

Operational return of experience

DNS Resolution Traffic Analysis Applied to Bot Detection

or how to detect even more infected computers on a big company network only with a PoC, a Co-op Engineering Student, some lines of codes and a lot of coffee...

...for about 10 000€

DZKSJEGGACJHPGFHACDGBGOGACCHFGGBGEGACEHIGJGDHMCACHHFGACNGBGJHACIG
BGGHFGACDHPGNGFGACDGPNGNGPGOGACJGOGEHFGCHFGDHEHBCACCHPGOGBGO
GOCNGPGFHDGIGPGFHIHAEDGJHCGFGCHBGDGEHJGPGOGMGBGCGOCOGFGEHDZJS

DISCLAIMER

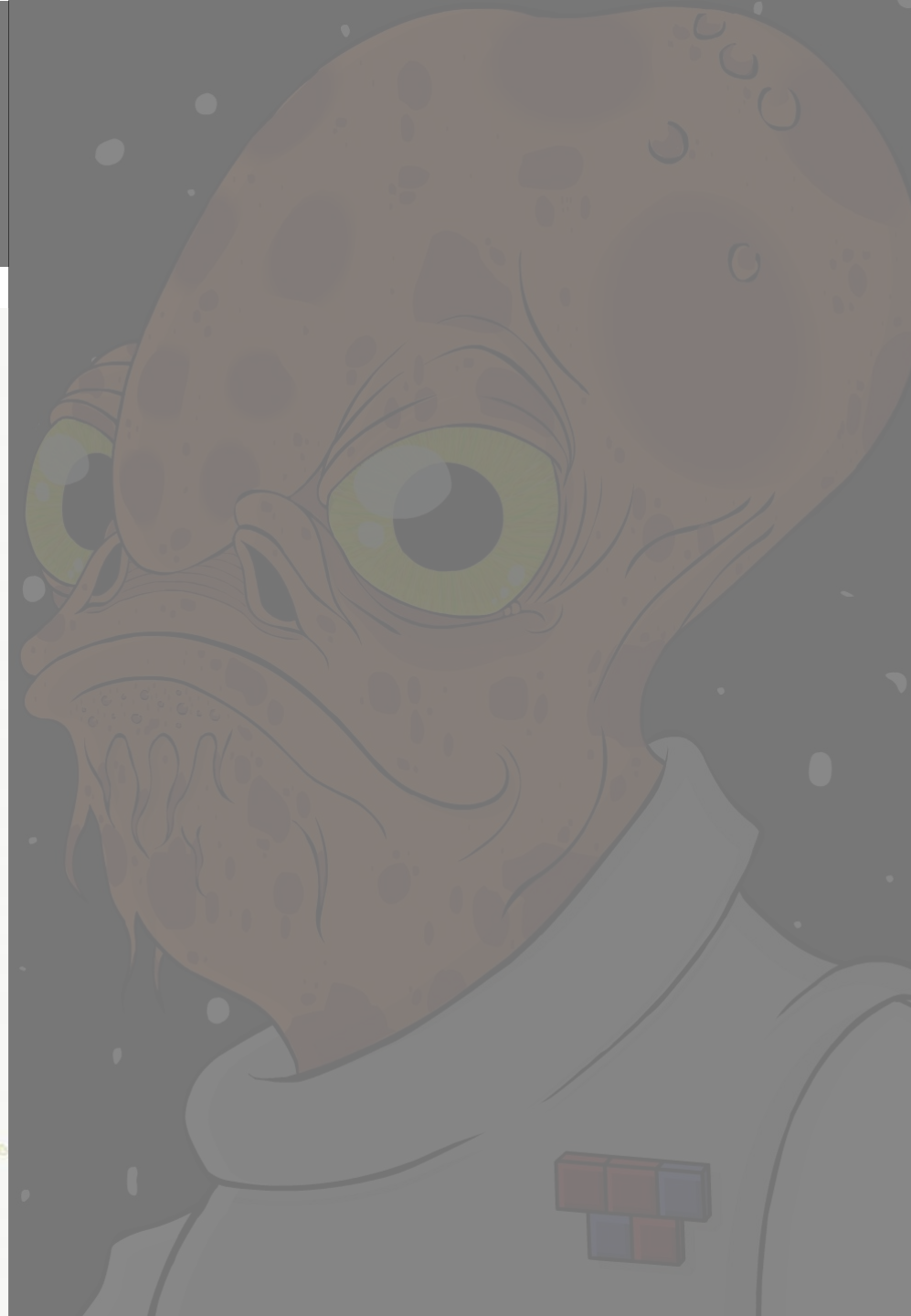
- Not a formal approach
- Response to a contextual problem
- Reinventing the wheel (for educational purpose only!)
- Some success, some mistakes, a bit of troll but a lot of fun

I'm a Security Kid, so I ask
kid's questions and provide
kid's responses



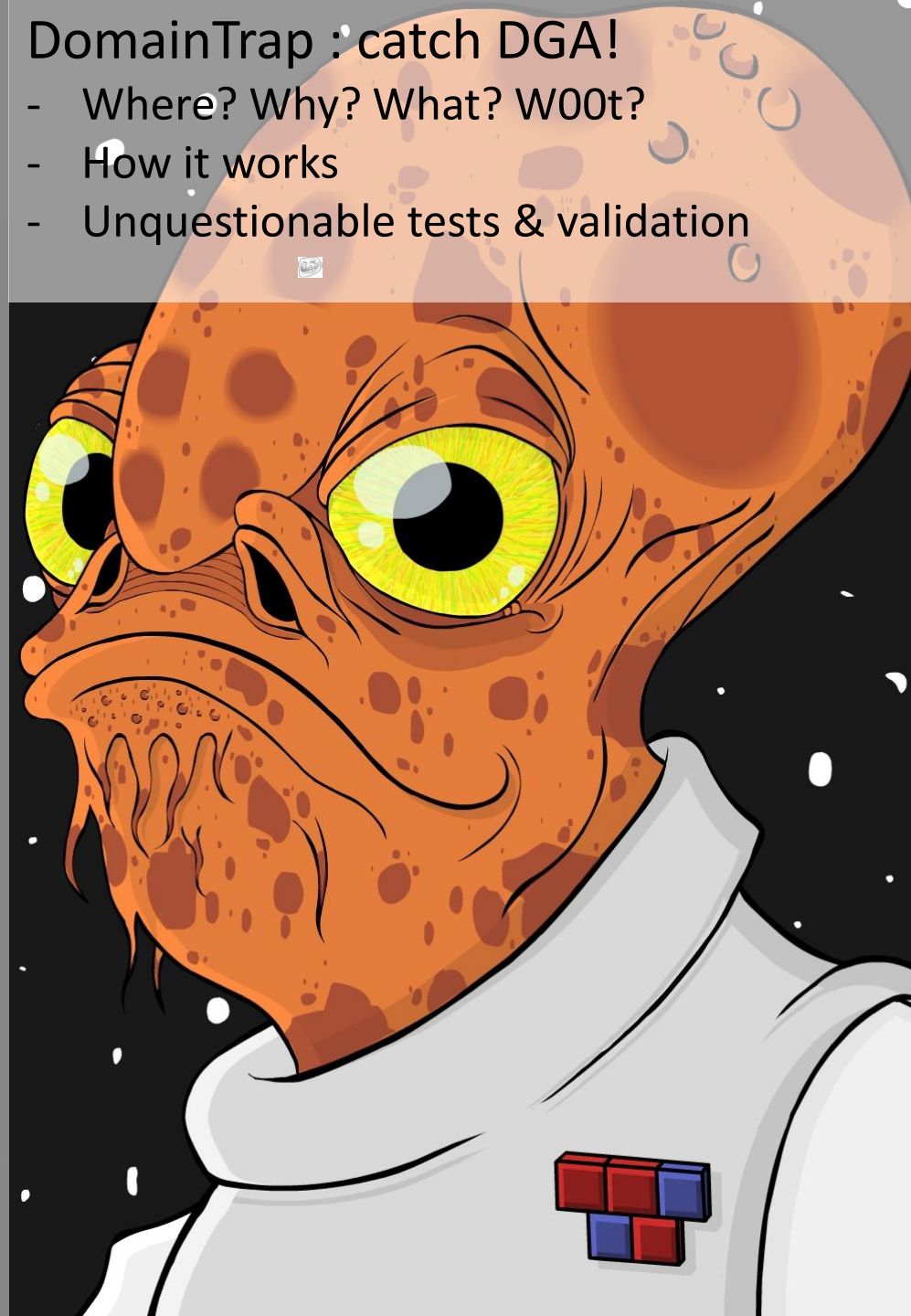
MalwareTrap : catch zombies!

- Where? Why? What? W00t?
- First PoC & Need of improvement
- MalwareTrapNG



DomainTrap : catch DGA!

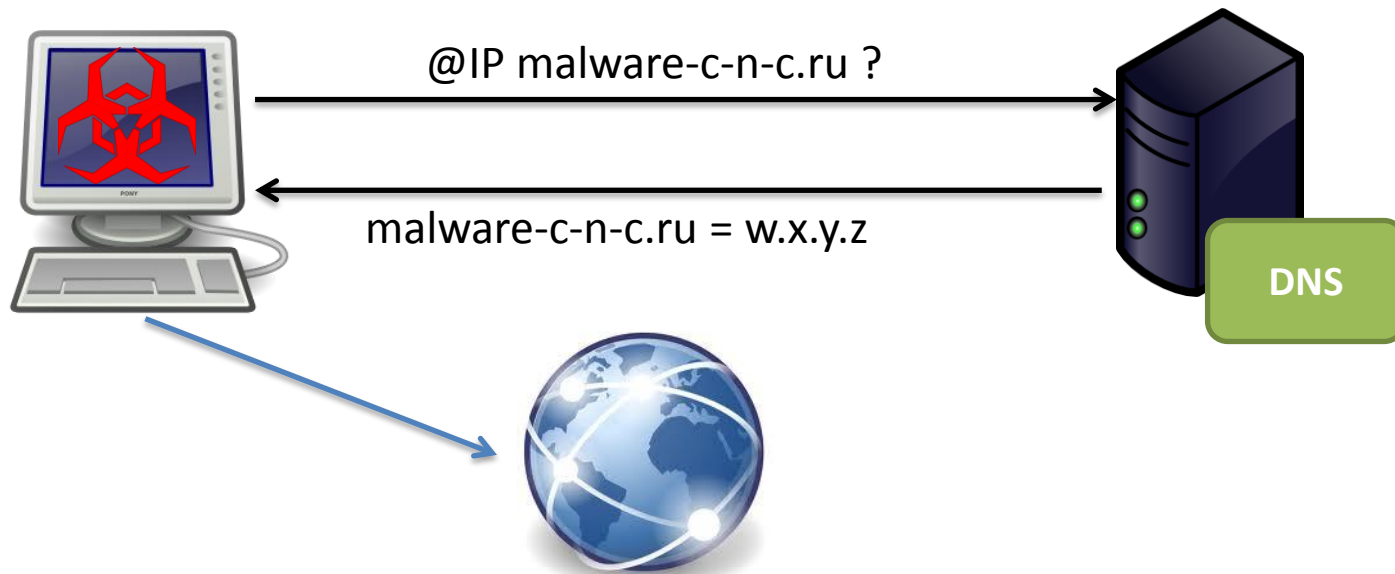
- Where? Why? What? W00t?
- How it works
- Unquestionable tests & validation



