’s vulnerable to this by default? Check!
- You are the passionate IT security professionals! Help me fight this battle!
- Network operators can block port 23 at the WAN and allow ‘opt-out’ for customers who want it enabled.
- Have other ideas on tackling the problems? Contact me! (email on next slide)


Thank You.
Questions? Please fill out session evaluation
Email: pparth@auscert.org.au
Twitter: http://twitter.com/pparth
Research Paper:
http://bit.ly/carna-paper
This presentation:
http://bit.ly/carna-sao-paulo
black hat
REGIONAL SUMMIT
SAO PAULO 2013
AUSCERT
++
Average
--
Relative IPv4 utilization observed using ICMP Ping requests
Source: Carna Botnet