Towards attack modelling thanks to honeypot data processing

Marc Dacier

Institut Eurécom

Sophia Antipolis, France

dacier@eurecom.fr

Overview

- Introduction
- *State of Knowledge*
- *Contributions of ReSIST Partners*
- *Conclusions*

Threats?

- *Fact:* New vulnerabilities discovered every day, new widespread attacks reported in the media.
- Questions:
  - Are these vulnerabilities actually exploited?
  - What are the “right” fault assumptions models that one should use to build intrusion tolerant systems?

Dahu: definition

source: http://www.vidonne.com/html/dahu-reignier.html

“The Dahu is an extremely shy animal living in the Alps of France and Switzerland.[…] It has adapted to its steep environment by having legs shorter on the uphill side and longer on the downhill side […]
“

“The Dahu, An endangered Alpine species”, *Science*, 2568, November 1996, pp.112:

Etude morphologique (planche 2)
Dahus rupicapra vacca montanus (vue de face)
oreille pleine de poils
oeil pétillant
pente 50°
hauteur 1 mètre 60
pattes
pattes plus courtes!!!?
1 beurre
1 pain
1 salade
Professeur Henri Henkor - 1862

Food for thoughts …

- *Dahus* are rare, bizarre, stimulating from an intellectual point of view but ...
- Does it justify the existence of *Dahusian research*?
- What about *Dahusian research* in security assessment?

Overview

- Introduction
- *State of Knowledge*
- *Contributions of ReSIST Partners*
- *Conclusions*

The basics

- « A Honeypot is an information system resource whose value lies in unauthorized or illicit use of that resource »

L. Spitzner, *Honeypots: tracking hackers*,
Addislon Wesley, 2002

The basics (ctd.)

- Low interaction honeypots:
  - emulate the existence of a potential target,
  - At various abstraction levels (network, OS, application)
- High interaction honeypots:
  - Use a real system as a potential target
  - Must be kept under close scrutiny.

Internet Telescopes

- Internet Telescopes observe empty address spaces:
  - CAIDA Telescope,
  - IMS,
  - iSink,
  - Minos,
  - Team Cymru,
  - Honeytank,
  - IUCC/IDC Internet Telescope (Israel),
  - Etc...
- The Honeynet Alliance promotes the use of high interaction honeypots.

Problems with current solutions

- **False positives**
  - **It may be difficult to discriminate true attacks from erroneous, yet legitimate behaviours, in data collected in real networks.**
- **Privacy**
  - **Data sets may contain private information (eg IP addresses, passwords, etc.). Anonymisation removes semantic and is therefore not always usable.**
- **Liability**
  - **Not stopping an ongoing attack may harm third parties. Major issue for high interaction honeypot.**

Problems with current solutions (ctd.)

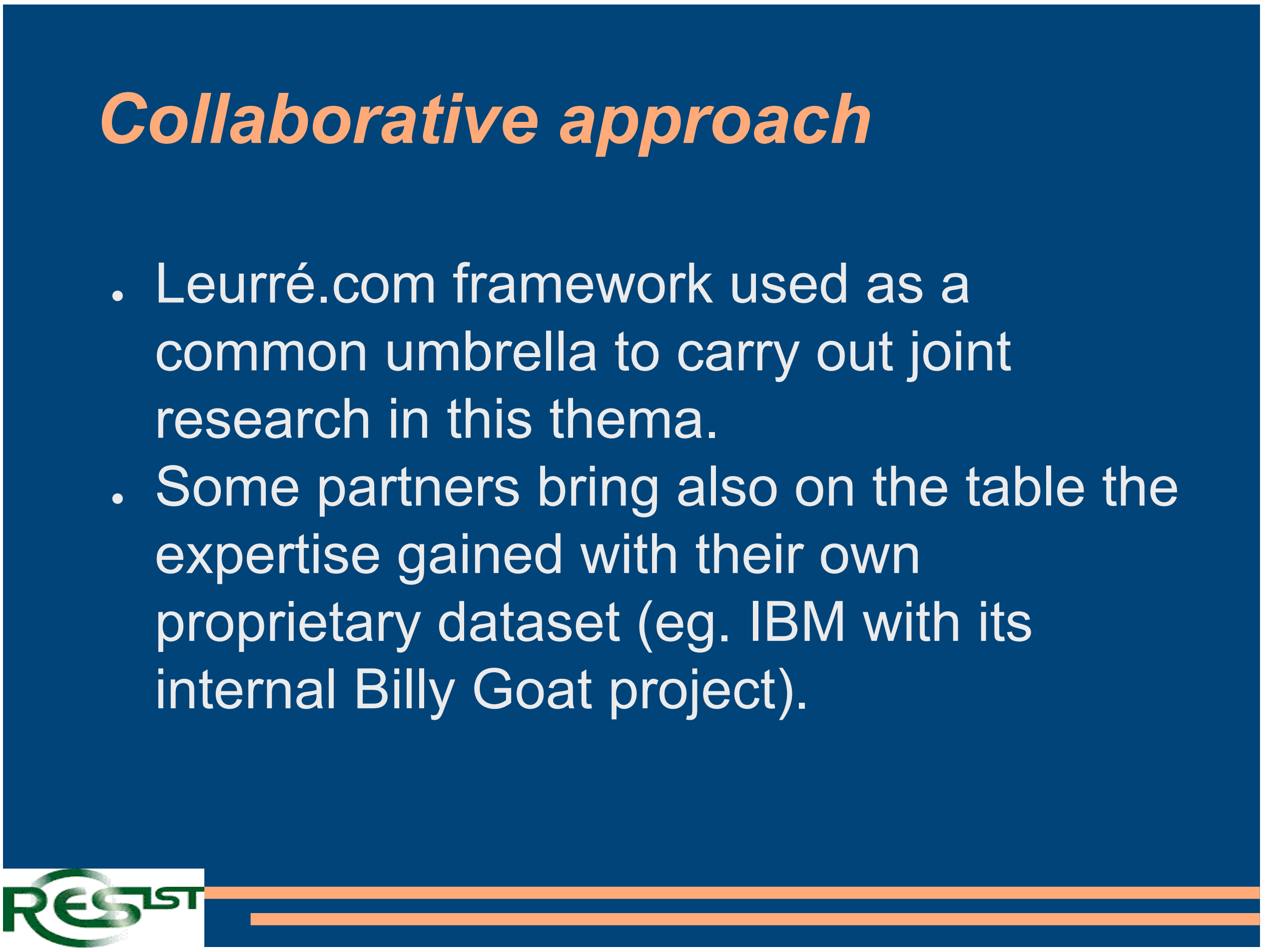
- Bias
  - Things may be different here and there.
  - Malicious users dislike to be observed and will avoid visiting known observation points (eg .mil, major corporate networks, etc..)
- Amount of data
  - Having access to a large amount of data is good
  - Having access to a rich amount of data is better.
  - Having access to a rich amount of complete and comparable data is even better!

