Data Collection and Analysis

At High Security Lab

Jérôme François, jerome.francois@inria.fr

MADYNES

Loria
Laboratoire lorrain de recherche en informatique et ses applications

cnrs

UNIVERSITÉ DE LORRAINE

1

Overview

Generalities and objectives

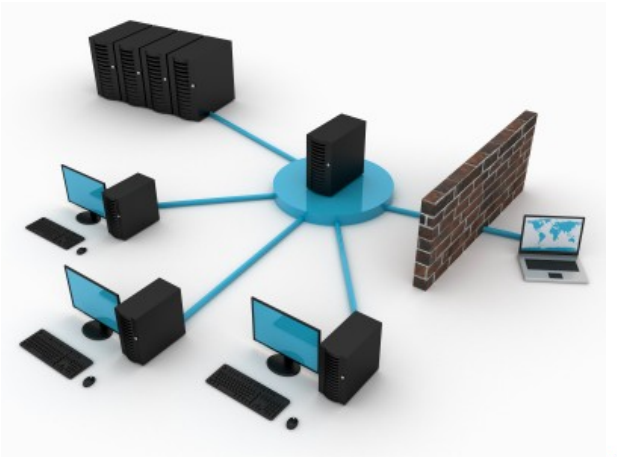
Objectives of the High Security Laboratory

A unique academic platform in France for

- Collecting and analyzing various security related data
- Hosting in a secure environment
- Providing point-of-presences in Internet for Security expreiments
- Contained sensitive execution

To reach major research results

- Pro-active defense against malwares and new threats
- Large scale experimentation and studies, publications
- Implementation and distribution of tools and software
- Validate and distribute research results
- http://lhs.loria.fr



Security of the LHS

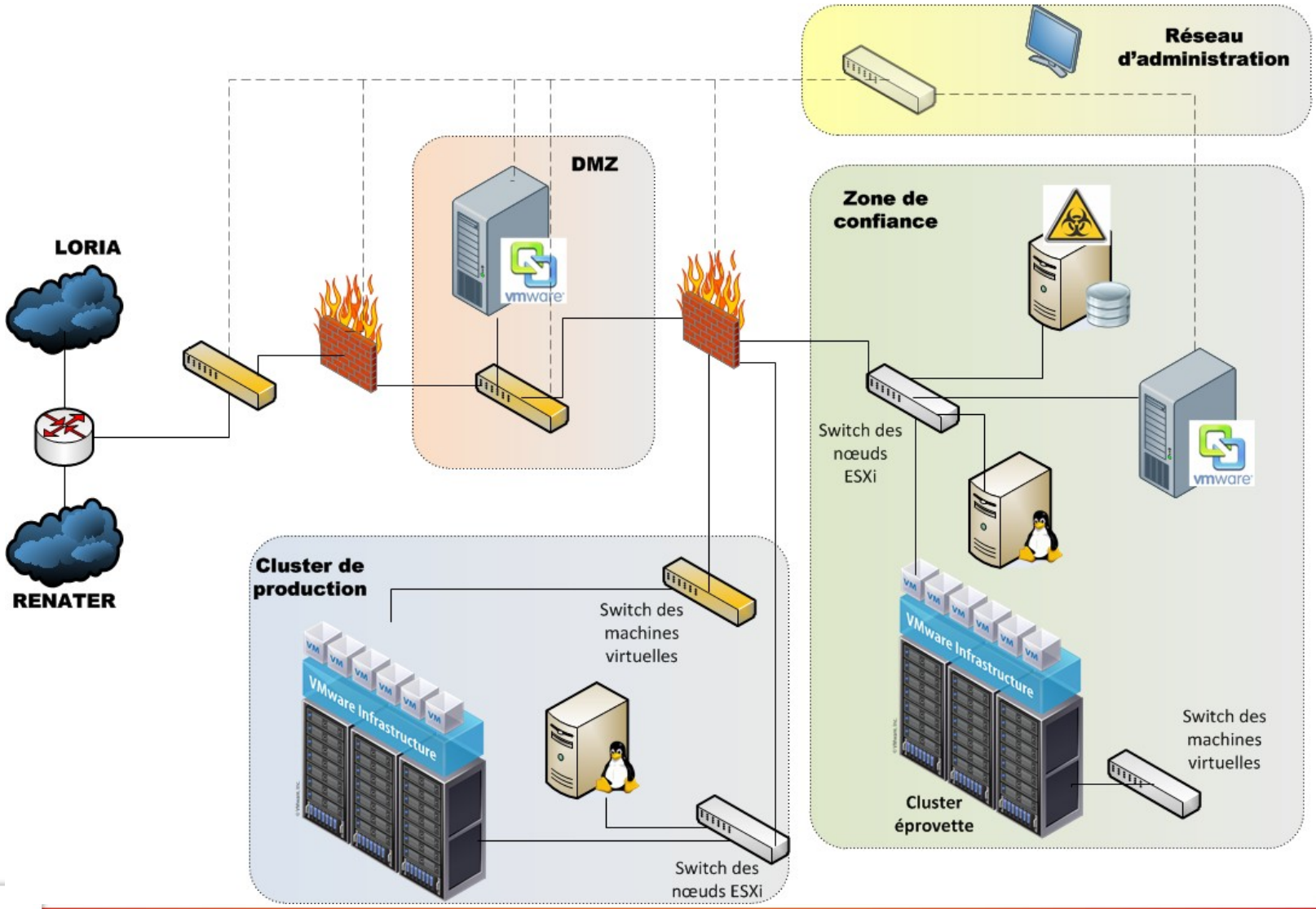
Dedicated and isolated infrastructure

- Workspace separated and “isolated”
- Self-sufficient (electricity, air conditionning)
- Separated network
  - Can simulate a virtual Internet
- **DMZ for results dissemination and collaborations**

Enhanced security

- Different areas with different security levels
  - Office > Servers room > “Red room”
  - “Red room” completely isolated, meant to store and treat sensitive information
- Strengthened access control
  - Strong authentication (entry pass + biométry)
  - Armoured doors and windows, alarms, airlock...

Virtualized and Isolated Architecture



2

Network telescope (Operational since 2008)

Network telescope objectives

Malicious code and binaries capture

- Vulnerabilities emulation
  - Avoid probes compromission and attacks propagation
- Malwares capture (binaries)
- Sandboxes and AV used to analyse and identify the malwares
- Collect all information regarding the attacks
  - Source IP, geographical location, server hosting the binary, preparations
- Zero-day attacks capture to define pro-active defenses

