

Blackhole Networks

an Underestimated Source for Information Leaks



CIRCL
Computer Incident
Response Center
Luxembourg

Alexandre Dulaunoy CIRCL -
TLP:WHITE

Team CIRCL - Team Restena

FIRST2017

Motivation and background

- IP darkspace or black hole is
 - **Routable non-used address space** of an ISP (Internet Service Provider),
 - incoming traffic is unidirectional
 - and **unsolicited**.
- Is there any traffic in those darkspaces?
- If yes, what and why does it arrive there?
 - And **on purpose** or **by mischance**?
- What's the security impact?
- What are the security recommendations?

The infinite value of crap

4 years in the life of a printer

from a series of packets hitting our darkspace

Printer sending syslog to the IP darkspace

2014-03-12 18:00:42

 SYSLOG lpr.error printer: offline
 or intervention needed

2014-03-23 21:51:24.985290

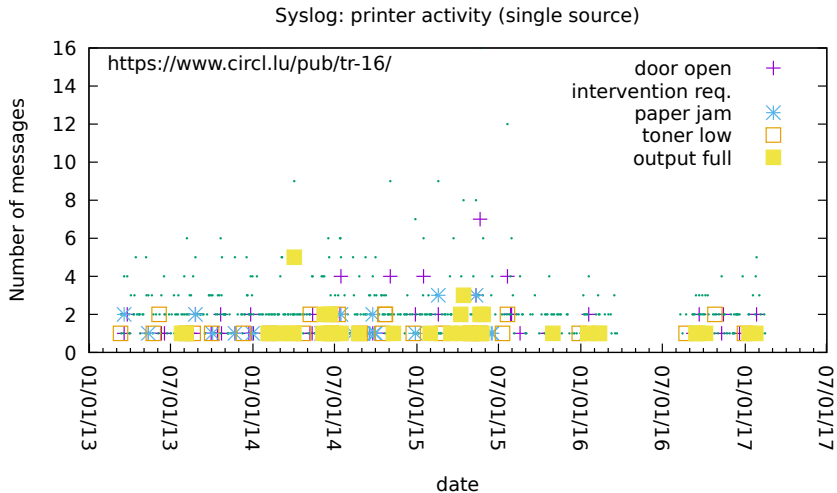
 SYSLOG lpr.error printer: paper out

...