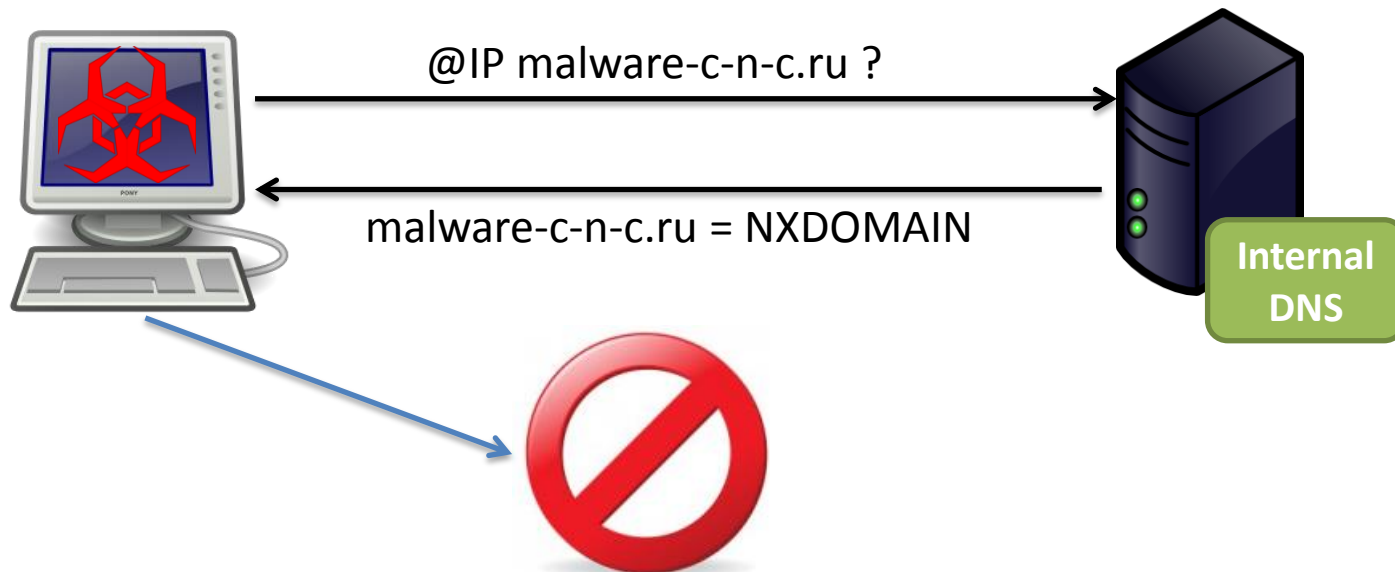
MalwareTrap

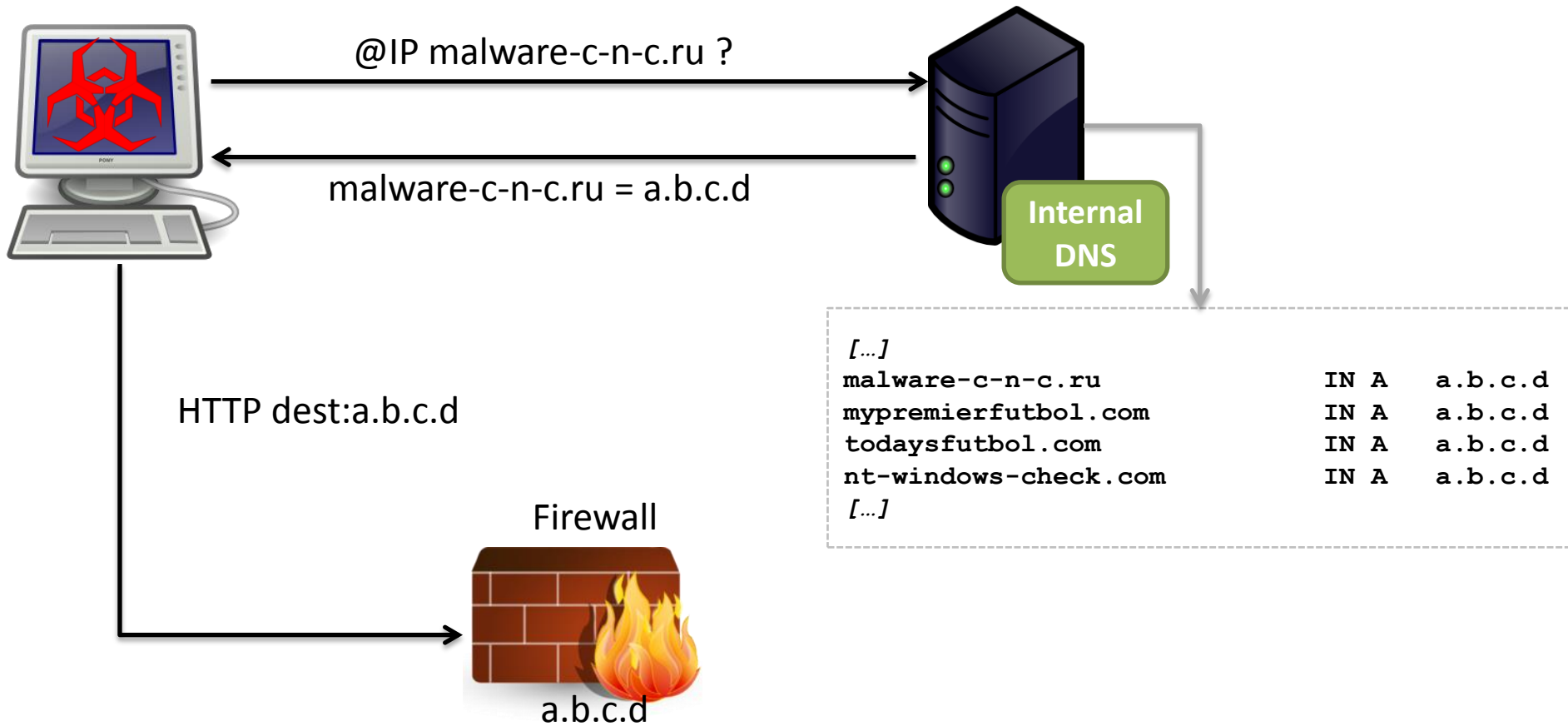
- Where : major French company (called *Canari* here)
nothing to do with security, IT or computer stuff
in small CSIRTeam with everything to do with security, IT or computer stuff
a very big network (~17 000 routers, 160 000 desktop units)
- What : Late 2010, discovering of a Mariposa epidemic (3% of the machine)
AV support unable to contain the pollution (from detection to cleaning)
AV updates sample by sample...coming from Canari's CSIRT
Idea to internally sinkhole C&C to locate the source
2 days of development
Unprecedented decrease of infected computer in one month

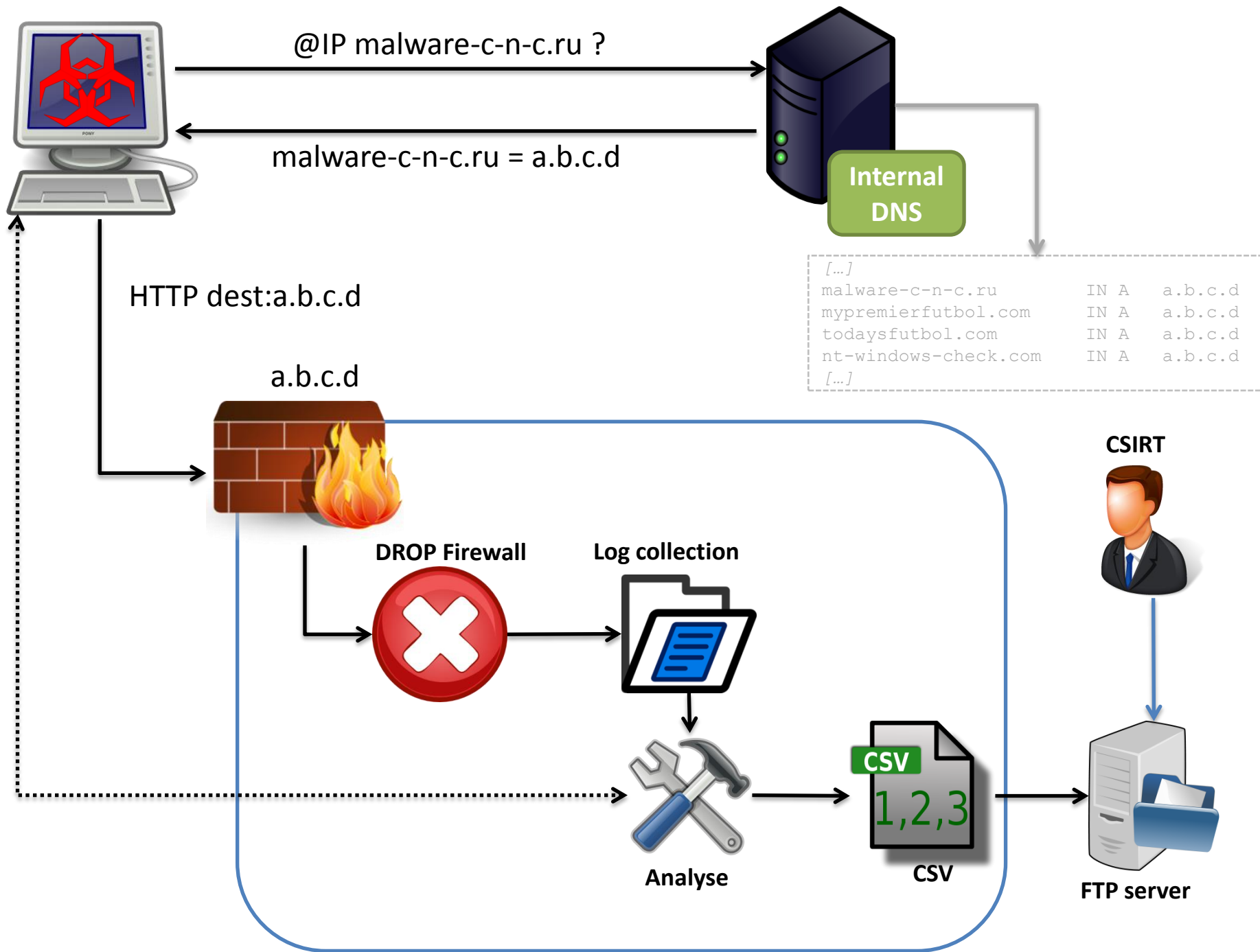




↑ Nominal
↓ Canari







Young Skywalker - rest now you must...



for someday save galaxy you will.

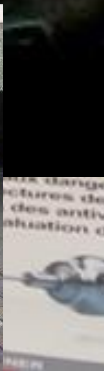
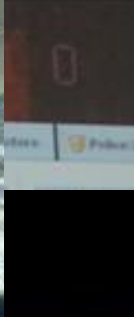
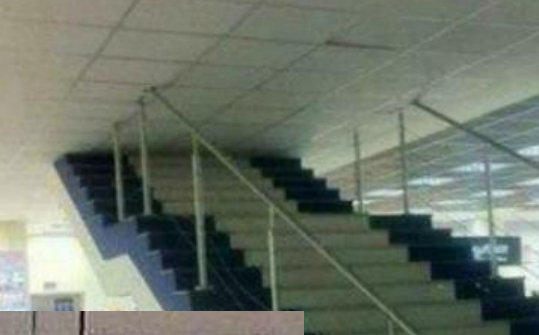
MalwareTrap's needs of improvment

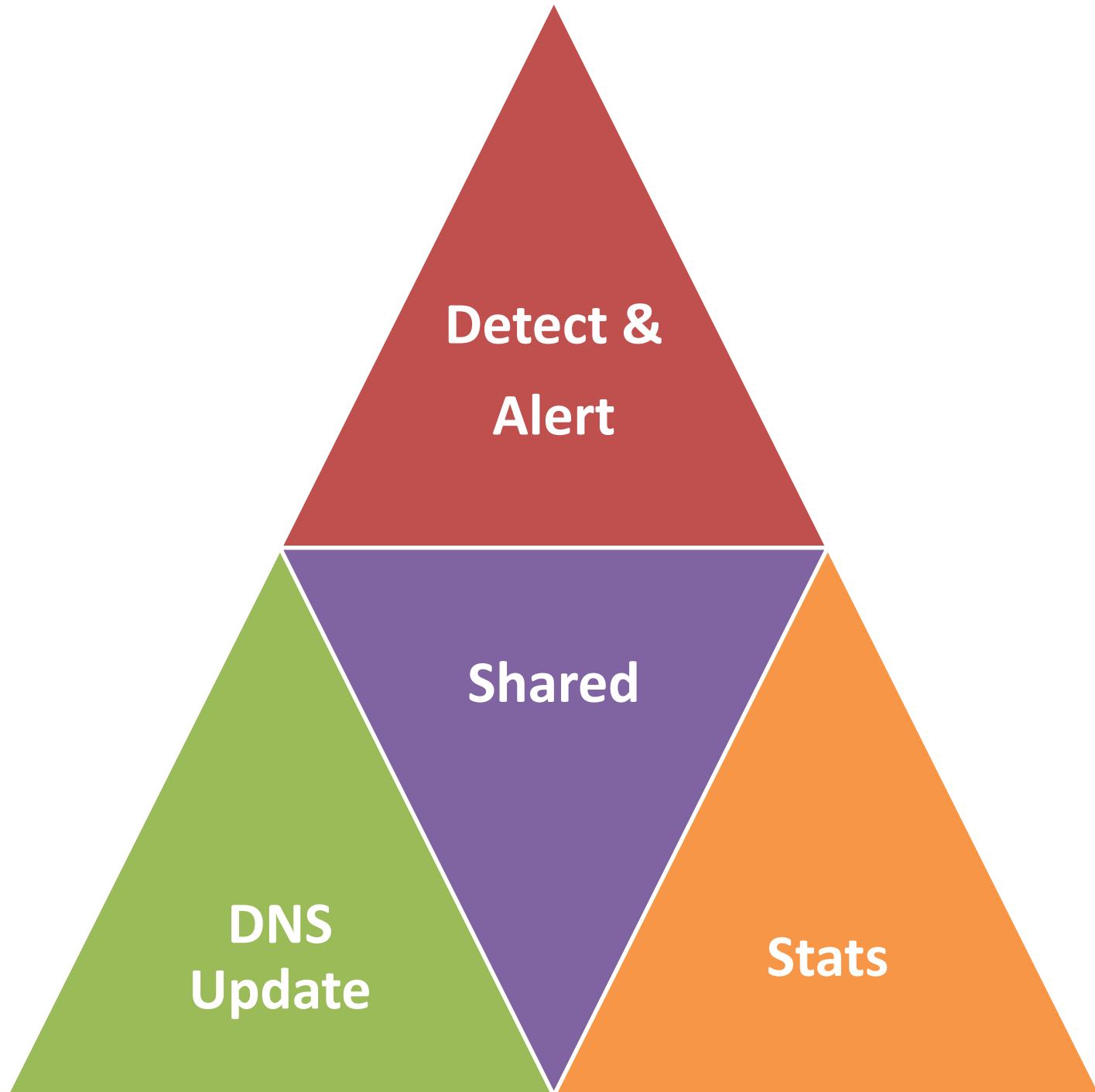
- Report on a monthly basis
- Polluting a production firewall's logs
- CSV format not very handy for quick incident assesment
- No check if FQDN were still malicious
- A single threat supervised

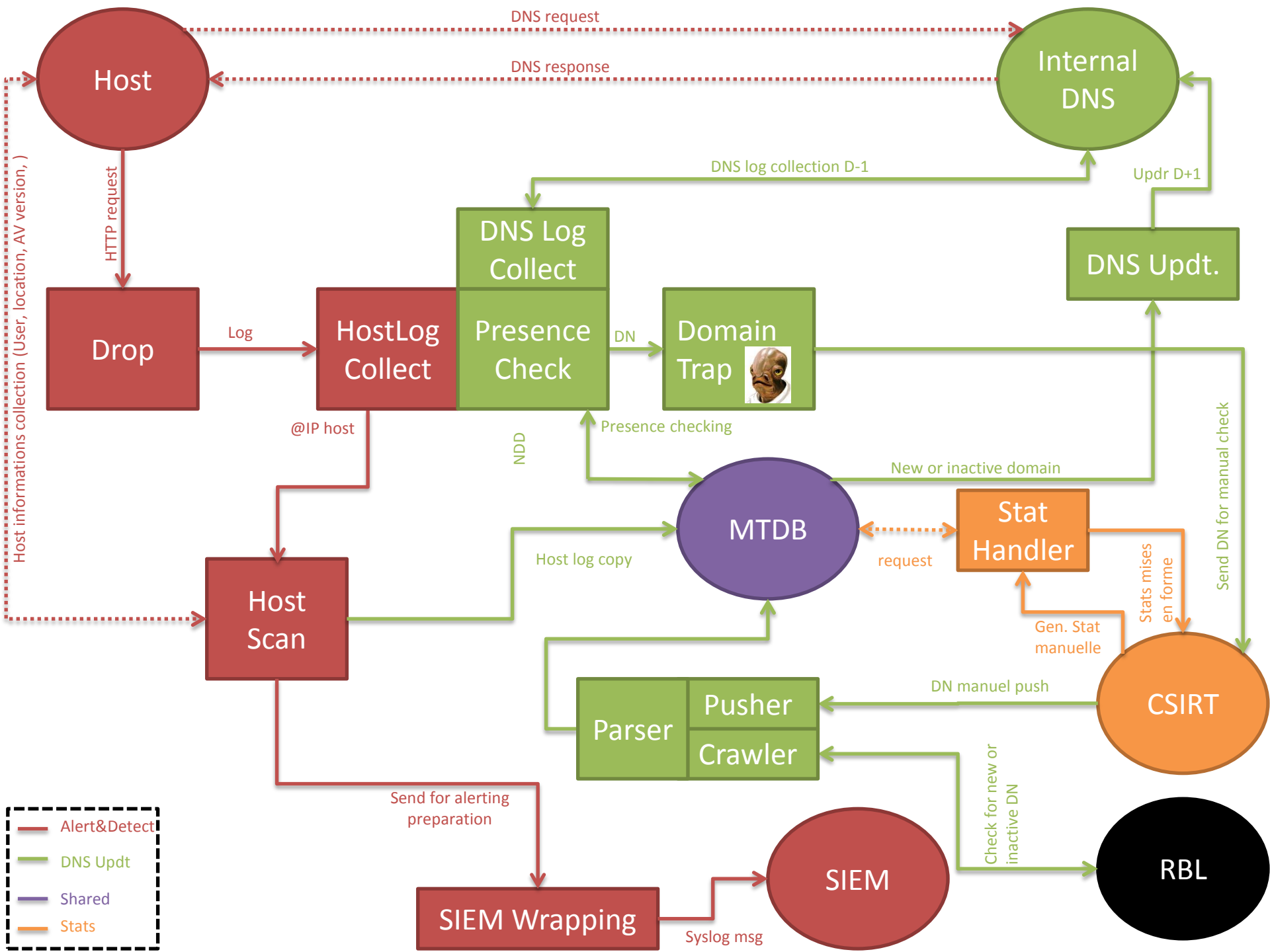


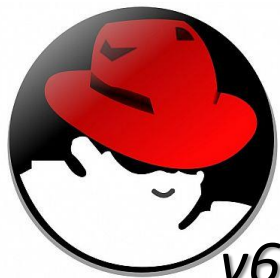
I WANT YOU TO

- 👉 Detect bot and notify the supervision team
- 👉 Identify currently used C&C for various threats
- 👉 Update DNS entries
- 👉 Wanna stats! (for my hierarchy you know ...)