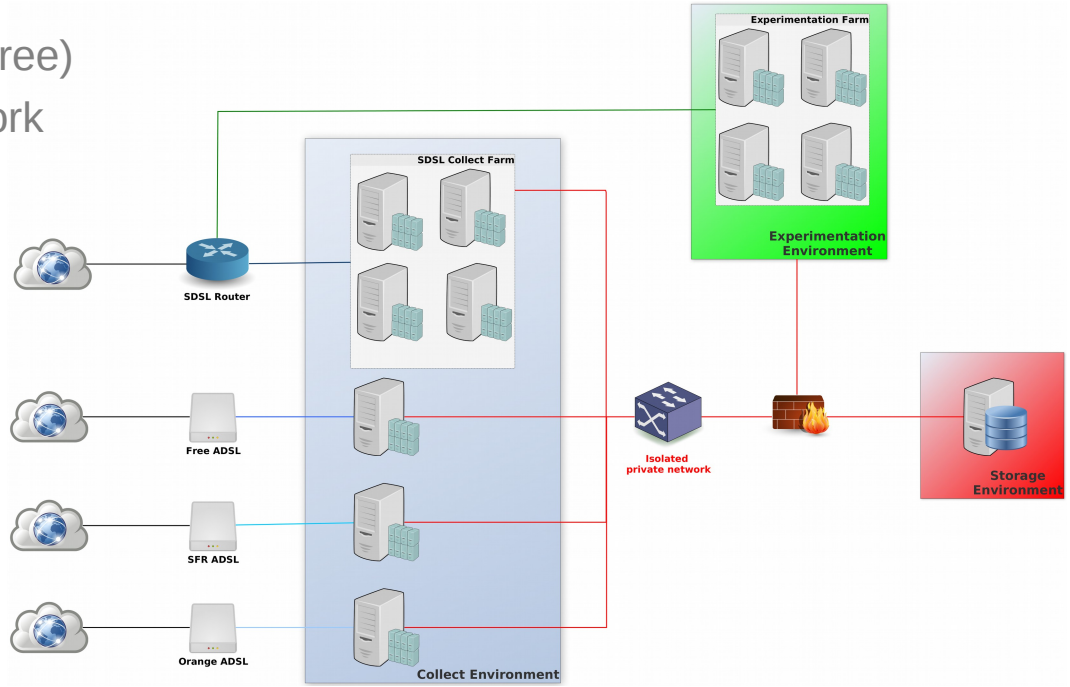
Network flows and traces capture

- Capture in PCAP and NetFlow of the attack traces
- Infection and propagation mechanisms analysis
- Objective
  - Definition of pro-active perimetric defenses
  - Block the attacks at their source

Network telescope

Large scale malwares and attacks traces collect

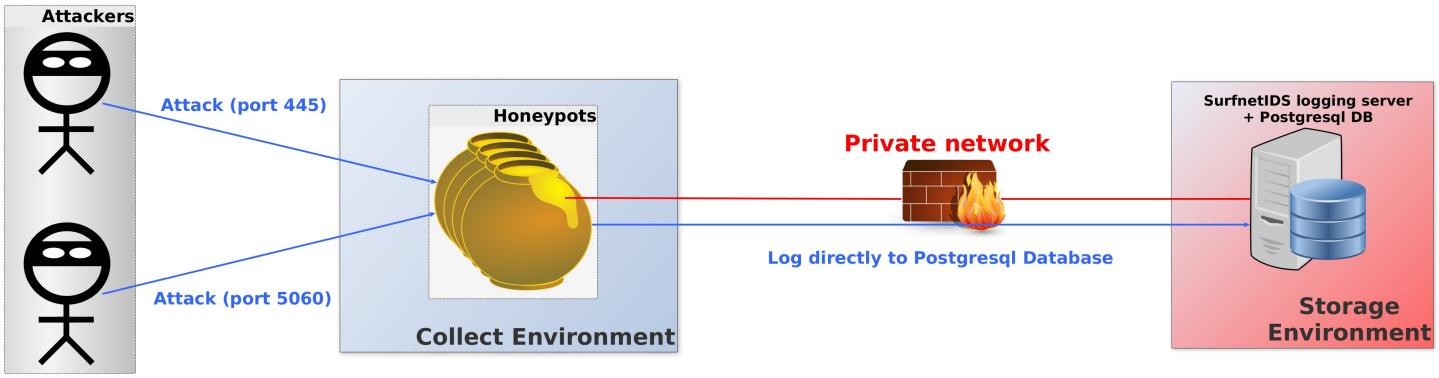
- Multi-provider architecture
  - 3 public ADSL (Orange, SFR, Free)
  - 1 SDSL 2Mbits with a /24 network
- Virtual and isolated architecture
- 3 distinct environments
  - Data collect
  - Data storage
  - Experimentations support



Network telescope v1

Large scale malwares and attacks traces collect using on low interaction honeypots

- Based on SurfNet IDS logging server (http://ids.surfnet.nl)
  - Data stored in a Postgresql DB + Web Interface (stats, maps...)
  - Attack information, geolocation, sandboxing, AV scans…
  - Up to 100 simultaneous honeypots
- Low interaction honeypots logging directly to the Postgresql DB via plugins
- Network traces
  - PCAP via custom scripts
  - NetFlow via fprobe + nfsen and FlowMon appliance



Honeypots and emulated services

Low interaction honeypots

- **25 instances** deployed (around 100 in the very first version)
- Dionaea
  - RPC/Netbios, HTTP, FTP/TFTP, SIP/VoIP, MSSQL
- Amun
  - Vulnerabilities emulated via python plugins
- Kippo
  - Brute-force SSH always works and access to minimalistic shell
  - Sessions and brute-force attempts are logged
- Leurrecom.org Honeypot project
  - Distributed honeypots project, hosting 2 probes
- Glastopf / Glaspot
  - WEB vulnerabilities
- Snort
  - Intrusion detection on the whole SDSL /24 IP range
- In the past
  - Nepenthes, Dionaea ancester
  - Hali in collaboration with the University of Luxembourg, SSH honeypot like Kippo

Some numbers

Operational since the 09th of September 2008

Total (29/10/2014)

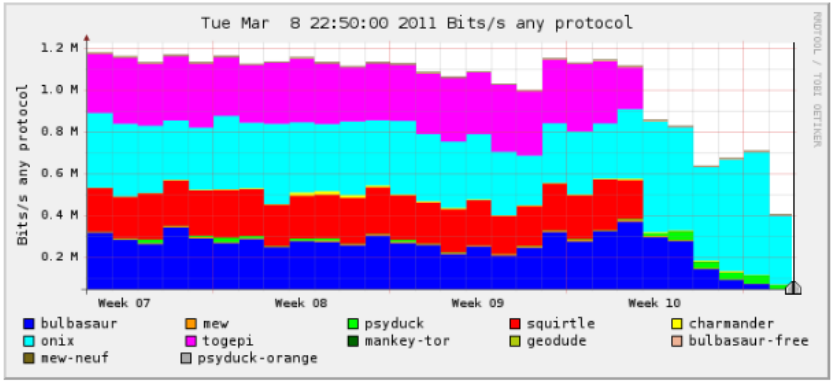
- 901 832 393 attacks
- **368 984 073 malicious attacks**
- 38 878 269 malwares captured
- **301 013 unique binaries**

Daily (on a 800 Kbit/s bandwidth)

- 500 000 attacks - 300 000 malicious
- **25 000 binaries captured**

Network traces

- **15 To of PCAP traces**
- **240 Go of NetFlow flows (v5 et v9)**
- **6 Go of anonymized Tor flows**



Limitations

Based on « old » technologies

- Database (very) slow
- Not scalable
- Outdated sensors integration (e.g. snort plugin)
- Difficult to integrate new data sources