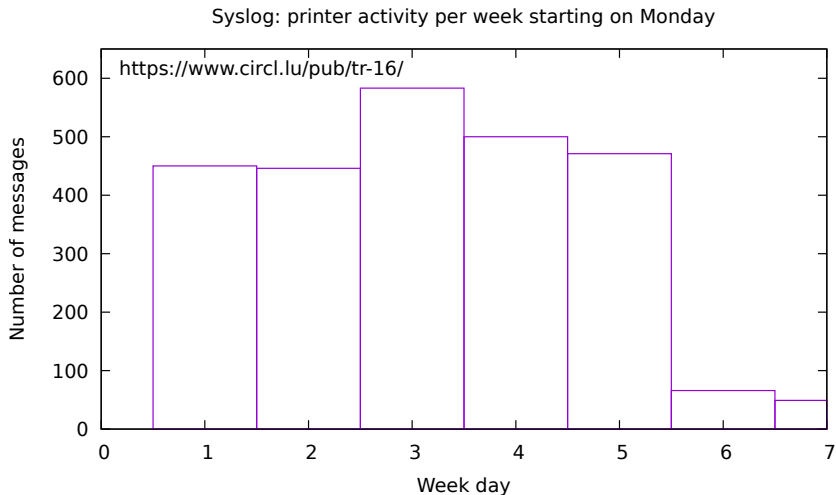
2014-08-06 19:14:57.248337

 SYSLOG lpr.error printer: paper jam

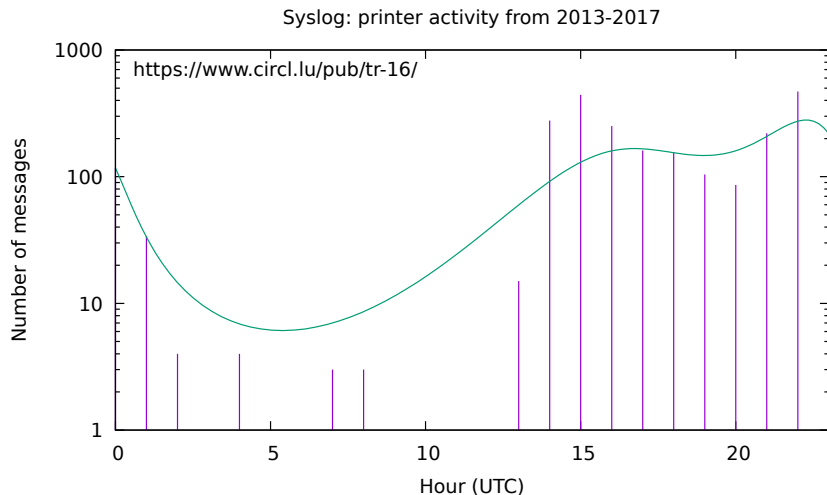
4 years in the life of a printer



Business days based on the printer activity



Printer activity and business hours



Origin of traffic in the black hole

- Attackers (and researchers) scan networks to find vulnerable systems (e.g. SSH brute-force)
- Backscatter traffic (e.g. from spoofed DoS)
- Self-replicating code using network as a vector (e.g. conficker, residual worms)
- Badly configured devices especially embedded devices (e.g. printers, server, routers)
 - → **Our IP darkspace is especially suited for spelling errors from the RFC1918 (private networks) address space**

Why is there traffic

Typing/Spelling errors with RFC1918 networks

- While typing an IP address, different error categories might emerge:

Hit wrong key	19 2 .x.z.y →	19 3 .x.y.z
	172.x.y.z	1 5 2.x.y.z
Omission of number	1 9 2.x.y.z →	12.x.y.z
Doubling of keys	10.a.b.c →	10 0 .a.b.c

Research activities related to spelling errors

Spelling errors apply to text but also network configuration

- 34% omissions of 1 character
 - Example: Network → Netork
- 23% of all errors happen on 3rd position of a word
 - Example: Text → Test)
- 94% spellings errors are single errors in word
 - And do not reappear

References

- Pollock J. J. and Zamora A., Collection and characterization of spelling errors in scientific and scholarly text. J. Amer. Soc. Inf. Sci. 34, 1, 51 58, 1983.
- Kukich K., Techniques for automatically correcting words in text. ACM Comput. Surv. 24, 4, 377-439, 1992.

How to write a Gartner market survey?

*What are the most common
antivirus software?*

by using the DNS queries hitting your darkspace

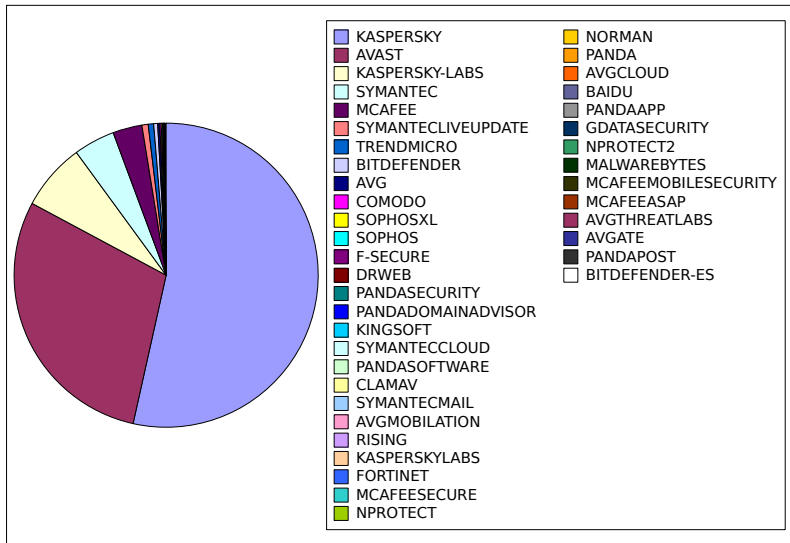
Sample subset of DNS queries towards antivirus vendors domains

```
1 0.0.0.16a8.20ae.2f4a.400.7d.igkhab8lsrnzhj726ngu8gbsev.  
   avqs.mcafee.com A INET 127.161.0.128  
2 0.0.0.16a8.20ae.2f4a.400.7d.sdszgsg5a6j516p9nui9jfz5mj.  
   avqs.mcafee.com A INET 127.161.0.128  
3 40.ucp-ntfy.kaspersky-labs.com  
4 46.ucp-ntfy.kaspersky-labs.com  
5 6.ucp-ntfy.kaspersky-labs.com  
6 dnl-06.geo.kaspersky.com.<COMPANYNAME>.local  
7 shasta-mr-clean.symantec.com  
8 shasta-mrs.symantec.com  
9 shasta-nco-stats.symantec.com
```

Scripting your statistics for antivirus installations

- Extract a **list of words** from VirusTotal (antivirus products supported)
- Match the DNS queries with extracted words (e.g. be careful with fake antivirus)
- **Filter per source IP address** (or aggregated subnets) to limit the result per organisation
- Plot the number of hits per aggregated words using in a single antivirus product name