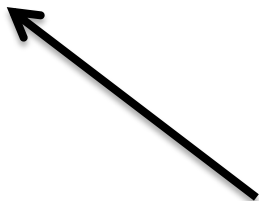




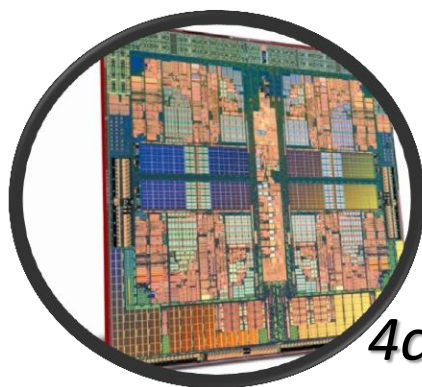
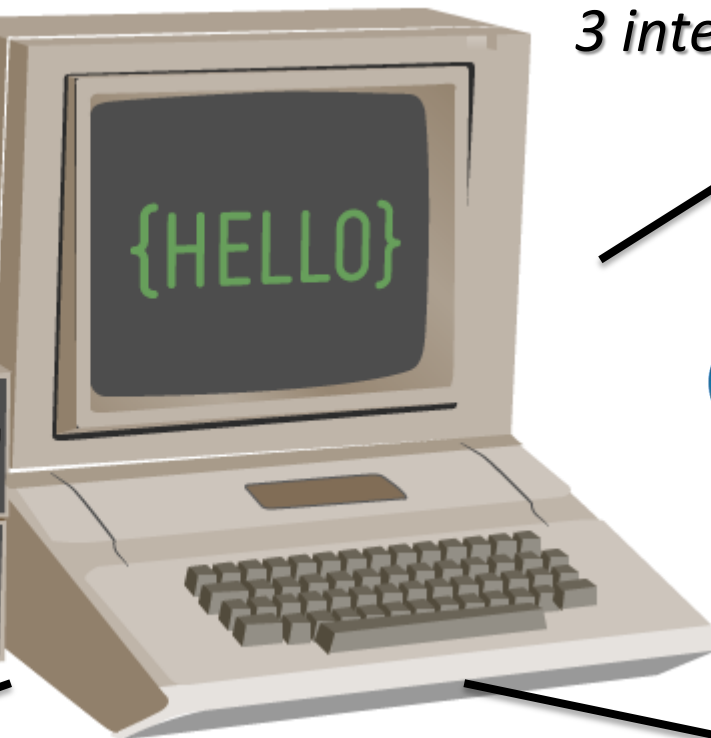
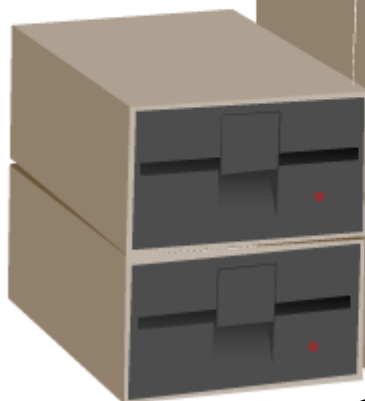
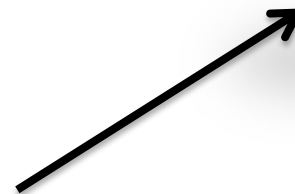




v6



3 interfaces



4cores
8 Go RAM



100 Go



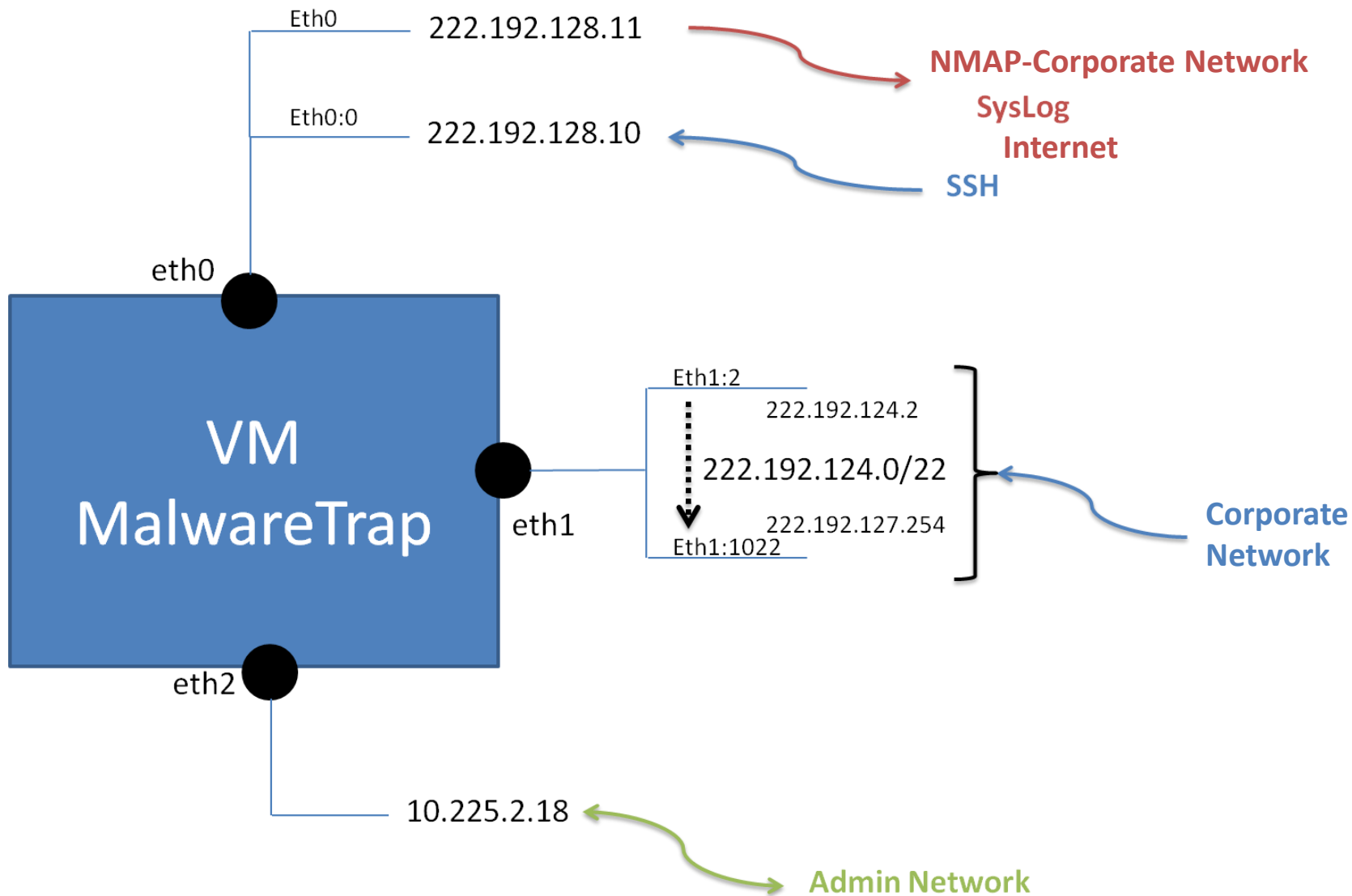
ROUND
1



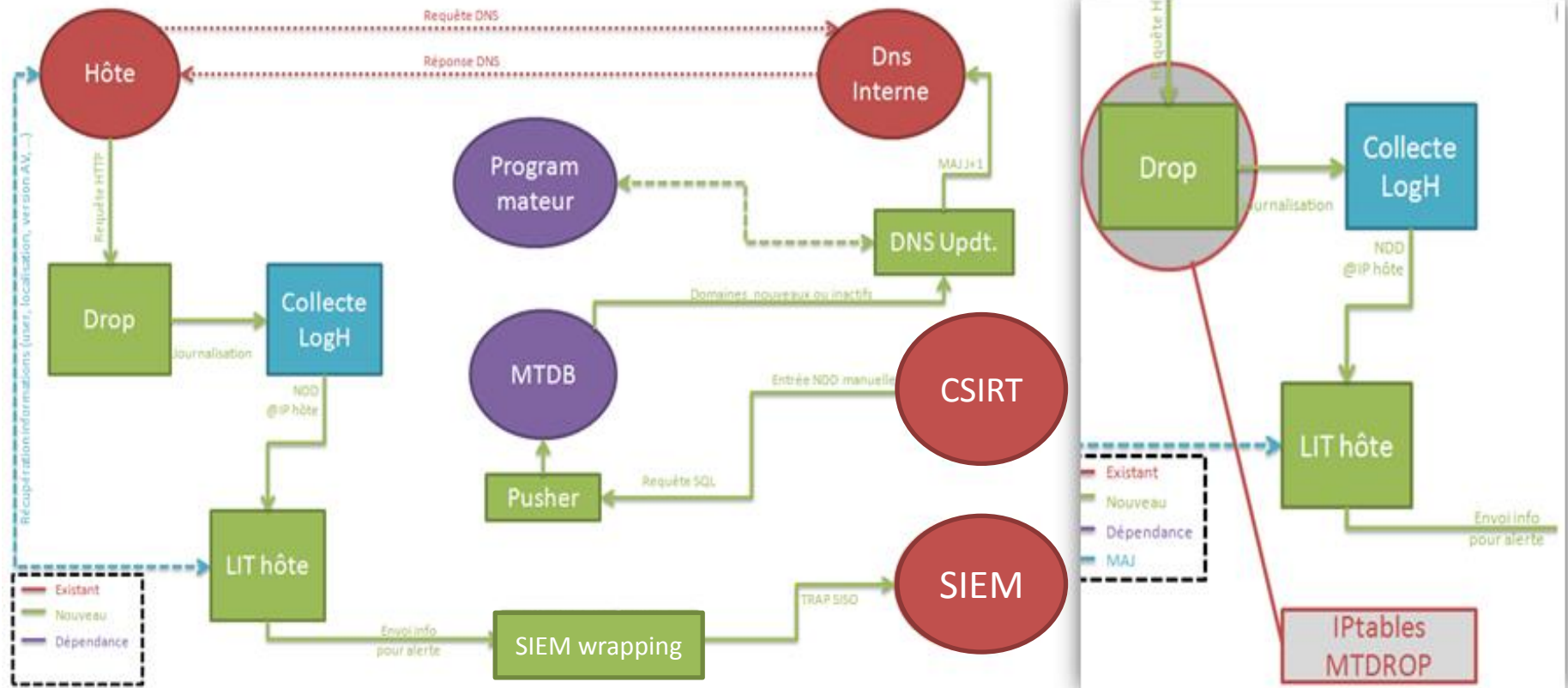
whitecollar  boxing.ie

whitecollar  boxing.ie





IPtables MTDROP



IPtable

```
# Creation d'une chaine de LOG&DROP (denomme MT-log-n-drop) d'une connexion
iptables -N MT-log-n-drop #cree une nouvelle chaine
iptables -A MT-log-n-drop -j LOG --log-prefix "MTDETECT " --log-level warning
iptables -A MT-log-n-drop -j ACCEPT
```

```
# ----- #
# ----- MT RULES : Regles metiers ----- #
# ----- #
# Regle LOG&DROP appliquee aux connexions entrantes
# Dans notre cas, les connexions ECRIN arrivent sur l'eth1
iptables -A INPUT -i eth1 -d 222.192.124.0/22 -j MT-log-n-drop
```