Not originally designed for this kind of deployment

- Meant as a realtime distributed IDS sending alerts
- Not designed to collect and store data over a long period
- Not meant to deal with 100+ sensors
- Loss of information due to SQL schema

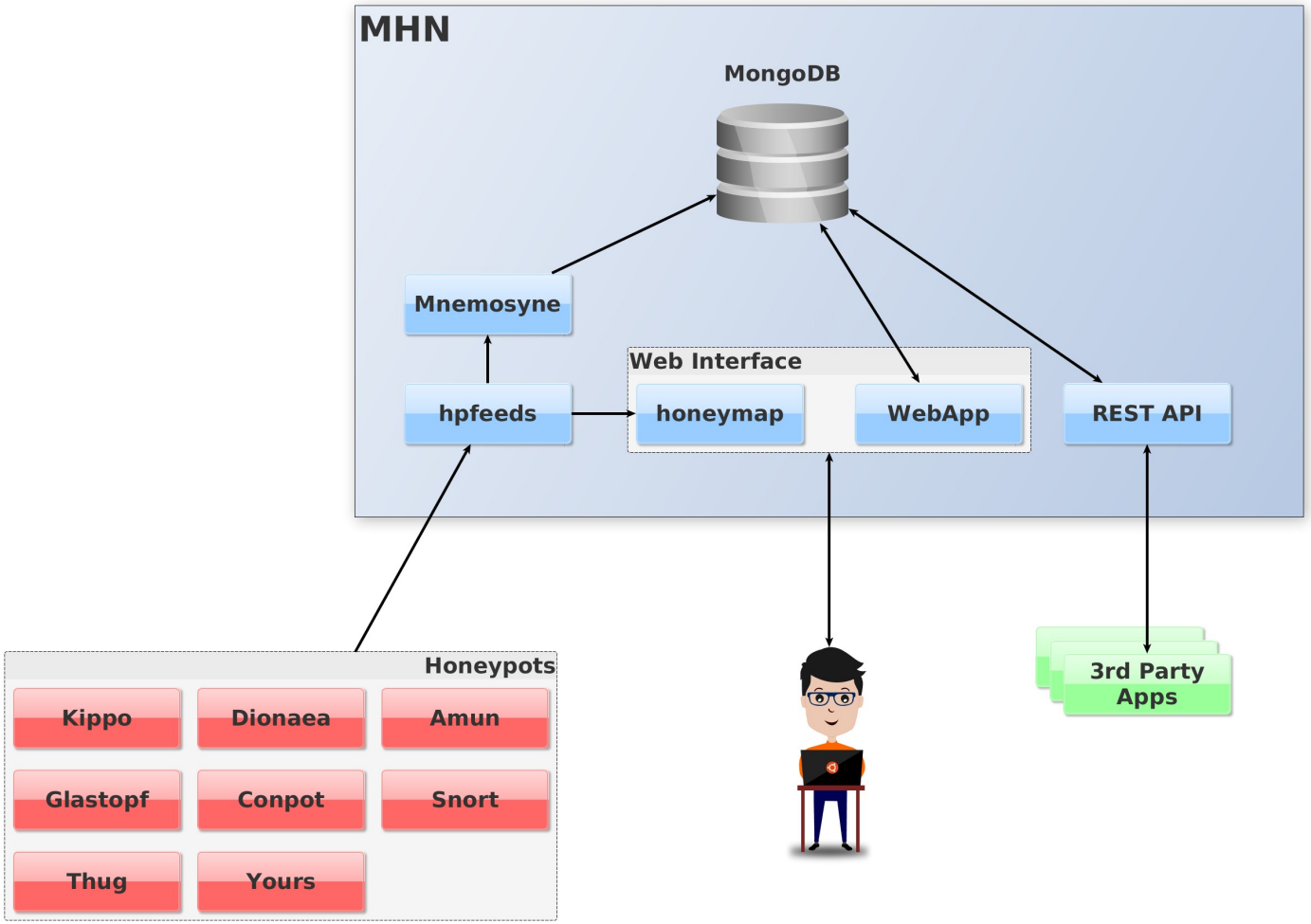
Need a new deployment based on modern technologies and solutions

Modern Honeypot Network - MHN

Centralized server and tools to manage honeypot networks

- Deploy and aggregate honeypots
- Designed for large and distributed honeypot networks
- Data stored in **MongoDB**
- Sensors log via **HPFeeds**
  - lightweight authenticated publish-subscribe protocol
  - supports arbitrary binary payloads
- Data normalized via **Mnemosyne**
  - Provides immutable persistence for hpfeeds
  - Normalization of data to enable sensor agnostic analysis
  - Expose the normalized data through a RESTful API
- Attacks stream visualized with **Honeymap**
  - Reads hpfeeds live stream
  - Displays GPS locations on a SVG world map
- http://threatstream.github.io/mhn/