A/V Statistics from Misconfigured Resolvers

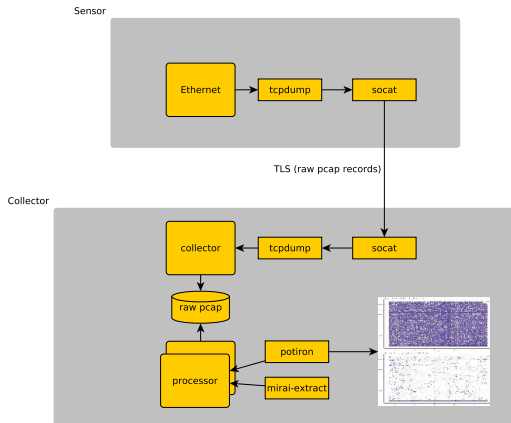


How do we collect all this crap?

by listening to the void

Collection and Analysis Framework

Collection and Analysis Framework



Collection and Analysis Framework

or to keep the collection as simple as possible

- Minimal sensor collecting IP-Darkspace networks (**close to RFC1918 address space**)
- Raw pcap are captured with the full payload
- Netbeacon¹ developed to ensure consistent packet capture
- Potiron² to normalize, index, enrich and visualize packet capture

¹<https://github.com/adulau/netbeacon/>

²<https://github.com/CIRCL/potiron>

Dataset collected and statistics

- From 2012-03-12 until Today (still active)
- More than 700 gigabytes of compressed raw pcap collected
- Constant stream of packets from two /22 network blocks
 - no day/night profile.
- Some peaks at 800kbit/s (e.g. often TCP RST from backscatter traffic but also from typographic errors)

General observations

- A large part of traffic is coming from badly configured devices (**RFC1918 spelling errors**)
 - Printers, embedded devices, routers or even server.
 - Trying to do name resolution on non-existing DNS servers, NTP or sending syslog messages.
- Even if the black hole is passive, payload of stateless UDP packets or even TCP (due to asymmetric routing on misspelled network) datagrams are present
- Internal network scanning and reconnaissance tool (e.g. internal network enumeration)

Observation per AS

Traffic seen in the darknet

N	Frequency	ASN
1	4596319	4134
2	1382960	4837
3	367515	3462
4	312984	4766
5	211468	4812
6	166110	9394
7	156303	9121
8	153585	4808
9	135811	9318
10	116105	4788

- Occurrences of activities related to the proportion of hosts in a country
- The Great Firewall of China is **not filtering leaked packets**
- Corporate AS number versus ISP/Telco AS number

How to build your "next" network reconnaissance tools?

by listening to the void

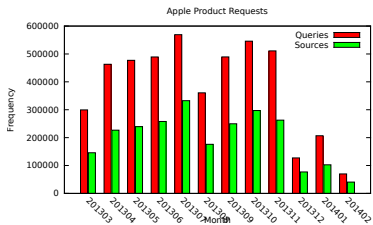
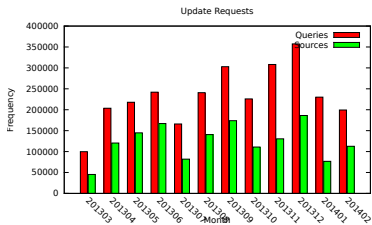
Network reconnaissance (and potential misuse): DNS

```
1 3684 _msdcs.<companyname>.local
2 1232666 time.euro.apple.com
3 104 time.euro.apple.com.<mylocaldomain>
4 122 ocsp.tcs.terena.org
5 50000+ ocsp.<variousCA>
```

- DNS queries to an incorrect nameserver could lead to major misuse
- **A single typographic error in a list of 3 nameservers is usually unnoticed**

Software Updates/Queries from Misconfigured Resolvers

- Discovering software usage (and vulnerabilities) can be easily done with passive reconnaissance
- Are the software update process ensuring the integrity of the updates?



Network Reconnaissance - A source for your smart DNS Brute-Forcer

ASTTF.NET	HELP.163.COM
ASUEGYI.INFO	HP_CLIENT1
ASUS1025C	MACBOOKAIR-CAD7
DEFAULT	MACBOOK-B5BA66
DELICIOUS.COM	MACBOOKPRO-5357
DELL	MAIL.AFT20.COM
DELL1400	S3.QHIMG.COM
DELL335873	SERVERWEB
DELL7777	SERVEUR
DELL-PC	SERVICE.QQ.COM
DELLPOP3	SMTP.163.COM

And many more ...