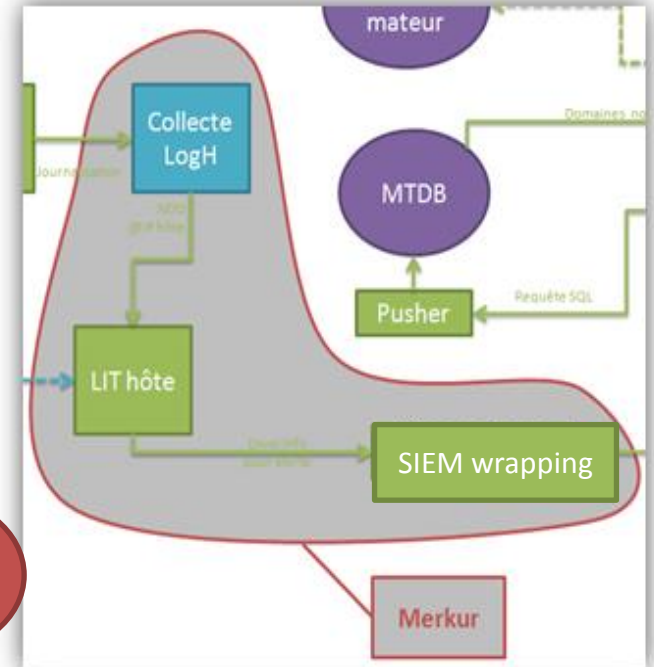
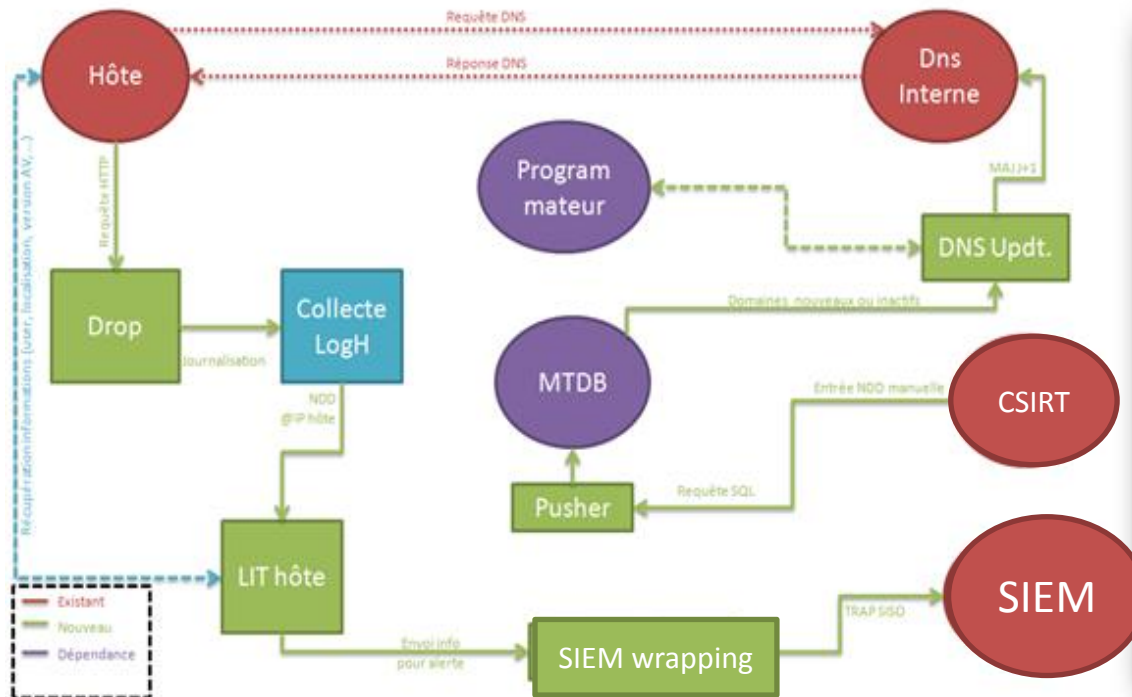
Log MTDROP

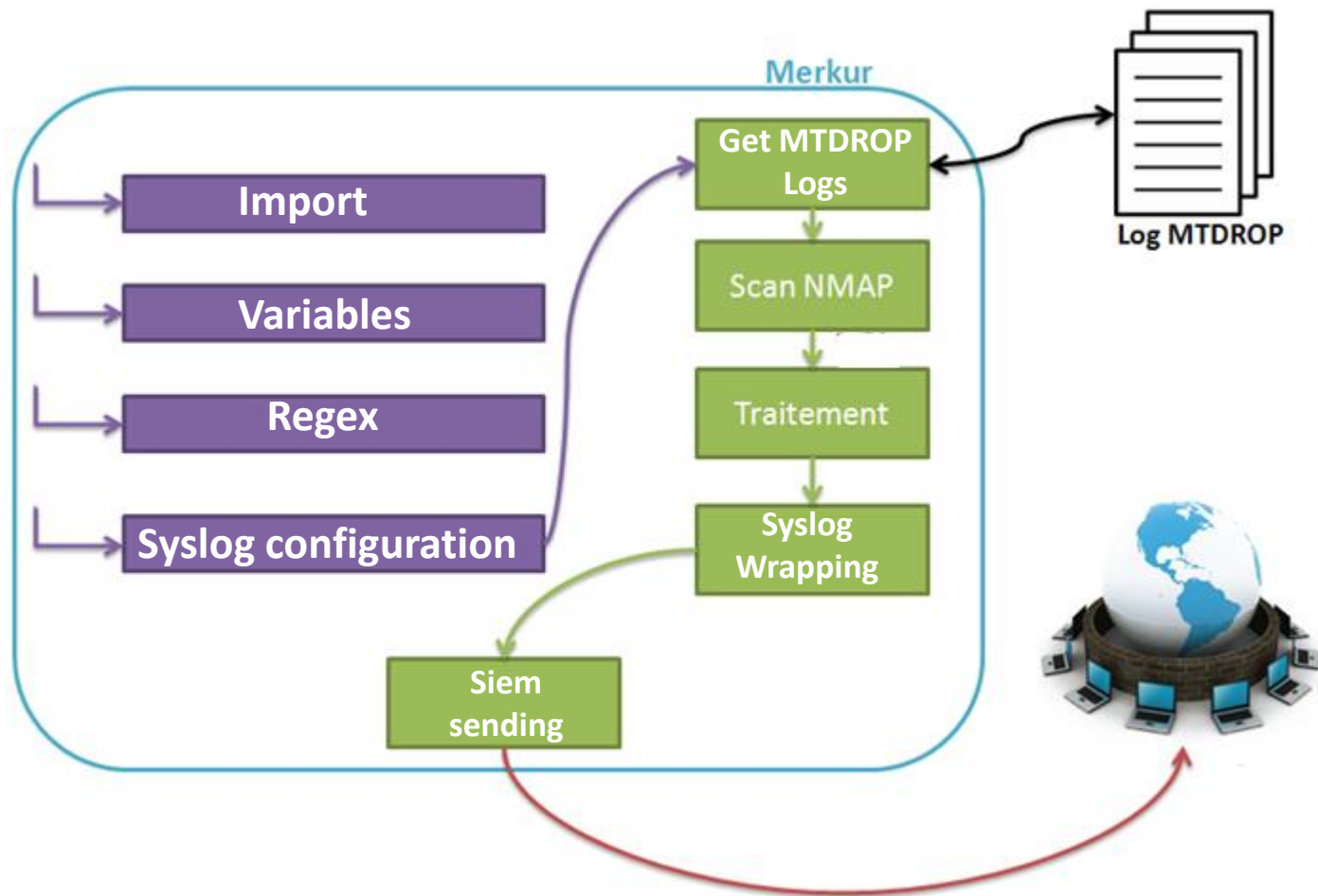
```
[root@Stat]# cat /etc/rsyslog.conf
#### RULES ####
# --- START MalwareTrap LOGnDROP forwarding rule --- #
:msg, startswith, "MTDETECT :" /MalwareTrap/LOG/drop.log
& ~
# --- STOP MalwareTrap LOGnDROP forwarding rule --- #

# --- START MalwareTrap GENERICDROPIN forwarding rule --- #
:msg, startswith, "MTGENERICIN :" /MalwareTrap/LOG/generic-in.log
& ~
# --- STOP MalwareTrap GENERICDROPIN forwarding rule --- #

# --- START MalwareTrap GENERICDROPOUT forwarding rule --- #
:msg, startswith, "MTGENERICOUT :" /MalwareTrap/LOG/generic-out.log
& ~
# --- STOP MalwareTrap GENERICDROPOUT forwarding rule --- #
```

Merkur





Scheduler

```
[root@xxxxxxxxxx Stat]# crontab -l  
*/5 * * * * /MalwareTrap/script/DetectnAlert/merkur-2-2.py &> /tmp/mttest.log
```

Index Search

Sources:

Last 3 Hours

Options

View



MALWARETRAPLOG

Run

MALWARETRAPLOG

Search Results

Search History

Search Filters

Clipboard



Clip selected message(s)

06/20/2013 09:58:00 - 12:58:00



Page

1

of 1

Time

Device IP

Device Source

Facility

Severity

06/20/13 12:33:58

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:12 iph

06/20/13 12:33:59

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:12 iph

06/20/13 12:34:07

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:20 iph

06/20/13 12:34:15

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:29 iph

06/20/13 12:34:16

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:29 iph
EPOStationName=444440APY36 EPOversion=4.6.0.2935 EPOS