MHN - Architecture



Honeypots and sensors

Low interaction honeypots and sensors

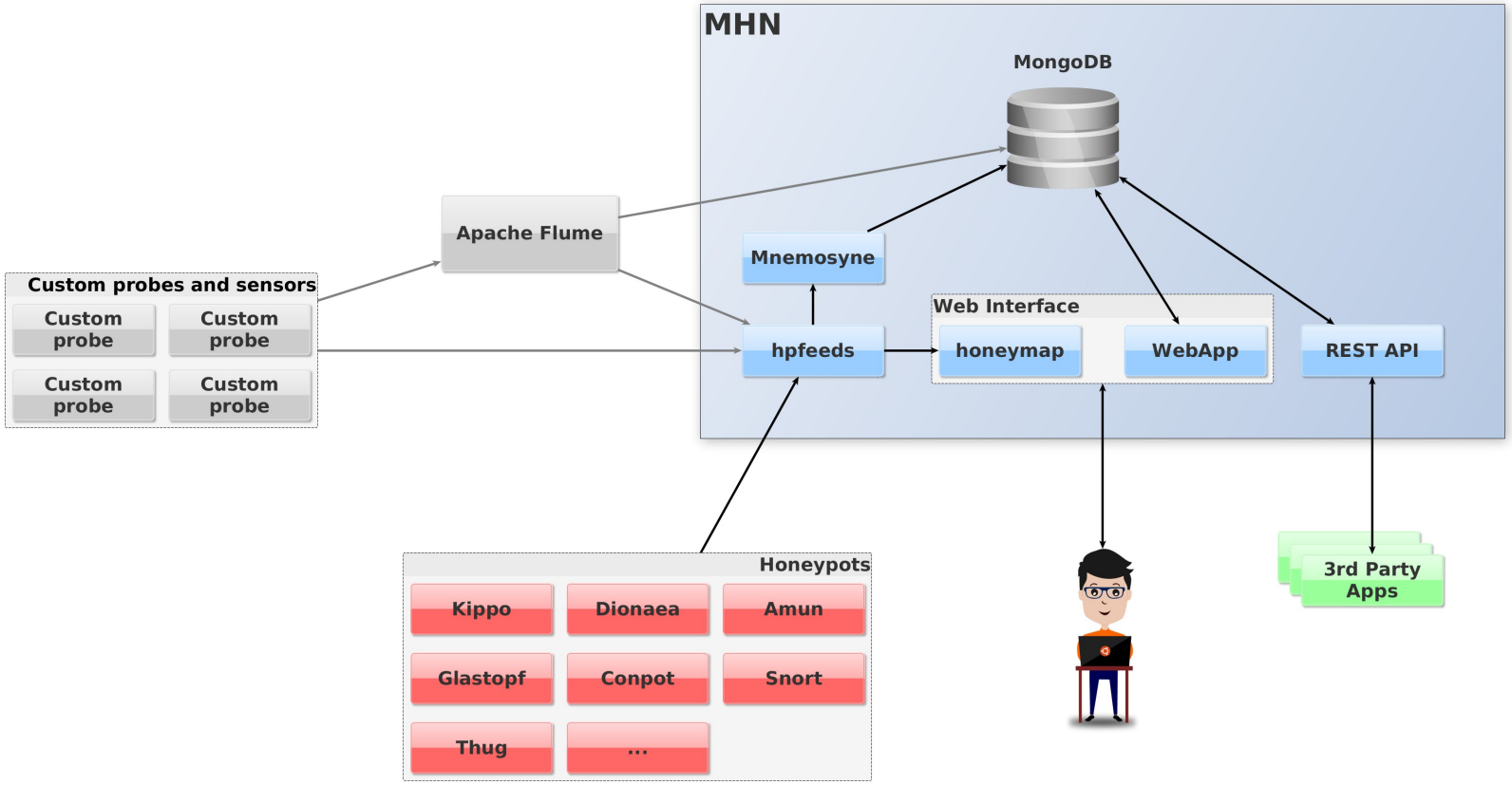
- **1 instance of each deployed** in the current deployment
- Automated deployment via puppet
- **Dionaea**
  - RPC/Netbios, HTTP, FTP/TFTP, SIP/VoIP, MSSQL
- **Amun**
  - Vulnerabilities emulated via python plugins
- **Kippo**
  - Brute-force SSH always works and access to minimalistic shell
  - Sessions and brute-force attempts are logged
- **Conpot**
  - ICS/SCADA Honeypot
- **Glastopf**
  - WEB applications honeypot
- **Snort + snort_hpfeeds**
  - Intrusion detection on the whole SDSL /24 IP range
  - Collector for shipping snort alerts using hpfeeds

Honeypots and sensors

Candidates

- **Thug**
  - Low-interaction honeyclient aimed at mimicking the behavior of a web browser in order to detect and emulate malicious contents
  - Automatic (via blacklists or spams) and manual submissions (portal) of URLs
- **Shockpot**
  - WebApp Honeypot for detecting Shell Shock exploit attempts
  - Working, but no attacks yet (need to investigate)
- **Wordpot**
  - Wordpress honeypot which detects probes for plugins, themes, timthumb and other common files used to fingerprint a wordpress installation.
- **p0f**
  - Passive traffic fingerprinting
- **Custom sensors**
  - DNS, Mobile networks

MHN – Extended architecture

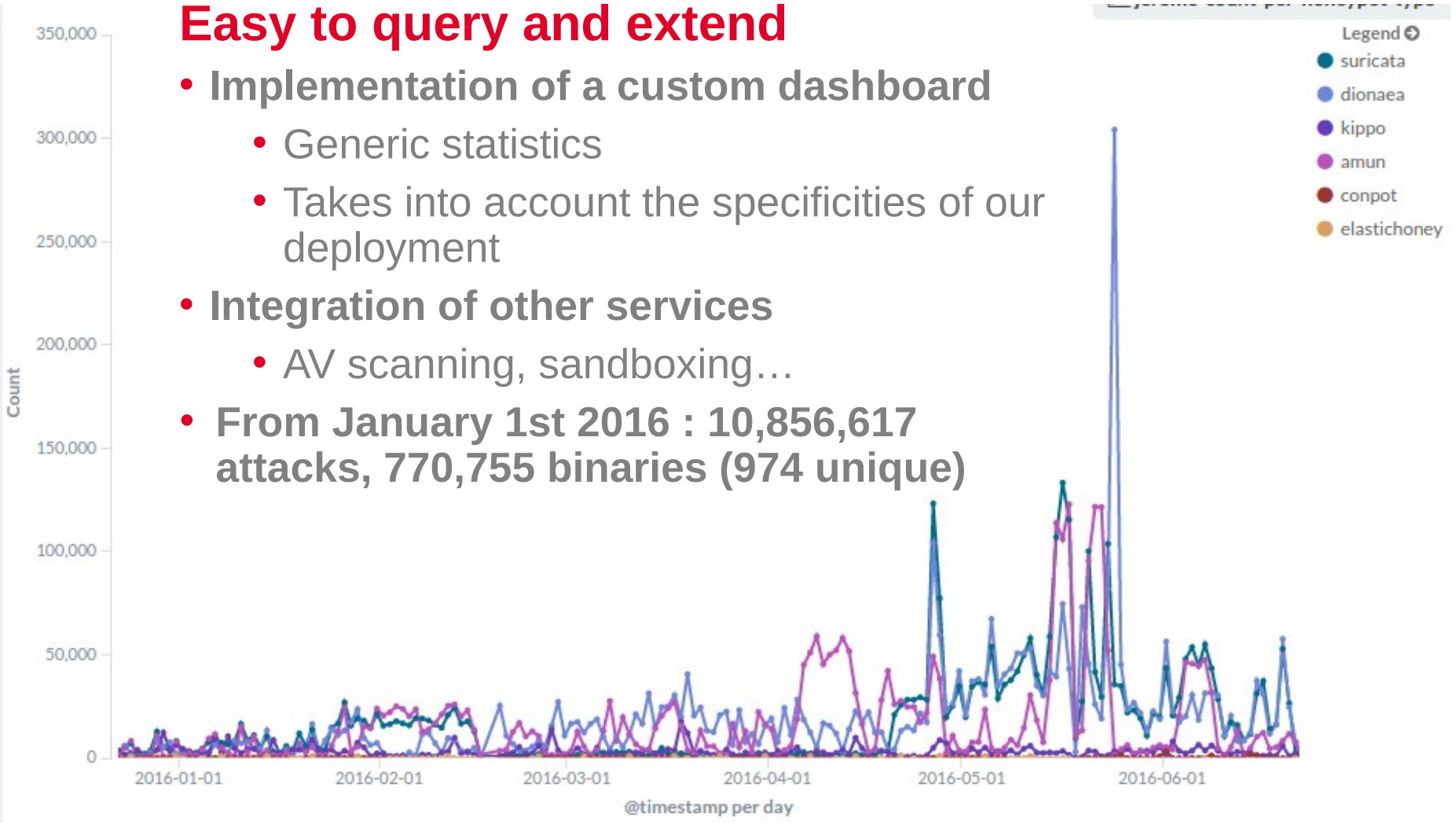


Makes possible the addition of new sensors or a geographical distribution of the honeypots