Building your DNS brute-forcer

- Smart DNS Brute-Forcer³⁴ uses techniques from natural language modeling with Markov Chain Models
- The processor relies on passive DNS data to generate the statistics and extract the features.
- The DNS queries seen in the **IP darkspace can be considered as a passive DNS stream** with a focus on internal network.
- Providing a unique way to create **internal DNS brute-forcers from external observations**.

³<https://www.foo.be/papers/sdbf.pdf>

⁴<https://github.com/jfrancois/SDBF>

Network Reconnaissance: NetBios Machine Types (1 week)

23	Browser Server
4	Client?
1	Client? M <ACTIVE>
21	Domain Controller
1	Domain Controller M <ACTIVE>
11	Master Browser
1	NameType=0x00 Workstation
1	NameType=0x20 Server
105	Server
26	Unknown
1	Unknown <GROUP> B <ACTIVE>
5	Unknown <GROUP> M <ACTIVE>
1322	Workstation
1	Workstation M <ACTIVE>

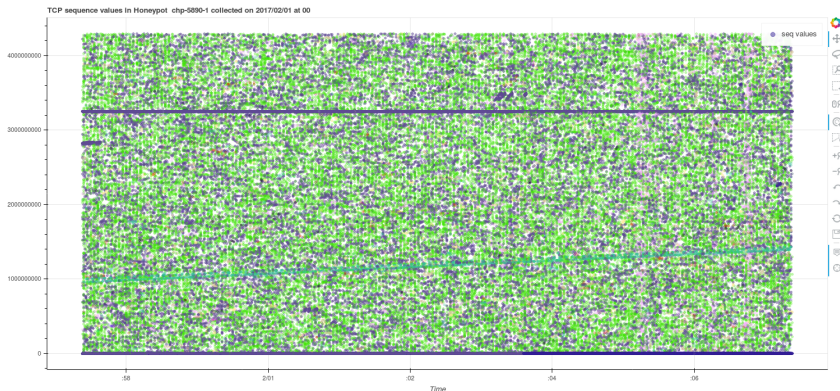
How to configure your router (without security)

Enable command logging and send the logs to a random syslog server

```
Aug 13 10:11:51 M6000-G5 command-log:[10:11:51 08-13-2012
  VtyNo: vty1  UserName: XXX IP: XXX ReturnCode: 1
  CMDLine: show subscriber interface gei-0/2/1/12.60
Aug 13 10:46:05 M6000-G5 command-log:[10:46:05 08-13-2012
  VtyNo: vty2  UserName: XXX IP: XXX  ReturnCode: 1
  CMDLine: conf t ]
Aug 13 10:46:10 M6000-G5 command-log:[10:46:10 08-13-2012
  VtyNo: vty2  UserName: XXX IP: XXX  ReturnCode: 1  CMD
Line: aaa-authentication-template 1100 ]
...
```

We will let you guess the sensitive part afterwards...

Finding origin of traffic by TCP sequence analysis



```

211 iph->id = rand_next();
212 iph->saddr = LOCAL_ADDR;
213 iph->daddr = get_random_ip();
214 iph->check = 0;
215 iph->check = checksum_generic((uint16_t *)iph, sizeof (struct iphdr));
216
217 if (i % 10 == 0)
218 {
219     tcph->dest = htons(2323);
220 }
221 else
222 {
223     tcph->dest = htons(23);
224 }
225 tcph->seq = iph->daddr;
226 tcph->check = 0;
227 tcph->check = checksum_tcpudp(iph, tcph, htons(sizeof (struct tcphdr)), sizeof (struct tcphdr));
228
229 paddr.sin_family = AF_INET;
230 paddr.sin_addr.s_addr = iph->daddr;
231 paddr.sin_port = tcph->dest;
232
233 sendto(rsck, scanner_rawpkt, sizeof (scanner_rawpkt), MSG_NOSIGNAL, (struct sockaddr *)&paddr, sizeof
234 )
---
```

Recommendations for operating an IP darkspace

- **Capture raw packets at the closest point**, don't filter, don't try to be clever, just store it as it.
- **Test your network collection mechanisms** and storage. Send test network beacons. Check the integrity, order and completeness of packets received.
- You never know in advance which features is required to distinguish a specific pattern.
- Did I mention to store **RAW PACKETS**?

Conclusions

- Security recommendations
 - **Default routing/NAT to Internet in operational network is evil**
 - Use fully qualified domain names (resolver search list is evil too)
 - Double check syslog exports via UDP (e.g. information leakage is easy)
 - Verify any default configuration with SNMP (e.g. enable by default on some embedded devices)
- Offensive usage? What does it happen if a malicious "ISP" responds to misspelled RFC1918 addresses? (e.g. DNS/NTP requests, software update or proxy request)
- Some research projects on this topic? Contact us <mailto:info@circl.lu>