06/20/13 12:34:16

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:30 iph
S

06/20/13 12:34:25

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:17:38 iph

06/20/13 12:37:42

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:20:55 iph
L vngnaw

06/20/13 12:37:42

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:20:56 iph

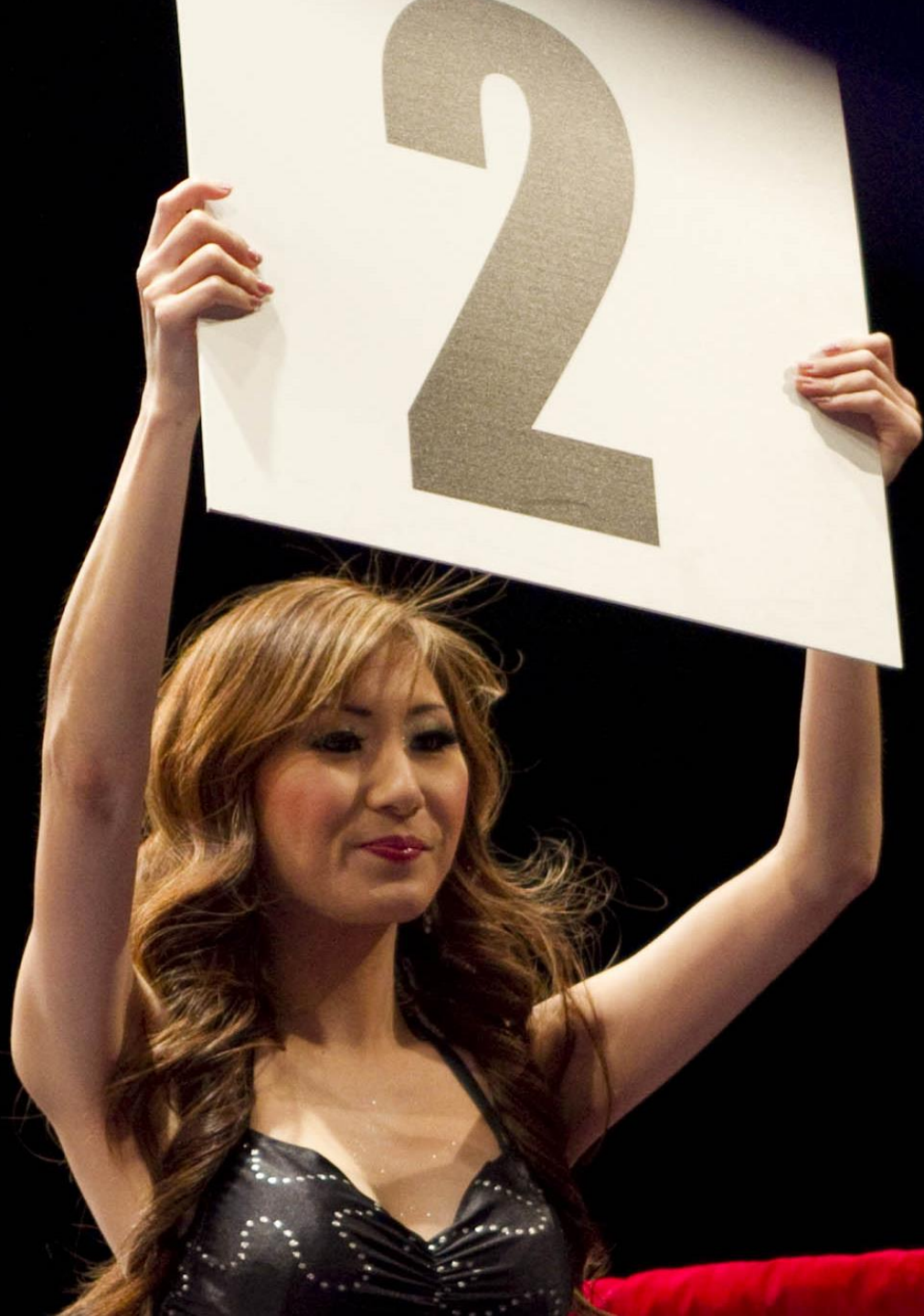
06/20/13 12:37:50

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:21:04 iph

06/20/13 12:37:58

Message: <12>MALWARETRAPLOG 20/Jun/2013:12:21:12 iph

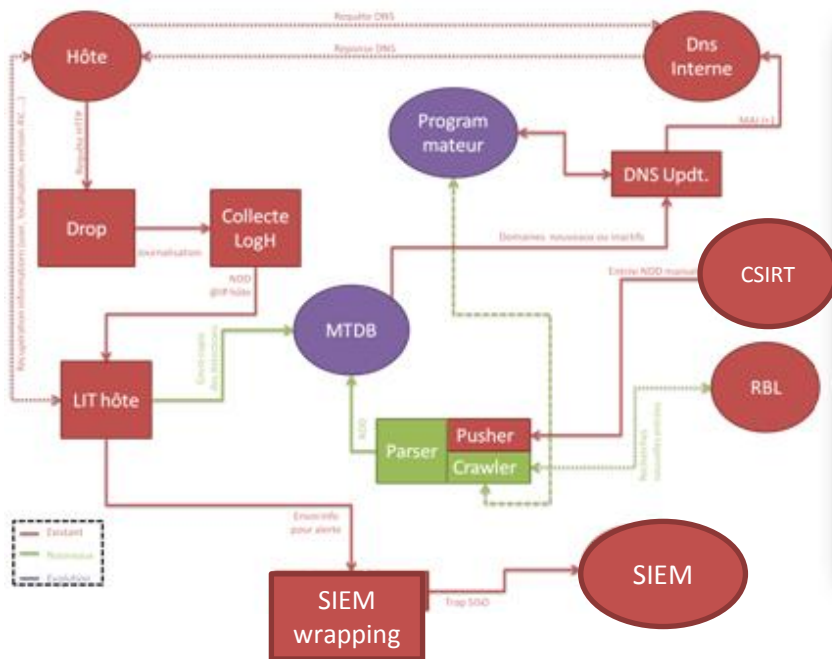
06/20/13 12:37:59

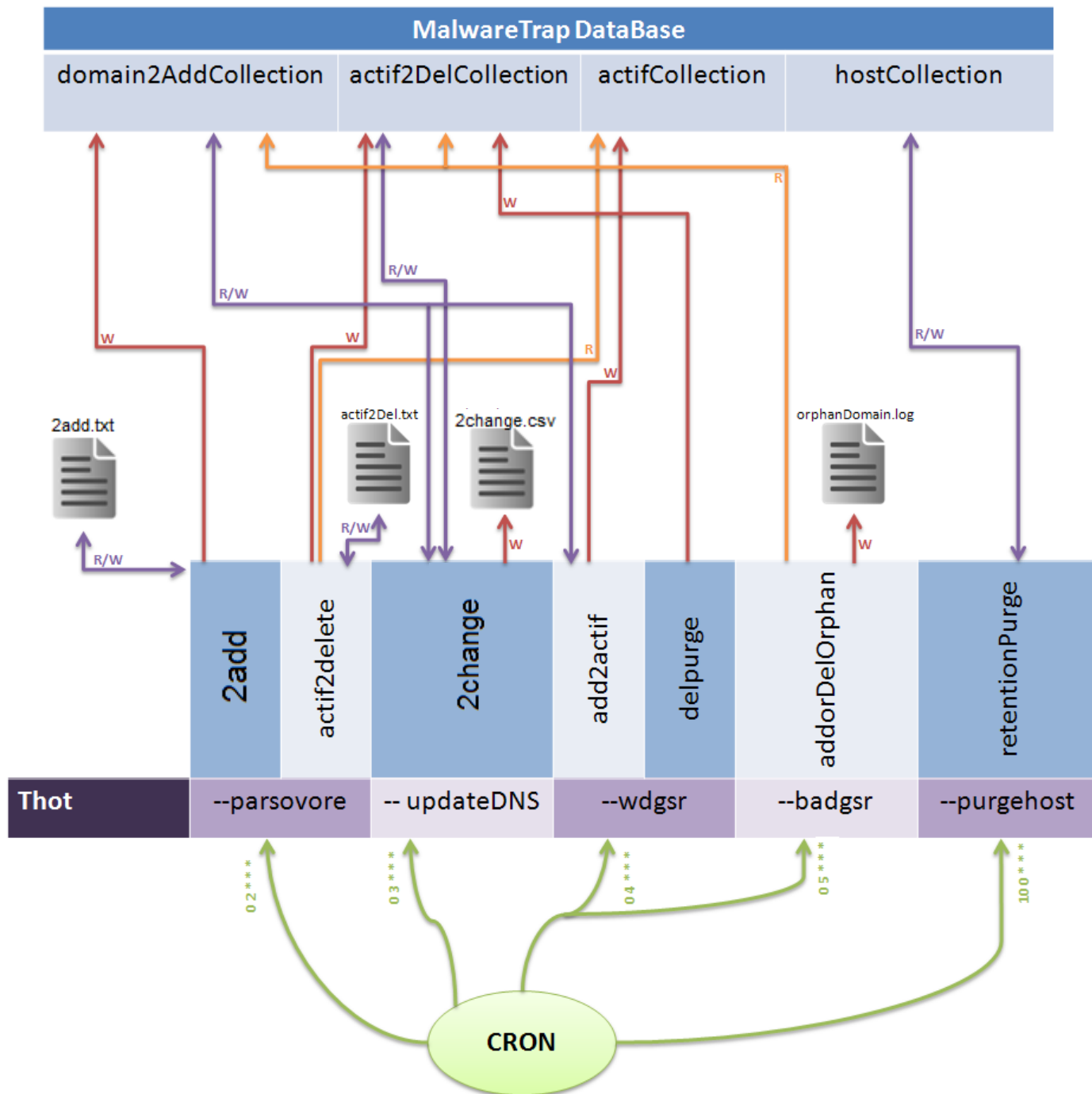


The diagram illustrates the NDD architecture. It features several components and their interactions:

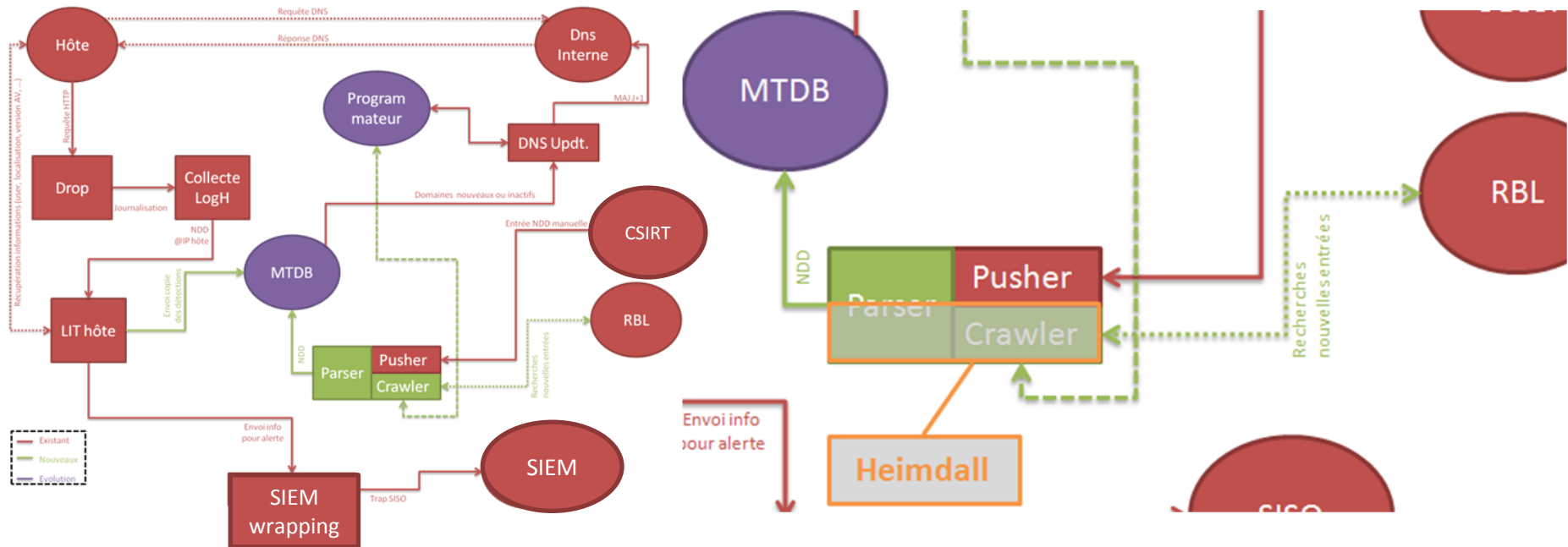
- Requête H**: An external request that enters the system.
- Drop**: A component that receives the request and logs it.
- Collecte LogH**: A component that collects logs from the Drop component.
- LIT hôte**: A component that receives data from the Drop component and sends it to the MTDB.
- MTDB**: A central database component that receives data from the LIT hôte and sends it to the Thot component.
- Thot**: A component that receives data from the MTDB and sends it to the Moteur.
- Moteur**: A component that receives data from the Thot component and sends it to the DNS Updt. component.
- DNS Updt.**: A component that updates the DNS based on the data received from the Moteur.
- Paraser**: A component that receives data from the MTDB and sends it to the Pusher component.
- Pusher**: A component that receives data from the Paraser component and sends it to the Crawler component.
- Crawler**: A component that receives data from the Pusher component and sends it to the Recherche des nouvelles entrées component.
- Recherche des nouvelles entrées**: A component that searches for new entries and sends the results to the Recherche des nouvelles entrées component.
- Recherche des nouvelles entrées**: A component that receives data from the Recherche des nouvelles entrées component and sends it to the Recherche des nouvelles entrées component.

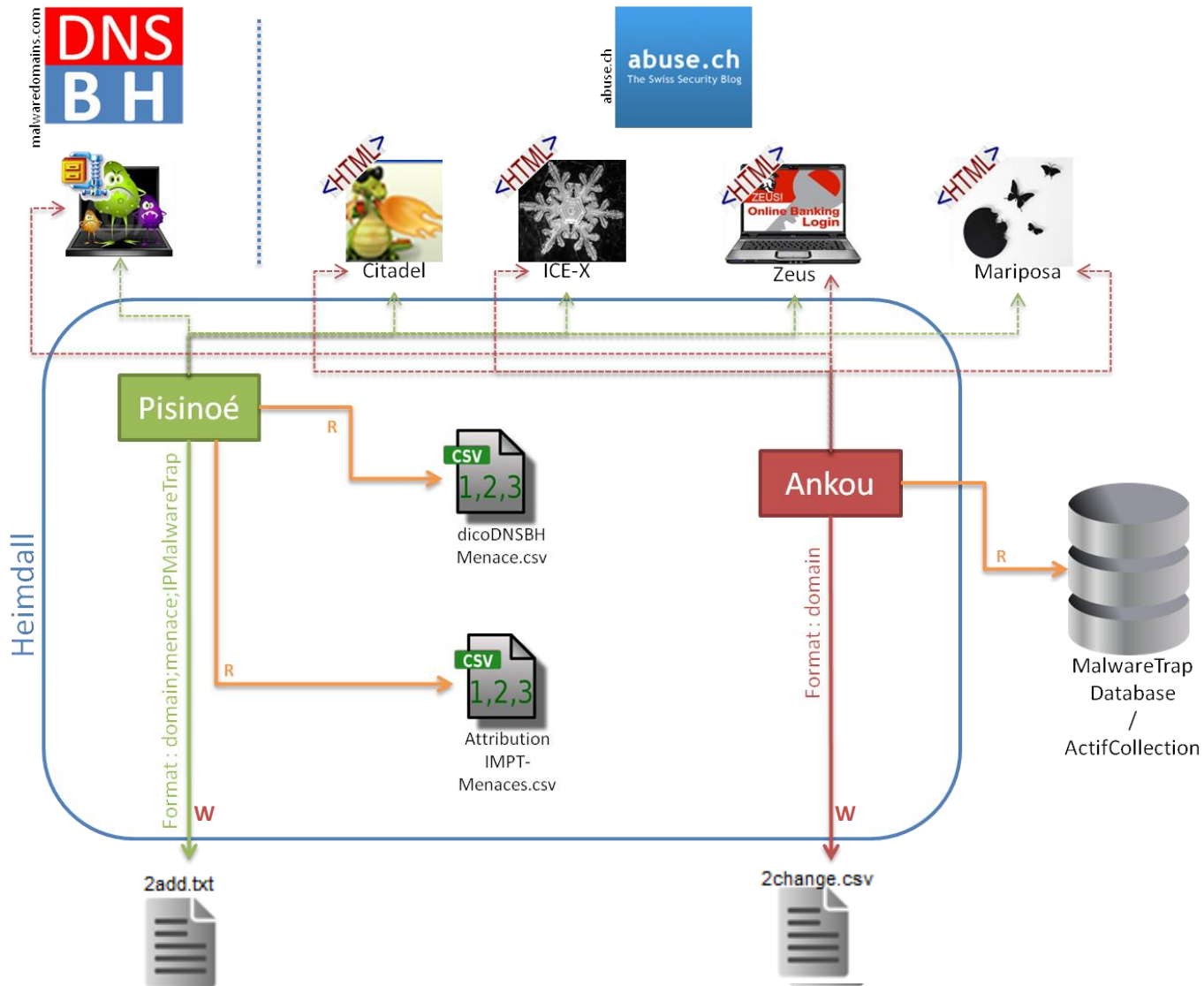
The diagram also shows a feedback loop from the Recherche des nouvelles entrées component back to the Paraser component.