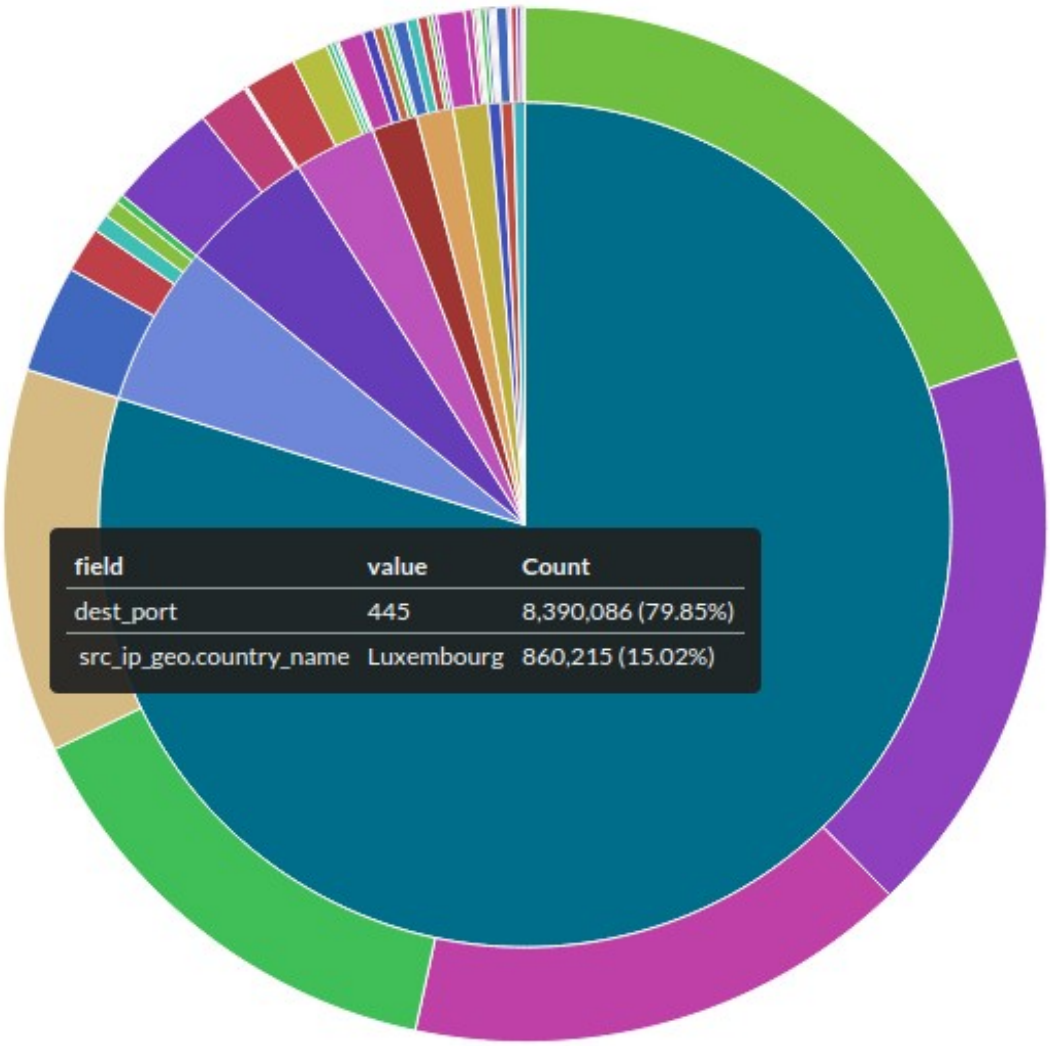
Dashboard

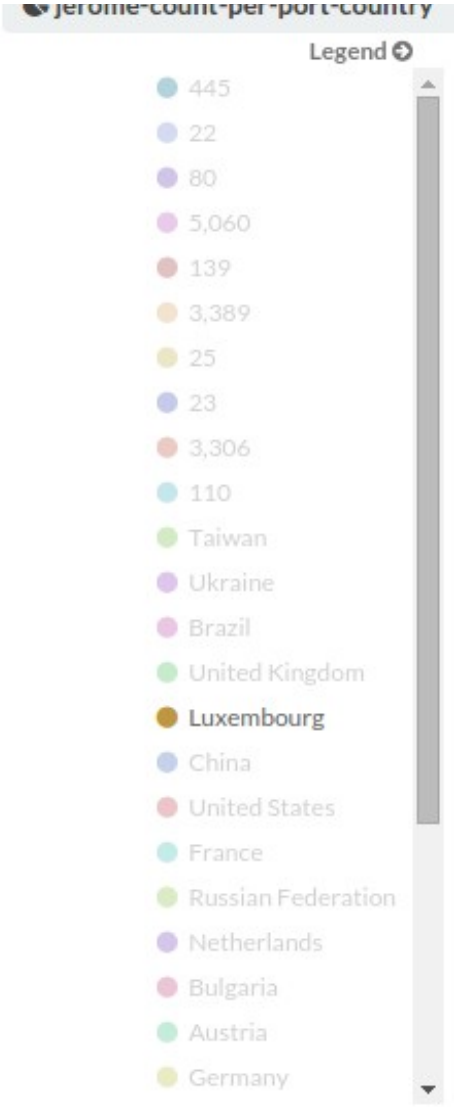
Easy to query and extend

- **Implementation of a custom dashboard**
  - Generic statistics
  - Takes into account the specificities of our deployment
- **Integration of other services**
  - AV scanning, sandboxing…
- **From January 1st 2016 : 10,856,617 attacks, 770,755 binaries (974 unique)**



Dashboard



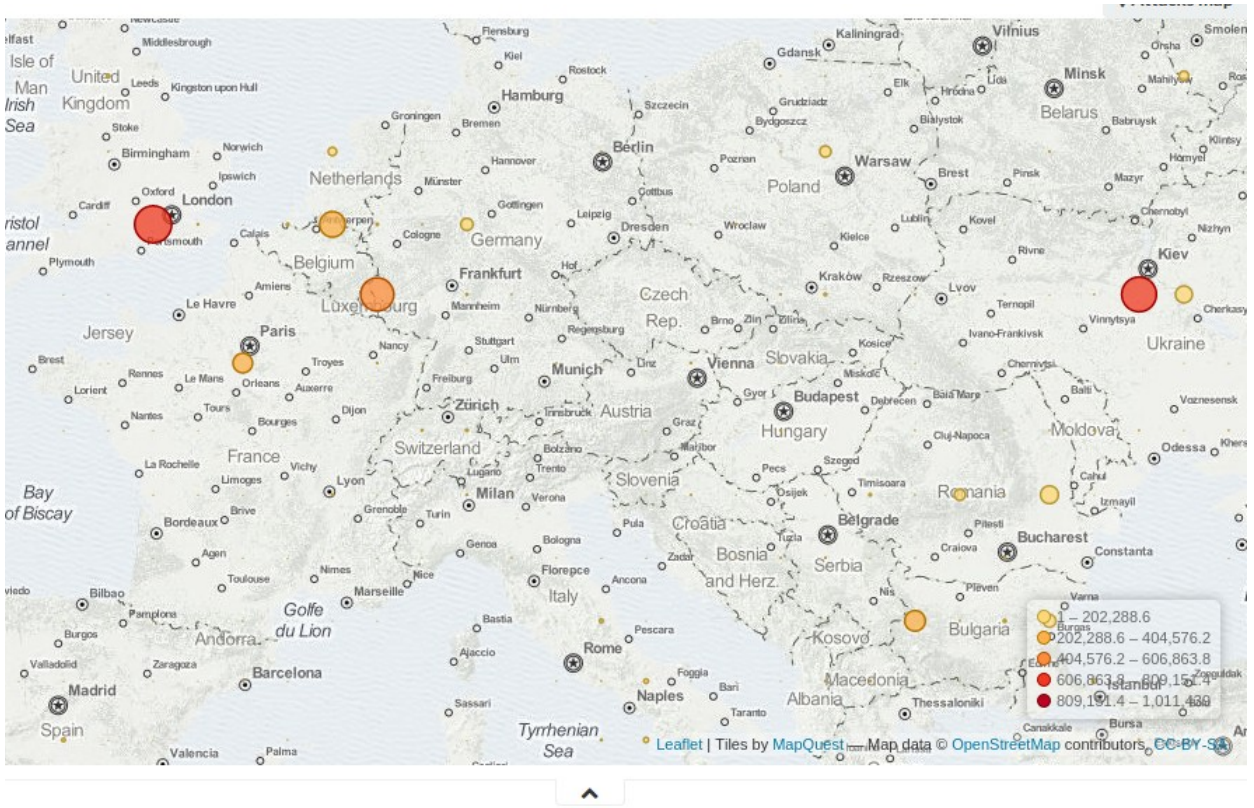


Most targeted ports ? From Where ?

Dashboard

password	Count
123456	7320
!@	5470
password	3641
1234	2481
ubnt	2071
12345	1707
123	1673
	1384
test	1375
1	1243
admin	1120
qwerty	1109
123qwe	1059

Geographic location of attacks



Most used SSH passwords

5

Beyond collected data

Security analytics

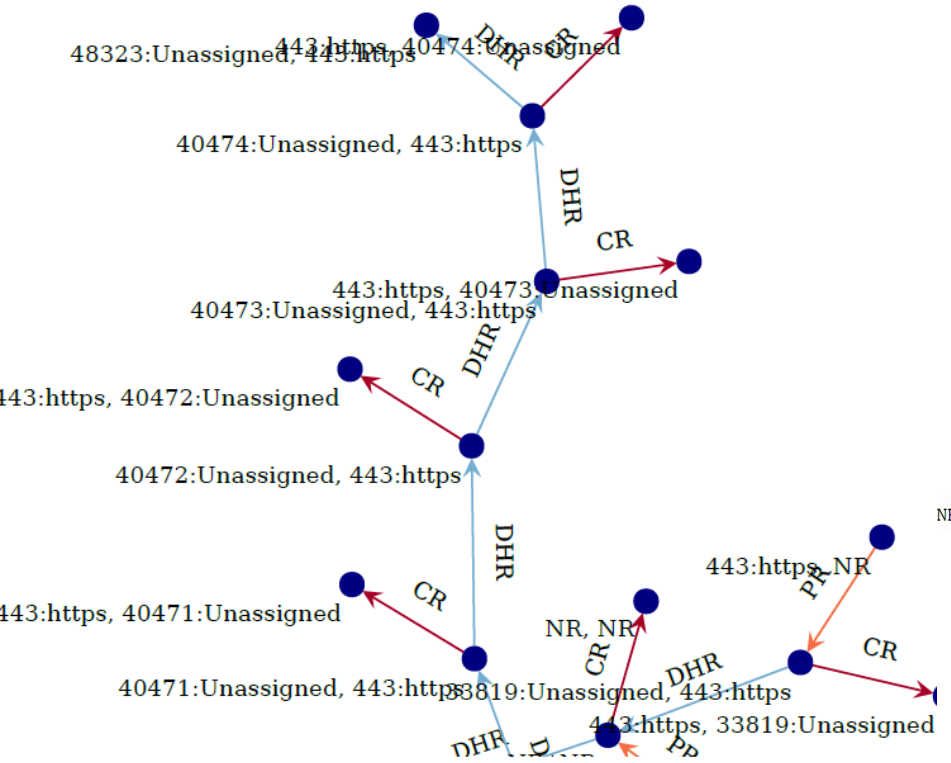
Flow analysis

Objective

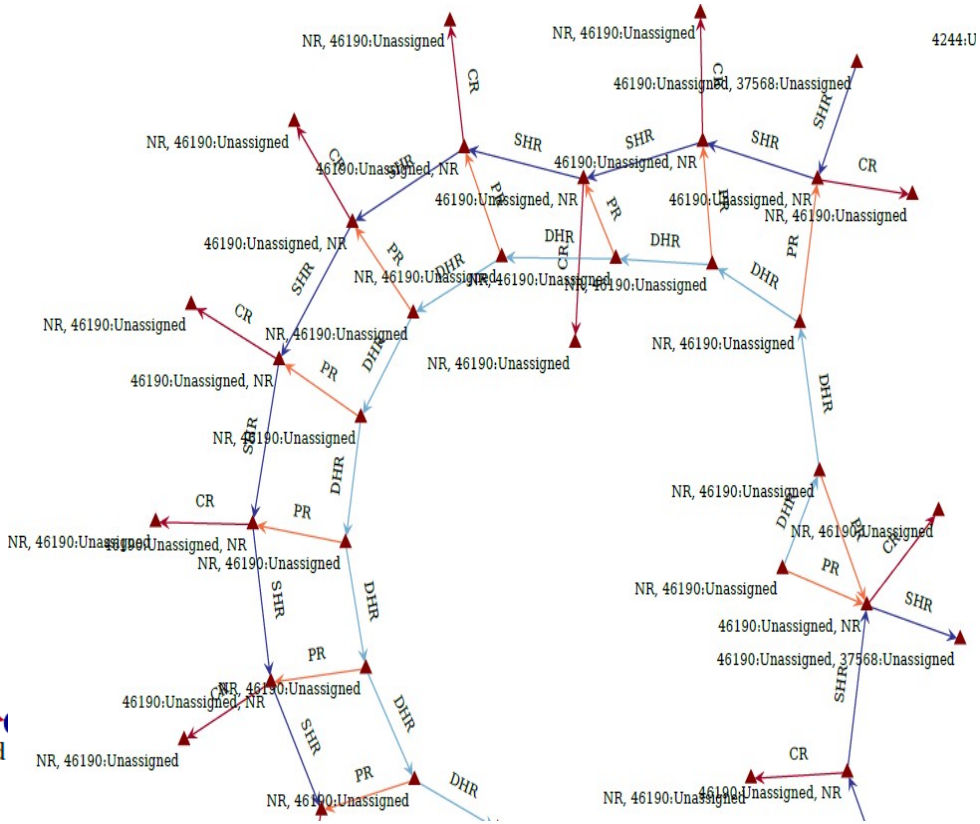
- Profile applications regarding trafic patterns (malware vs benign)
- Aggregation of network trafic
  - *Asai, H.; Fukuda, K.; Esaki, H., ”Traffic causality graphs: Profiling network applications through temporal and spatial causality of flows,” Teletraffic Congress (ITC), 2011*
- Model
  - Vertices are flows and edges are the relations between them (not relations between hosts)
  - Four types of relation based on IP address and port numbers: communication, propagation, dynamic port, static port
  - + reduction rules: limit the number of edges