Heimdall

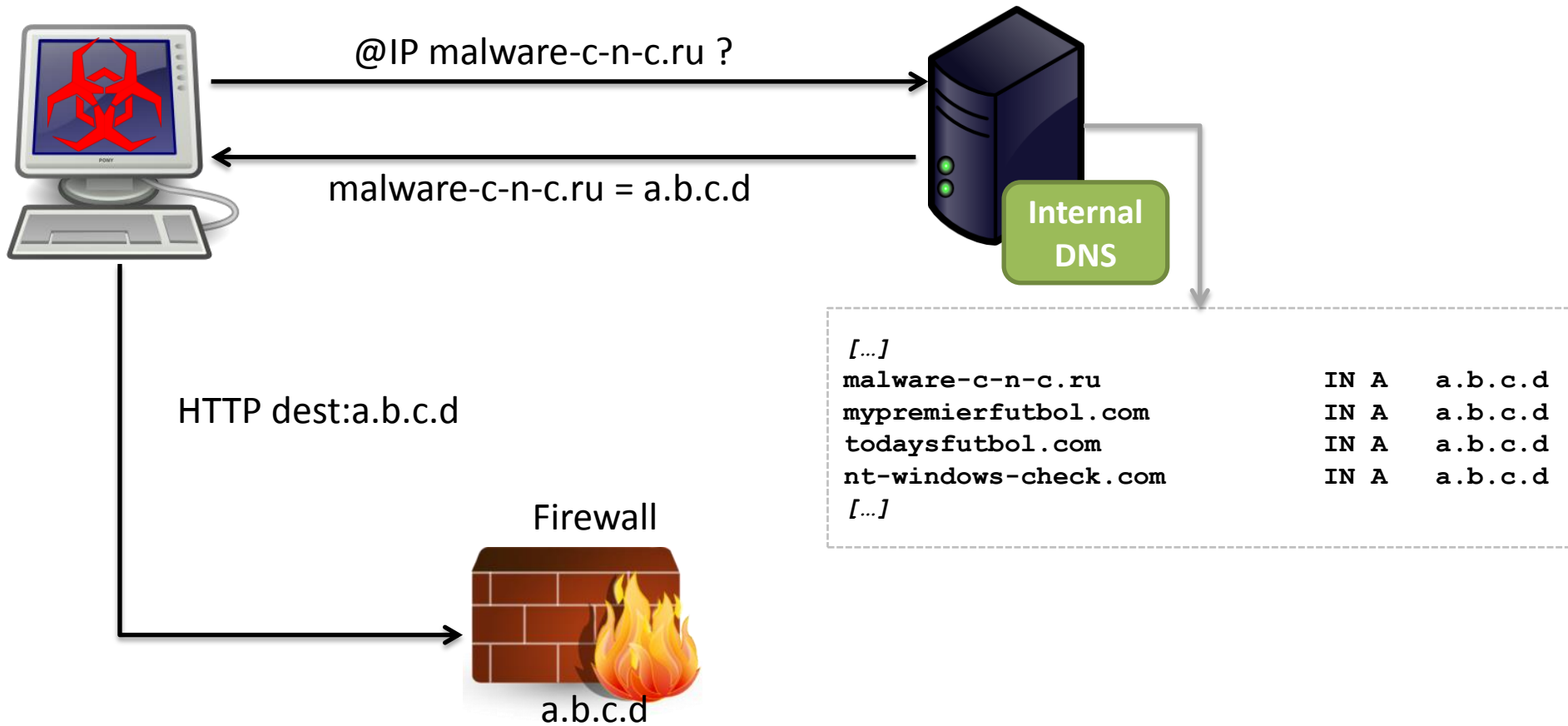


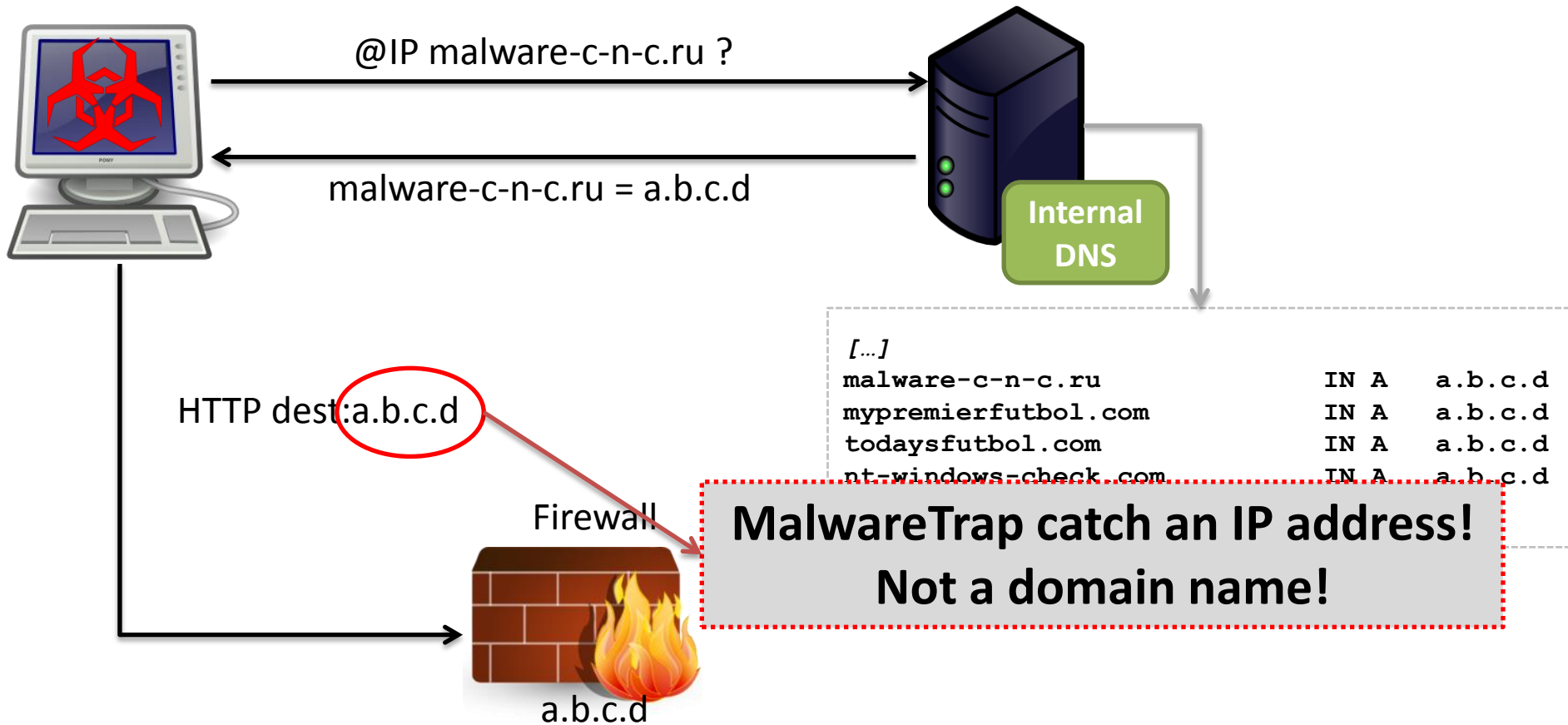




ThreatIP-
Attribution.csv

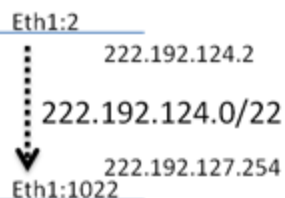
?





VM MalwareTrap

eth1



ThreatIP-
Attribution.csv

*.124.0/24 :

Cybercrime Threat

*.125.0/24 :

Advanced Persistent
Threat (APT)

*.12(6|7).0/24 :

Free

222.192.124.1	x	222.192.125.1	VARIOUSAPT	222.192.126.1	x	222.192.127.1	x
222.192.124.2	DEFAULT	222.192.125.2	MANDIANT-APT1	222.192.126.2	x	222.192.127.2	x
222.192.124.3	MARIPOSA	222.192.125.3	COMMENTCREW	222.192.126.3	x	222.192.127.3	x
222.192.124.4	ZEUS	222.192.125.4	FIN FISHER	222.192.126.4	x	222.192.127.4	x
222.192.124.5	ICE-X	222.192.125.5	GAUSS	222.192.126.5	x	222.192.127.5	x
222.192.124.6	CITADEL	222.192.125.6	KEYBOY	222.192.126.6	x	222.192.127.6	x
222.192.124.7	ZEUSGAMEOVER	222.192.125.7	LADYBOYLE	222.192.126.7	x	222.192.127.7	x
222.192.124.8	FARFLI	222.192.125.8	NETTRAVELER	222.192.126.8	x	222.192.127.8	x
222.192.124.9	BETABOT	222.192.125.9	RBN	222.192.126.9	x	222.192.127.9	x
222.192.124.10	CDORKED	222.192.125.10	x	222.192.126.10	x	222.192.127.10	x
222.192.124.11	COOLWEBSEARCH	222.192.125.11	x	222.192.126.11	x	222.192.127.11	x
222.192.124.12	CRIDEX	222.192.125.12	x	222.192.126.12	x	222.192.127.12	x
222.192.124.13	EXPIROZ	222.192.125.13	x	222.192.126.13	x	222.192.127.13	x
222.192.124.14	GATAK	222.192.125.14	x	222.192.126.14	x	222.192.127.14	x
222.192.124.15	GEINIMI	222.192.125.15	x	222.192.126.15	x	222.192.127.15	x
222.192.124.16	GHOSTRAT	222.192.125.16	x	222.192.126.16	x	222.192.127.16	x
222.192.124.17	SHIZINFOSTEALER	222.192.125.17	x	222.192.126.17	x	222.192.127.17	x
222.192.124.18	KELIHOS	222.192.125.18	x	222.192.126.18	x	222.192.127.18	x
222.192.124.19	ZEUSKNEBER	222.192.125.19	x	222.192.126.19	x	222.192.127.19	x
222.192.124.20	KOOBFACE	222.192.125.20	x	222.192.126.20	x	222.192.127.20	x
222.192.124.21	KULUQZ	222.192.125.21	x	222.192.126.21	x	222.192.127.21	x
222.192.124.22	MEDFOS	222.192.125.22	x	222.192.126.22	x	222.192.127.22	x
222.192.124.23	NAIKON	222.192.125.23	x	222.192.126.23	x	222.192.127.23	x
222.192.124.24	PONY	222.192.125.24	x	222.192.126.24	x	222.192.127.24	x
222.192.124.25	PONMOCUP	222.192.125.25	x	222.192.126.25	x	222.192.127.25	x
222.192.124.26	PUSHDO	222.192.125.26	x	222.192.126.26	x	222.192.127.26	x
222.192.124.27	RANBYUS	222.192.125.27	x	222.192.126.27	x	222.192.127.27	x
222.192.124.28	PWS	222.192.125.28	x	222.192.126.28	x	222.192.127.28	x
222.192.124.29	RAWIN	222.192.125.29	x	222.192.126.29	x	222.192.127.29	x
222.192.124.30	ZEROACCESS	222.192.125.30	x	222.192.126.30	x	222.192.127.30	x
222.192.124.31	RUBOBO	222.192.125.31	x	222.192.126.31	x	222.192.127.31	x
222.192.124.32	SINOWAL	222.192.125.32	x	222.192.126.32	x	222.192.127.32	x
222.192.124.33	SUTRADIS	222.192.125.33	x	222.192.126.33	x	222.192.127.33	x
222.192.124.34	TDSS	222.192.125.34	x	222.192.126.34	x	222.192.127.34	x
222.192.124.35	TRAVNET	222.192.125.35	x	222.192.126.35	x	222.192.127.35	x
222.192.124.36	KRYPT	222.192.125.36	x	222.192.126.36	x	222.192.127.36	x
222.192.124.37	ZEGOST	222.192.125.37	x	222.192.126.37	x	222.192.127.37	x
222.192.124.38	ZBOT	222.192.125.38	x	222.192.126.38	x	222.192.127.38	x
222.192.124.39	RUNFORESTRUN	222.192.125.39	x	222.192.126.39	x	222.192.127.39	x

OM



www.ONEFC.com

ONE



634 056



100



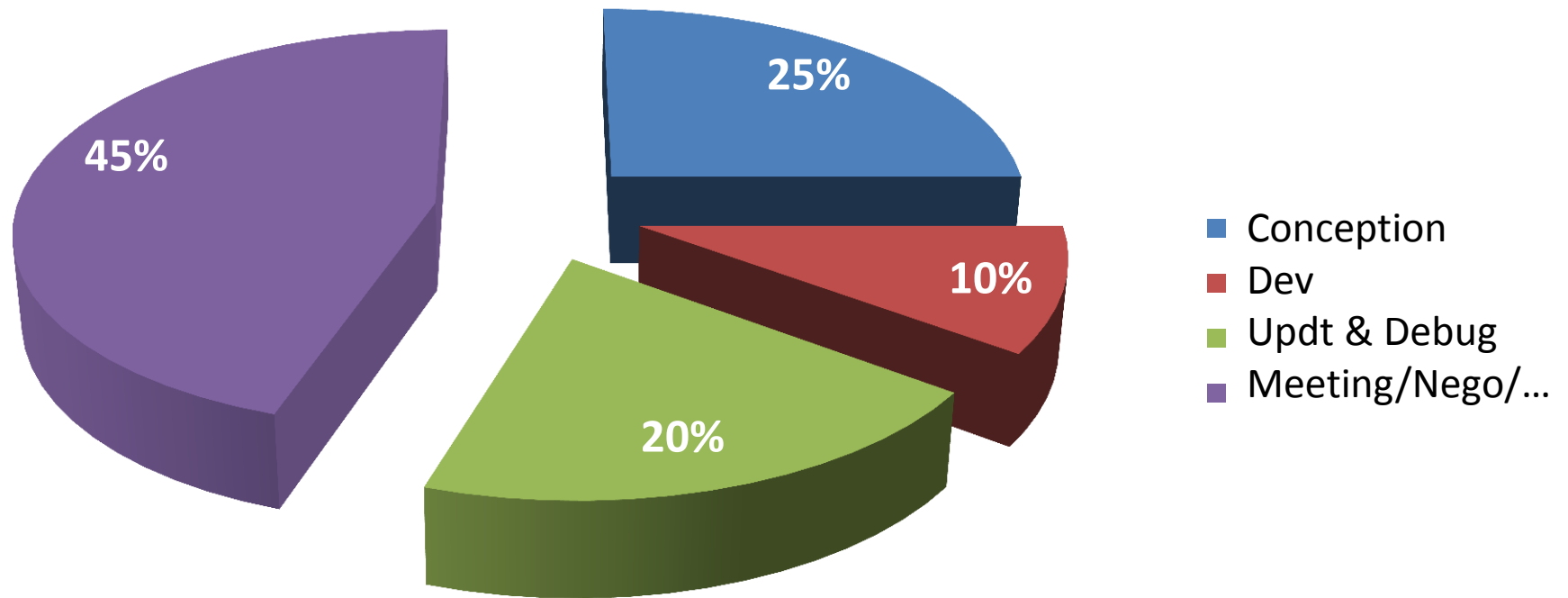
1

08/07/2013 -> 08/08/2013

~90%

Adresse IP	Events	Moy/j
0.106	174167	5 618
3.19	169542	5 469
3.45	39658	1 279
0.229	34502	1 113
7.134	34276	1 106
72	25435	820
0.52	23971	773
0.29	23819	768
0.28	20742	669
0.29	14775	477
TOTAL	560887	18 093

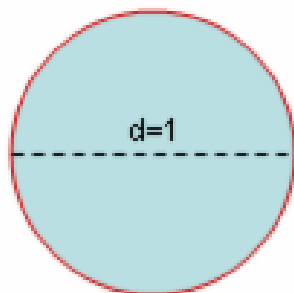
470h of work over 5 months



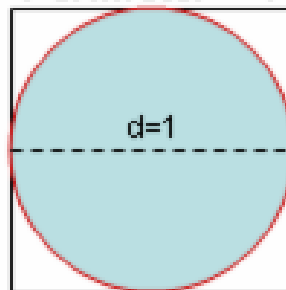
*It's time to talk a
bit about me*



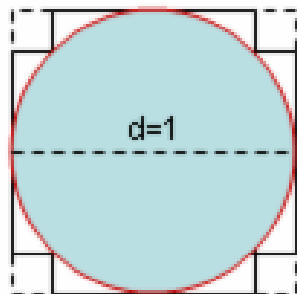
Draw a circle



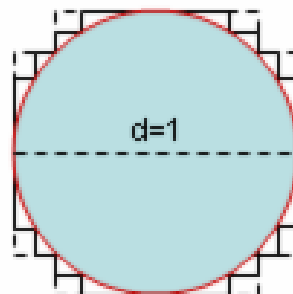
Draw a square around it
Perimeter = 4



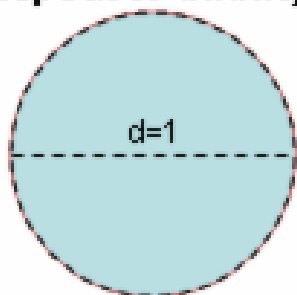
Remove corners.
Perimeter is still 4!



Remove more corners.
Perimeter is still 4!



Repeat to infinity

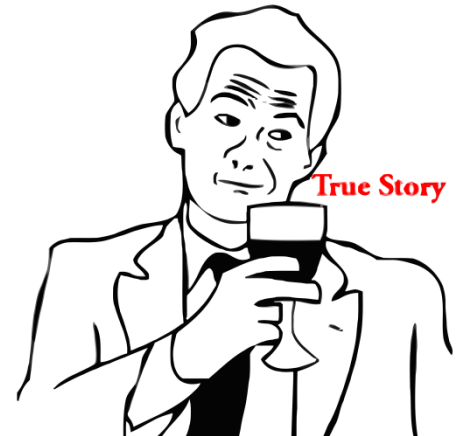


$$\pi = 4!$$



Problem Archimedes?

xyjvufatvwjrviw[.biz]



Cryptolocker

jyyfmnefedjogsh.biz
dxejhoplbgymgld.com
oiokidamwjythao.info
qayuttfyvdsufol.net
bgtyvxkyemflyjo.co.uk
ujtohypxdfvrtor.org
fposjduxloiiurh.net
srjeviklelcqdbl.biz
hhydutakkicjusf.ru
ngxdvwiwmlfxegs.info

RunForestRun

*.qxpnhnrvrkqewurq.waw.pl (Sat Jul 21 2012 00:00)
*.keefqnfsgqxrzlrw.waw.pl (Sat Jul 21 2012 01:00)
*.ekkugeunekaxqolz.waw.pl (Sat Jul 21 2012 07:00)
*.svndeqsqughepaye.waw.pl (Sat Jul 21 2012 13:00)
*.aksfkuuozvfqprms.waw.pl (Sat Jul 21 2012 19:00)
*.zpqkervzziqffvas.waw.pl (Sun Jul 22 2012 00:00)
*.uiuxumxroflzpfxr.waw.pl (Sun Jul 22 2012 01:00)

type	time dependent	deterministic	example
TID	no	yes	Kraken
TDD	yes	yes	Conficker
TDN	yes	no	Torpig

→ Not a word!



jyyfmnefedjogsh
dxejhopldgymgld
oiokidamwjythao
qayuttfyvdsufol
bgtyvxkyemflyjo
ujtohypxdfvrtor
fposjduxloiiurh
srjeviklelcqdbl
hhydutakkicjusf
ngxdvwiwmlfxegs

- Identical letter or number series
- Consonant series
- Too many '-'
- Too many numbers
- Too many '.'
- Too many rare letters



- Not closed to a french word
- Not closed to an english word
- Too much characters diversity
- This tld is know to be evil
- This domain is know to be evil

- Identical letter or number series
- Consonant series
- Too many '-'
- Too many numbers
- Too many ‘.
- Too many rare letters



Sashav1 :

- Group 9 infractions
- Constant or percentage trigger
- One afternoon for the PoC
- Detected 65% of DGA
- 10% of FP

- Not closed to a french word
- Not closed to an english word
- Too much characters diversity

Shannonv1 :

- 2 algorithms
 - Levenshtein
 - Shannon
- 15 000 random English words
- 15 000 random French words
- One day for the PoC
- TP : 50%
- FP : 20%



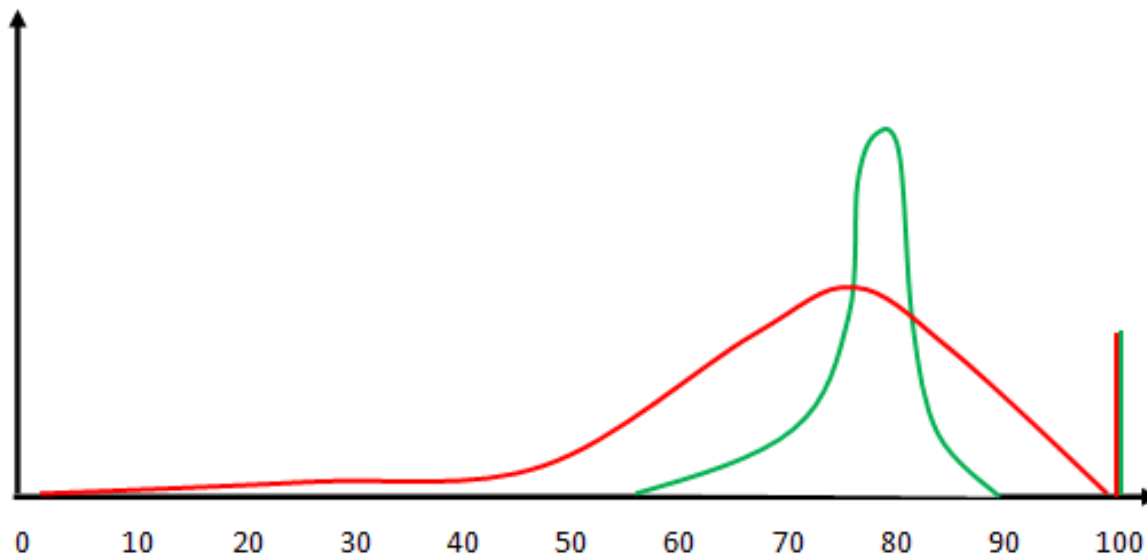
Shannon entropy

Quantity of information contain in a message :

- toto $\rightarrow \{o,t\} \rightarrow 2$ informations
- $\text{len}(\text{toto}) = \text{message size} = 4$
- entropy : $2/4 = 0,5$

Low entropy : redondancy

High entropy : equiprobable random generation



Levenshtein distance

Number of modifications to go from one string to another

toto -> tatao = ?

t / t = 0

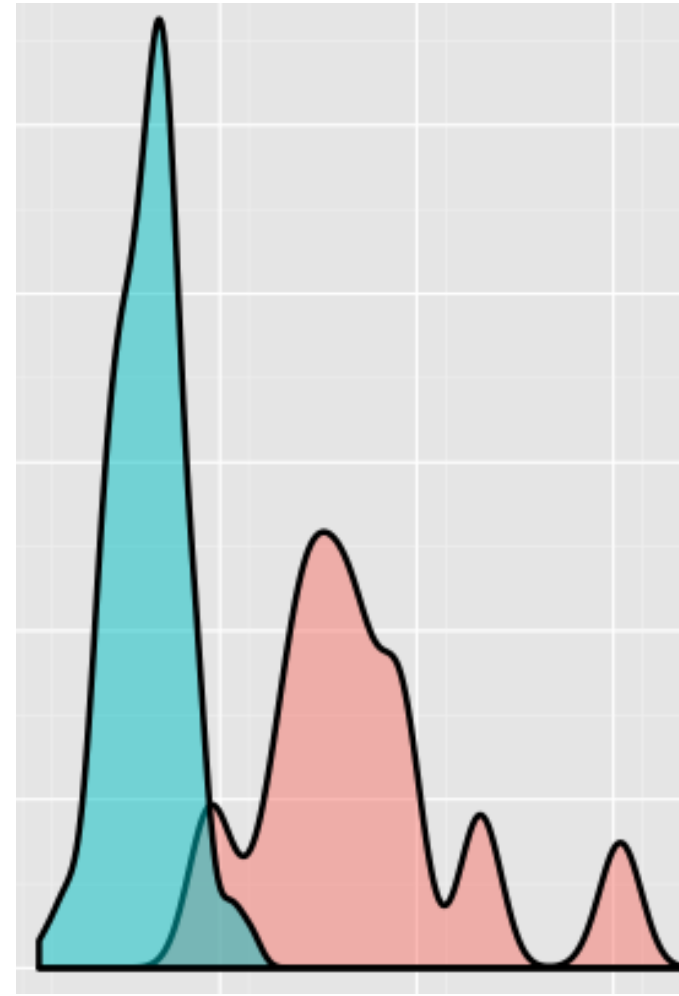
to / t[a->o] = 1

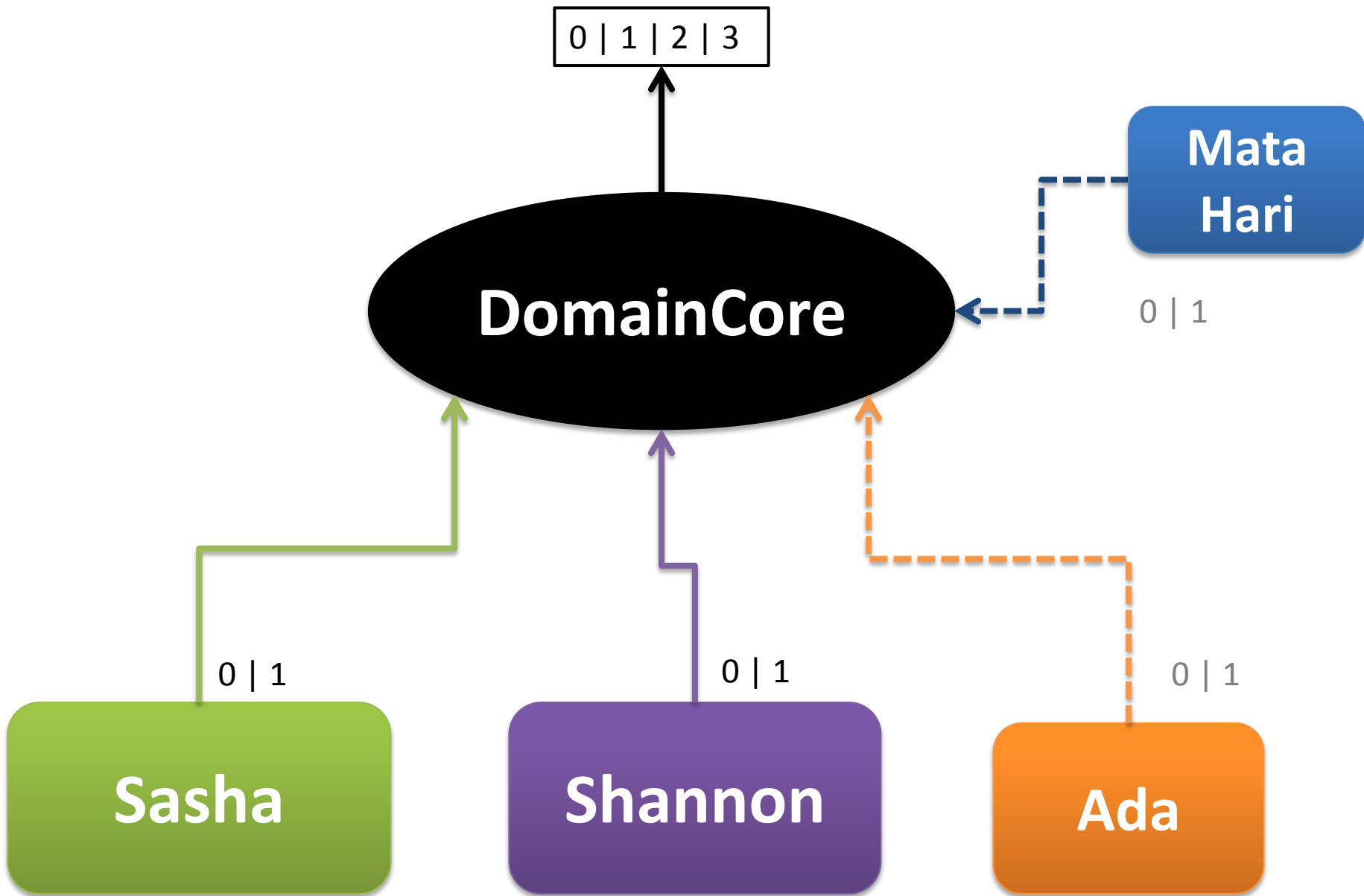
tot / tot = 1

toto / tot[a->o] = 2

toto / toto[+o] = 3

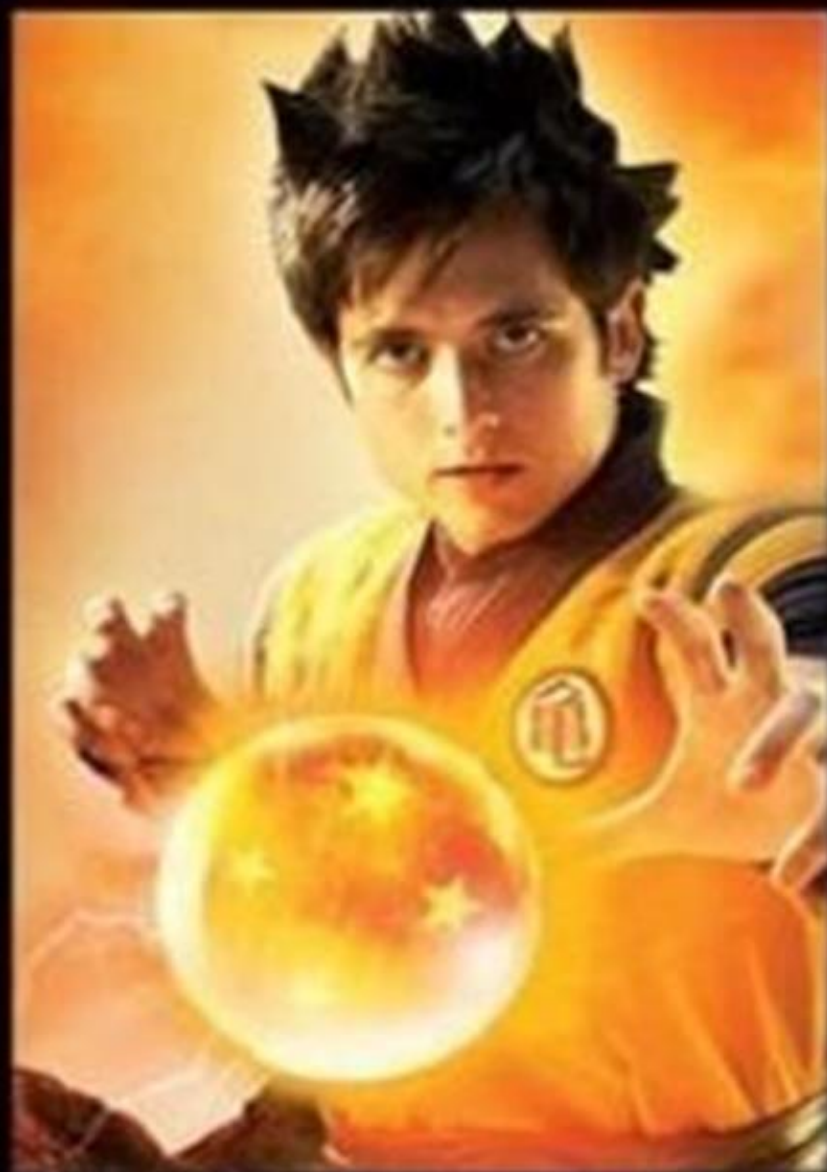
Levenshtein(toto, tatao) = 3







V1



V2

- This tld is know to be evil
- This domain is know to be evil

Ada :

$$P(S|M) = \frac{P(M|S)}{P(M|S) + P(M|H)}$$

Example: The probability for a domain to be malicious (S), knowing that its tld is 'fr' (M) equals to **the probability that 'fr' is malicious** divided by **the probability that 'fr' is malicious** plus the **probability that 'fr' is legit (H)**



TLD	Hit Spam	Hit Ham
.fr	3	7
.co.cc	9	1
.edu	0	10
.com	5	5

Calculs:

$$('fr' = \text{spam}) = \frac{\frac{3}{3+9+0+5}}{\frac{3}{3+9+0+5} + \frac{7}{7+1+10+5}} = \frac{0.17}{0.47} = 0.36$$