Flow analysis

Application on Android Applications



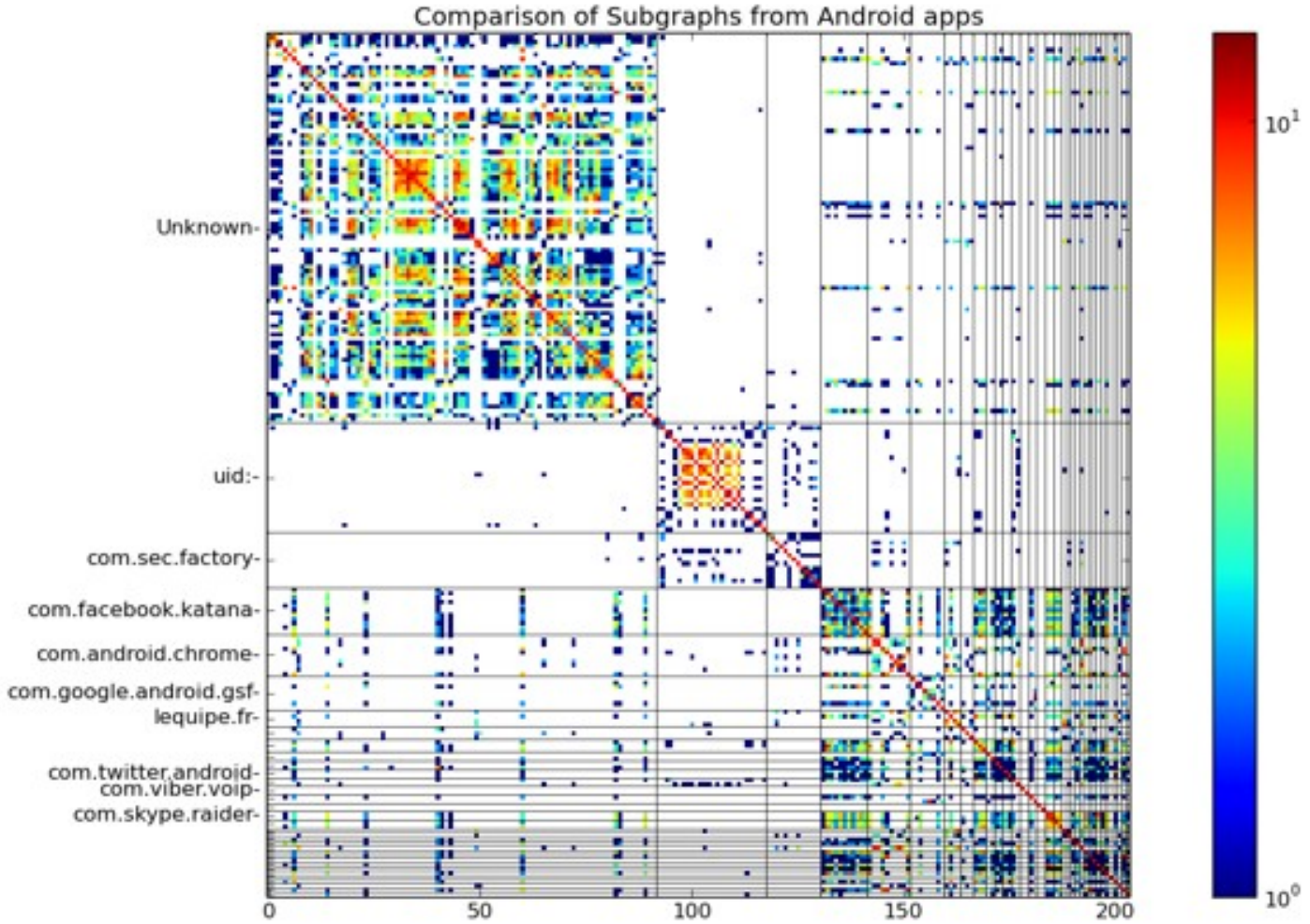
Facebook



Viber

Flow analysis

Frequent substructure mining and comparison



On this example, we rely on benign applications

→ It is necessary for research purposes

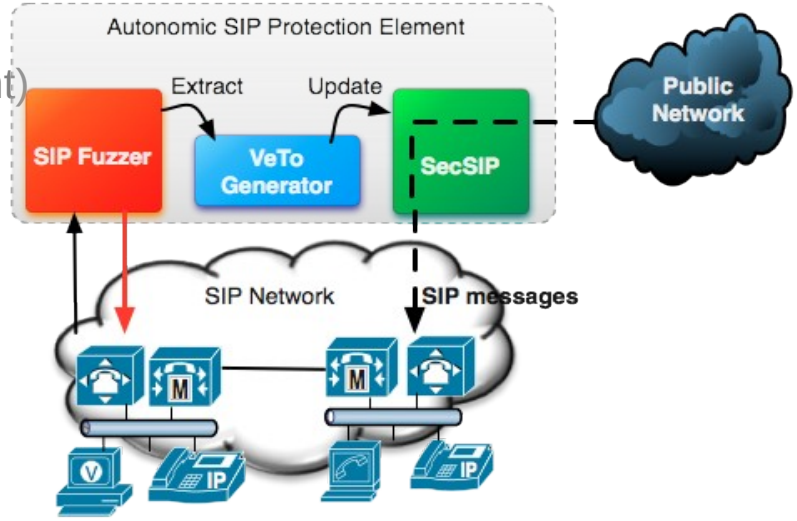
And it was based on Android logs

→ **LHS is not only a telescope, we have and are continuously extending its capability to serve our research and the research of our partners (collecting new data, hosting new types of probes, ….)**

Experimentations support

Experiments

- Vulnerabilities assesssment
  - Fuzzing (KiF), VoIP (SecSIP, Risk management)
- Network monitoring
  - Pedophilia in P2P networks (KAD, Bittorrent)
  - I2P anonymous P2P networks
- Services monitoring
  - Realtime analysis of malicious DNS requests
- Protocols et network mechanisms
  - IPv6, Botnets...
  - e.g. NDPMon, IPv6 Neighbor Discovery Monitor



SCADA platform integration

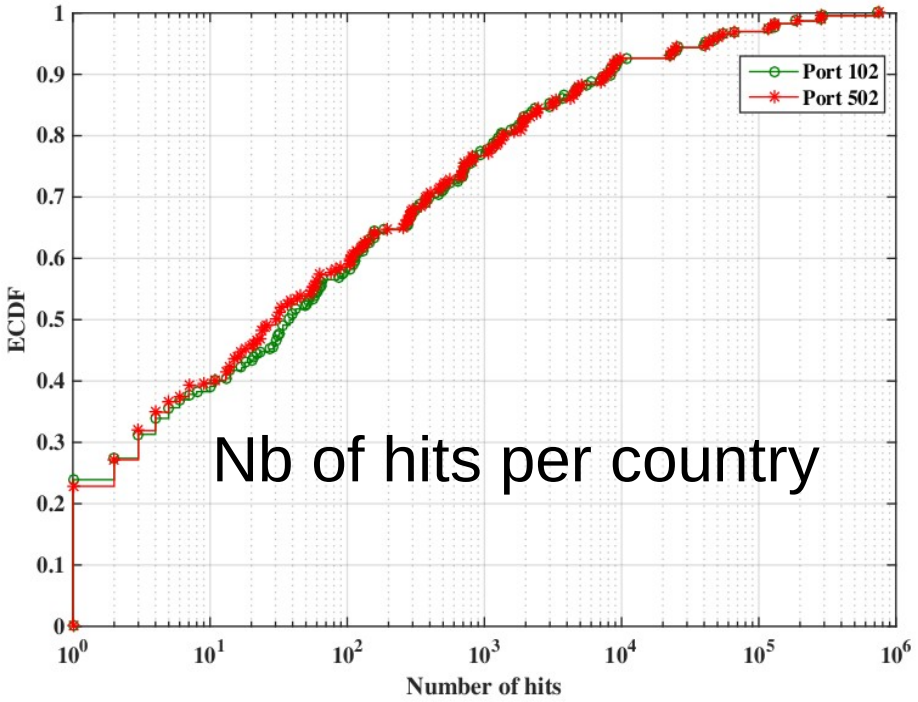
- Simulate physical processes with automated control (PLC Siemens)
- Communication protocols between I/O and controlers analysis (protocol Profinet)
- Attacks scenarios identification, vulnerabilities assessment and counter-measures

Being more active!

New types of experiments

- Most of existing experiments relies on passively collected data
- Doing active security measurement on Internet
  - Ethical and legal issues
  - → we have a special committee at Inria

- First really active experiment : IPV4 scanning
  - Industrial system exposition
  - *Optimizing Internet Scanning for Assessing Industrial Systems Exposure, Jérôme François, Abdelkader LahmadiValentin Giannini, Damien Cupif, Frederic Beck and Bertrand Wallrich,, 7th International Workshop on TRaffic Analysis and Characterization, 2016*



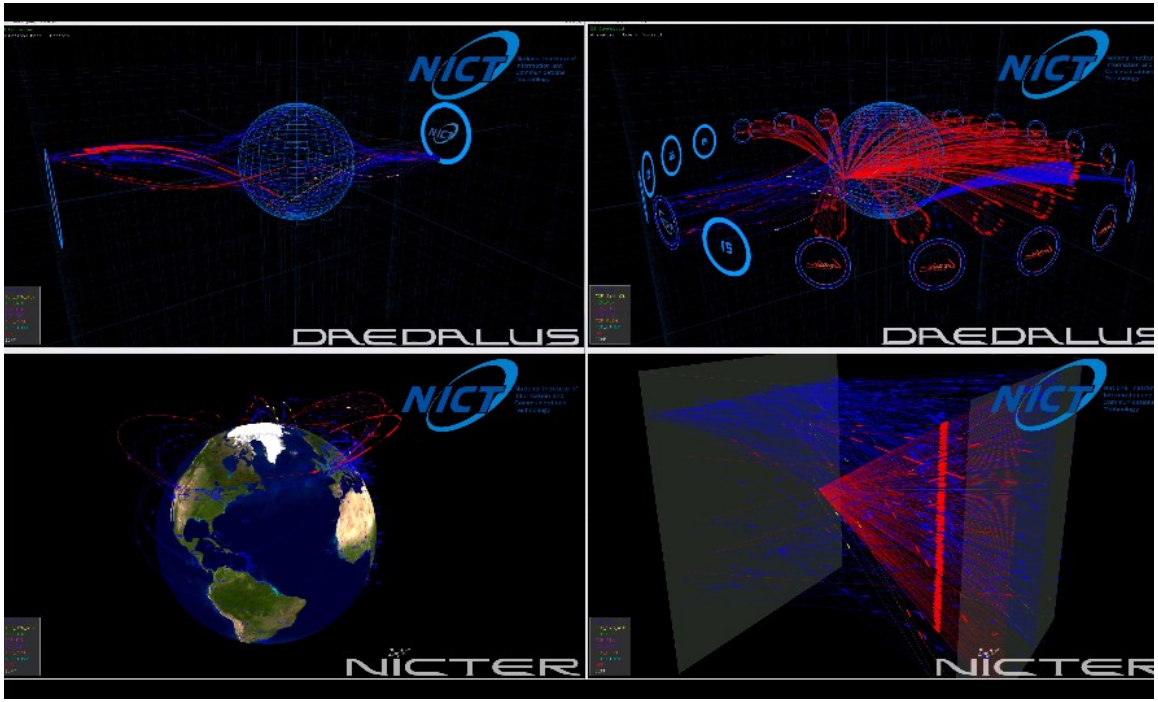
We observe our own scan and the other ones

→ Profiling and correlation

Still observing the Internet... but with larger scope

Darknet

- Actually also known as a telescope, sinkhole
- A large subnetwork which is announced over Internet but with no host
- Input trafic only
- Useful to observe large phenomenons : DDoS, scan, botnets
- In cooperation with NICT, Japan : data sharing and visualisation tools



5

Conclusion

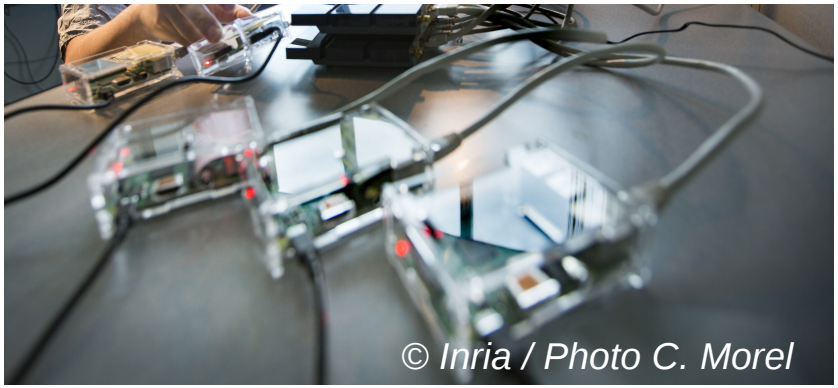
Current and future work

Conclusion

- **Server to store and analyze various large datasets**
- **Next platform updates**
  - Full platform upgrade and new security services
  - High-interaction / active honeypots

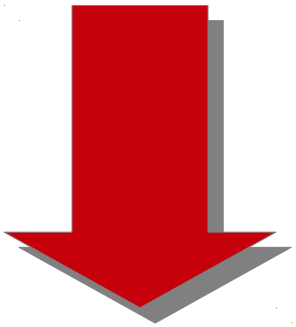
- **Dissemination / dataset sharing**
  - **Anonymize the traces / remove private information**
  - **Correlate the various information captured and offer full attack packages**

- **Distribute the sensors**
  - **Provide a secure platform for data storage and sharing**
  - **deploy sensors in partners networks, Raspberry Pi (or equivalent)**



Thank you for your attention

More detail about our activities ?
Access data ?
Join our team (PhD, engineer,...) ?



Contact us!