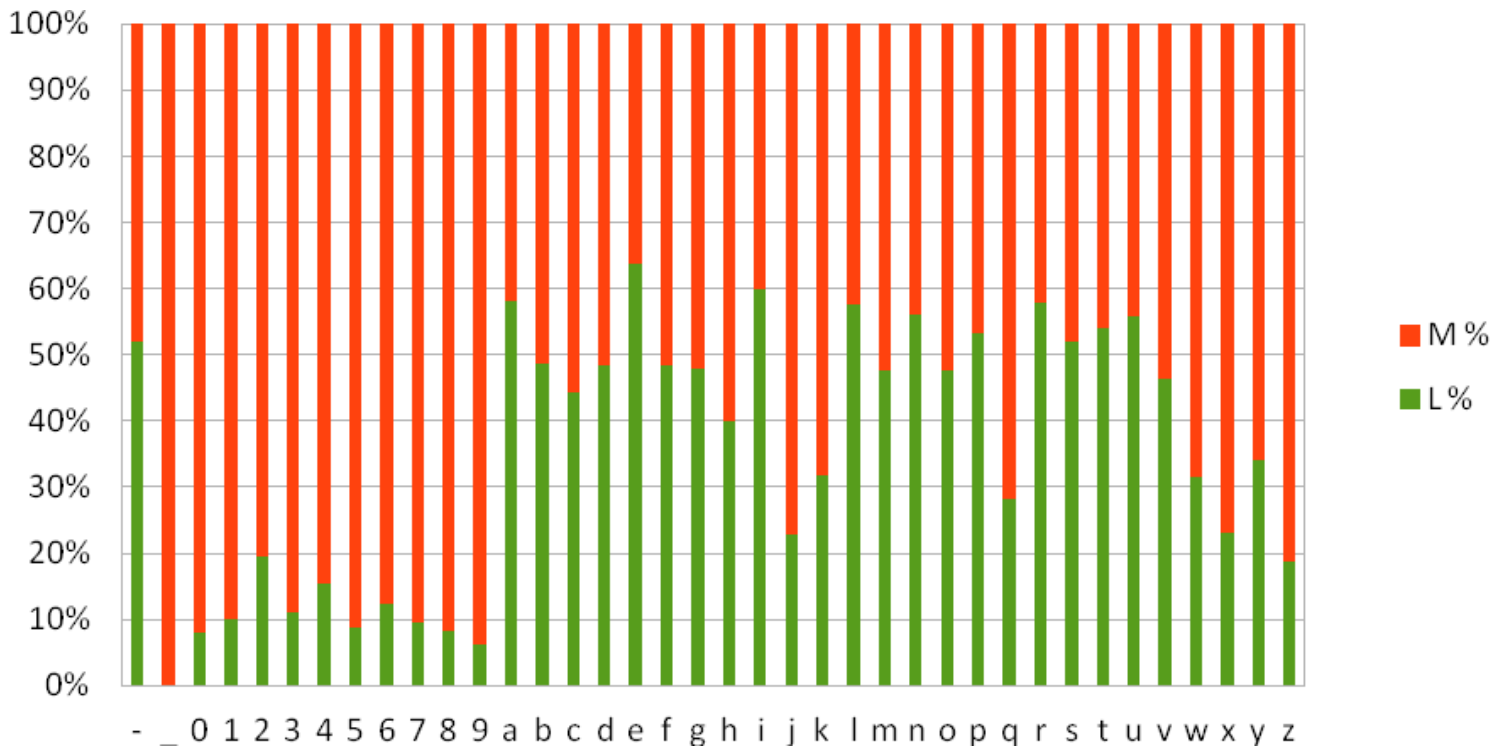
According to the internal Knowledge Base, a 'fr' tld has 36% chance of beeing malicious

DomainCleaner

- Remove the 'www' subdomain
- Transform IDN into latin string
 - xn--pp-bjab.fr > pépé.fr > pepe.fr
- Remove the TLD

Sasha v2

- Speed improvment
 - From 300 ms/FQDN -> 175 ms/FQDN
- Trigger improvment



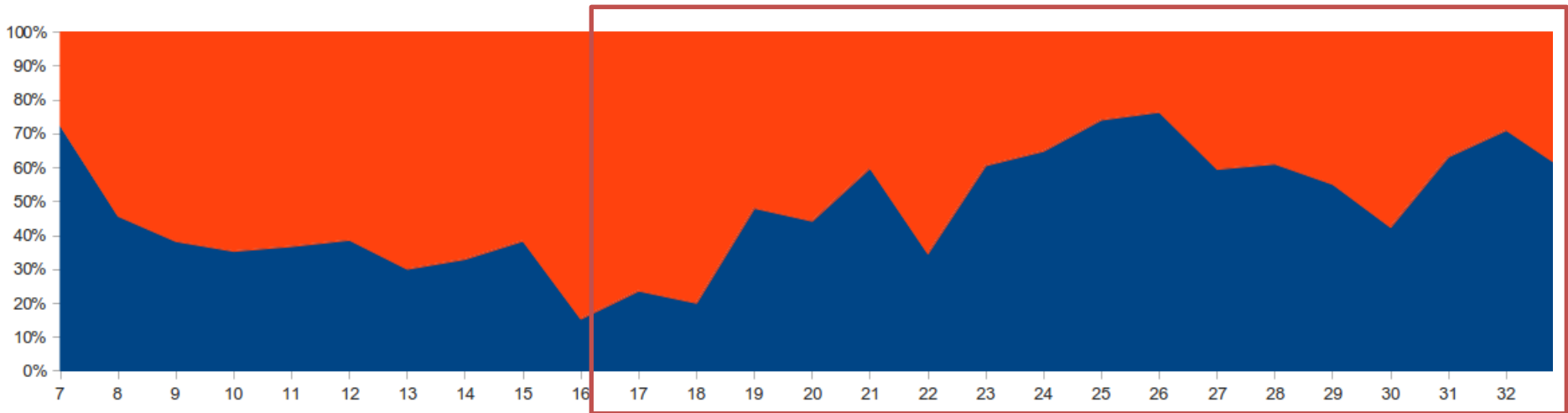
Shannon v2

1/ Levenshtein



Shannon v2

1/ Levenshtein Hum



Shannon v2

2/ Looking for new algorithms

- *Levenshtein Distance*
- *Damerau-Levenshtein Distance*
- *Jaro Distance*
- *Jaro-Winkler Distance*
- *Match Rating Approach Comparison*
- *Hamming Distance*

- *American Soundex*
- *Metaphone*
- *NYSIIS (New York State Identification and Intelligence System)*
- *Match Rating Codex*

Shannon v2

3/ Dictionary Approach improvement

- *Take a string*
- *Split it into bigrams*
- *If bigrams match score >xx% with a dict word*
 - *Append dict word to a shortlist*

stexankmpplex > st,te,ex,xa,an,nk,km,mp,pp,pl,le,ex (len : 12)

justanexample > ju,us,st,ta,an,ne,ex,xa,am,mp,pl,le

match : st, ex, xa, an, mp, pl, le, ex (match = 8)

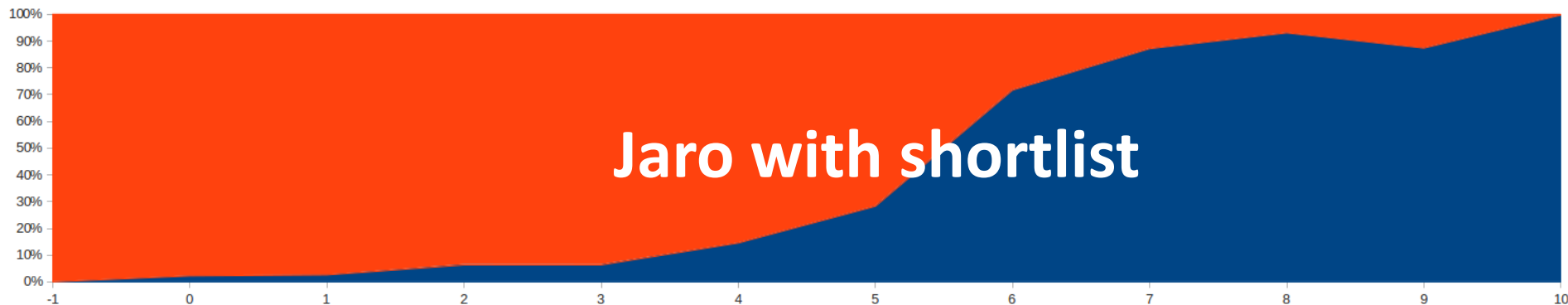
*66%*12 = 8; 8 >= match*

justanexample append to the shortList

Shannon v2

3/ Dictionnary Approach improvment

- *$\text{len}(\text{shortlist}) = 0?$ -> suspicious!*



Shannon v2

4/ The One(s)

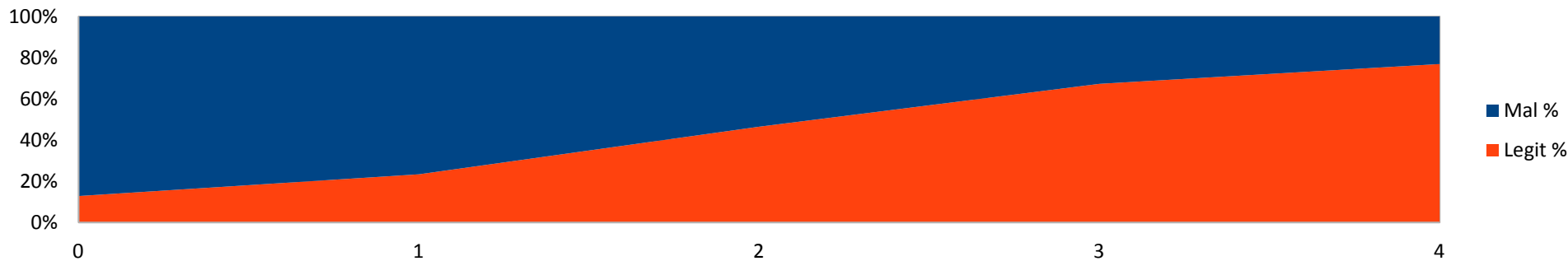
- Damerau Levenshtein (Bigram)
- Jaro + MRC

MRC(stexankmpplex) = STXPLX

MRC(justanexample) = JSTMPL

jaro(stexankmpplex, justanexample) = 80%

jaro(STXPLX, JSTMPL) = 77%



Shannon v2

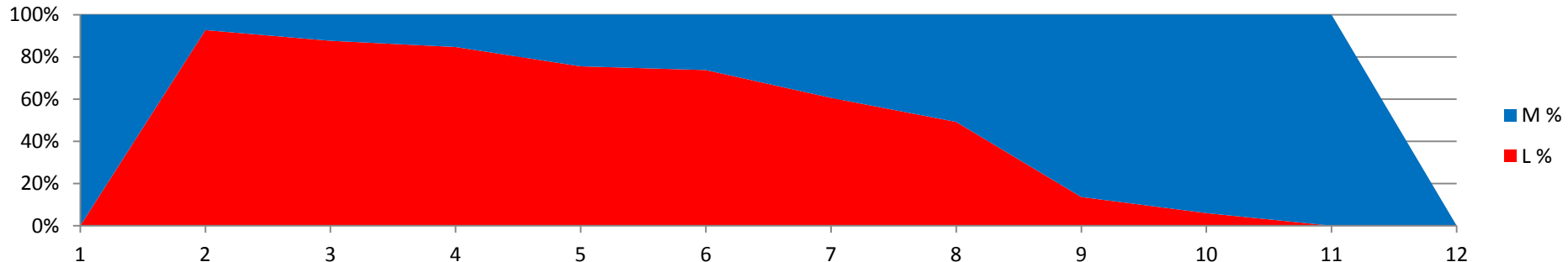
4/ The One(s)

- Jaro (Bigram)
- Match Rating Comparaison (Bigram)

`match_rating_comparison(money,maunnie)` = True

`match_rating_comparison(stexankmpplex,justanexample)` = False

- Jaccard (Bigram)

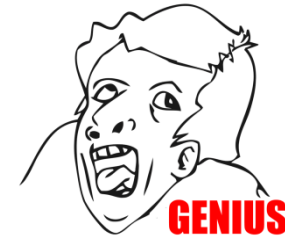


Results

- In Canari's context :
 - 100% DGA
 - 80% overall malicious domain name
 - 10% False positive
 - Hash in subdomain
 - 118712.fr
- European network traffic dump :
 - 40% False Positive
 - East European domain name
- Post development test (RunForestRun 2) :
 - 100% of detection

Looking back ...

- Use stat dataset for validation ...
- Too much improvment -> forget the target (DGA)
 - ADA useless in DGA's case
- But :
 - RunForestRun 2 detection
 - Not a lot of work in 2012(open source PoC, research, technical paper, ...)

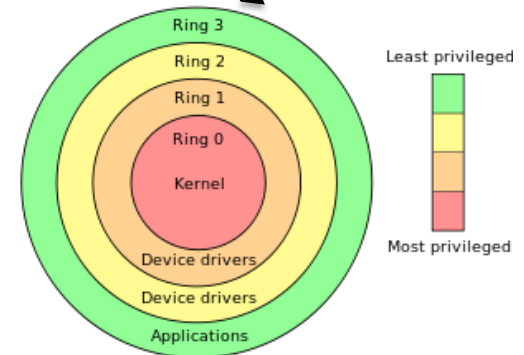
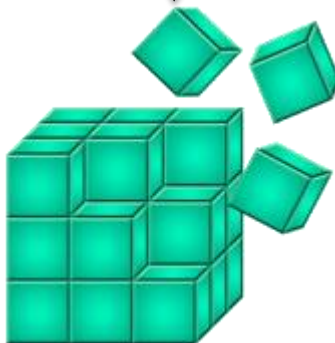


Some leads

- Data Mining
- Multi-Criteria Decision Analysis
- Nothing powerfull as *Common Sense*
 - > *Lessons learned from Robotics applied to CyberSecurity*
by Teresa Escrig (Jaume University) & Sam Chung (University of Washington)
International Journal of Computer Applications – vol 74 n°8 – July 2013
- Don't be blinded by technics, thrill, ... aim your target!
- Take a look at Wordstats (Splunk's Add -on)



HOW

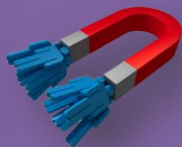


Intelligence

Why?



How?

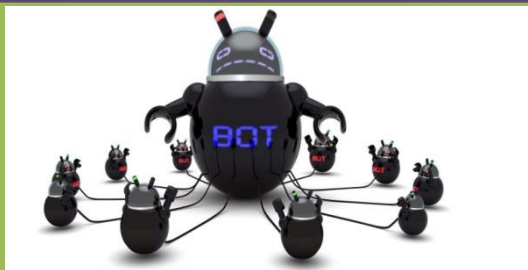


Why?

How?

Network

Why?



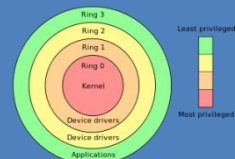
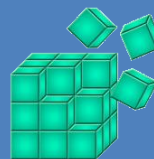
How?

Why?



How?

System



Criminology



Intelligence

↑ Why?

↓ How?

↑ Why?

↓ How?

Network

↑ Why?

↓ How?

System

↑ Why?

↓ How?

Botconf



Thank you!

Contact :

DZKSJEGGACJHPGFHACDGBGOGACCHFGFBGEGACEHIGJGDHMCACHHFGACNGBGJHACIG
BGGHFGACDHPGNGFGACDGPNGNGPGOGACJGOGEHFGCHFGDHEHBCACCHPGOGBGO
GOCNGPGFHDGIGPGFHIHAEDGJHCGFGCHBGDGEHJGPGOGMGBGCGOCOGFGEHDZJS



*Votre adversaire n'est pas le malware,
mais la personne qui est derrière*