Summary

- What we need is:
  - an environment to collect unbiased, rich, complete and comparable data about attacks without facing **liability** or privacy issues.
- To do so, we have deployed:
  - the very same **low interaction** honeypots in a large number of diverse locations using each time a very limited amount of IP addresses. We collect all packets sent to or from these machines, including payload.

Overview

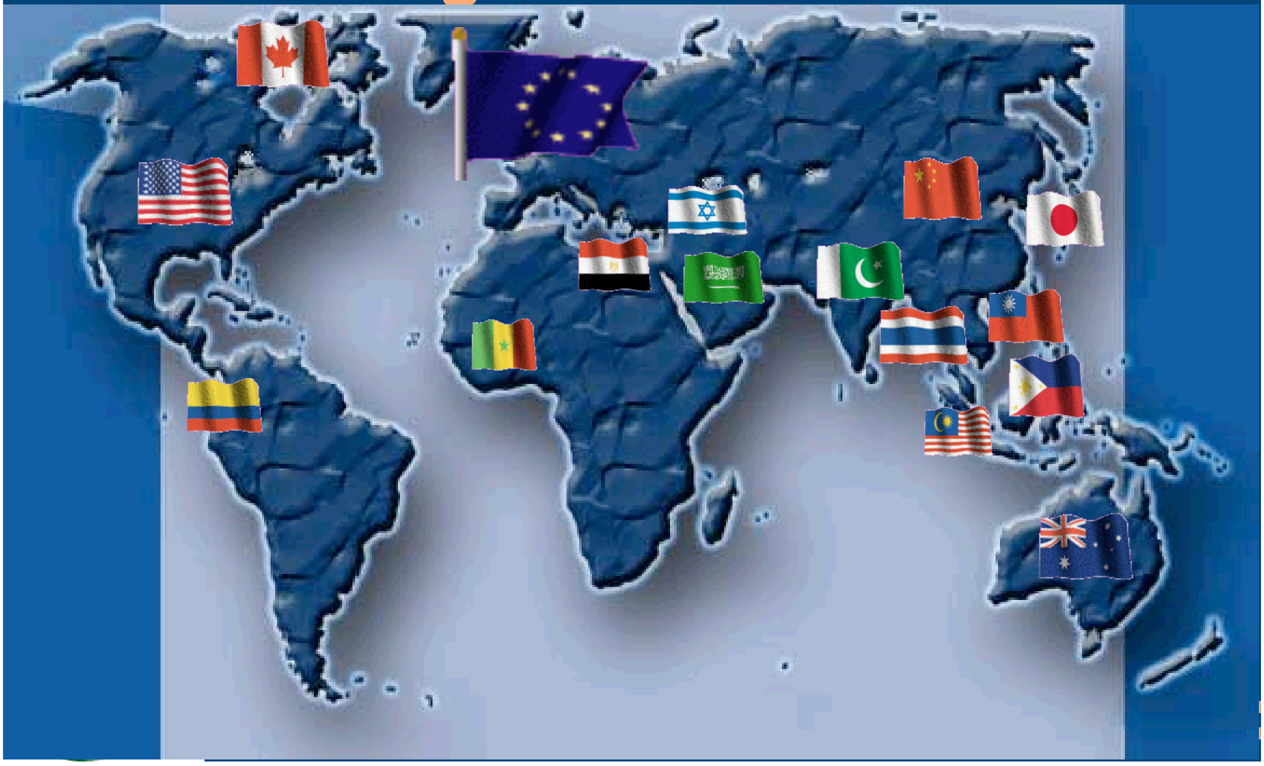
- Introduction
- *State of Knowledge*
- *Contributions of ReSIST Partners*
- *Conclusions*

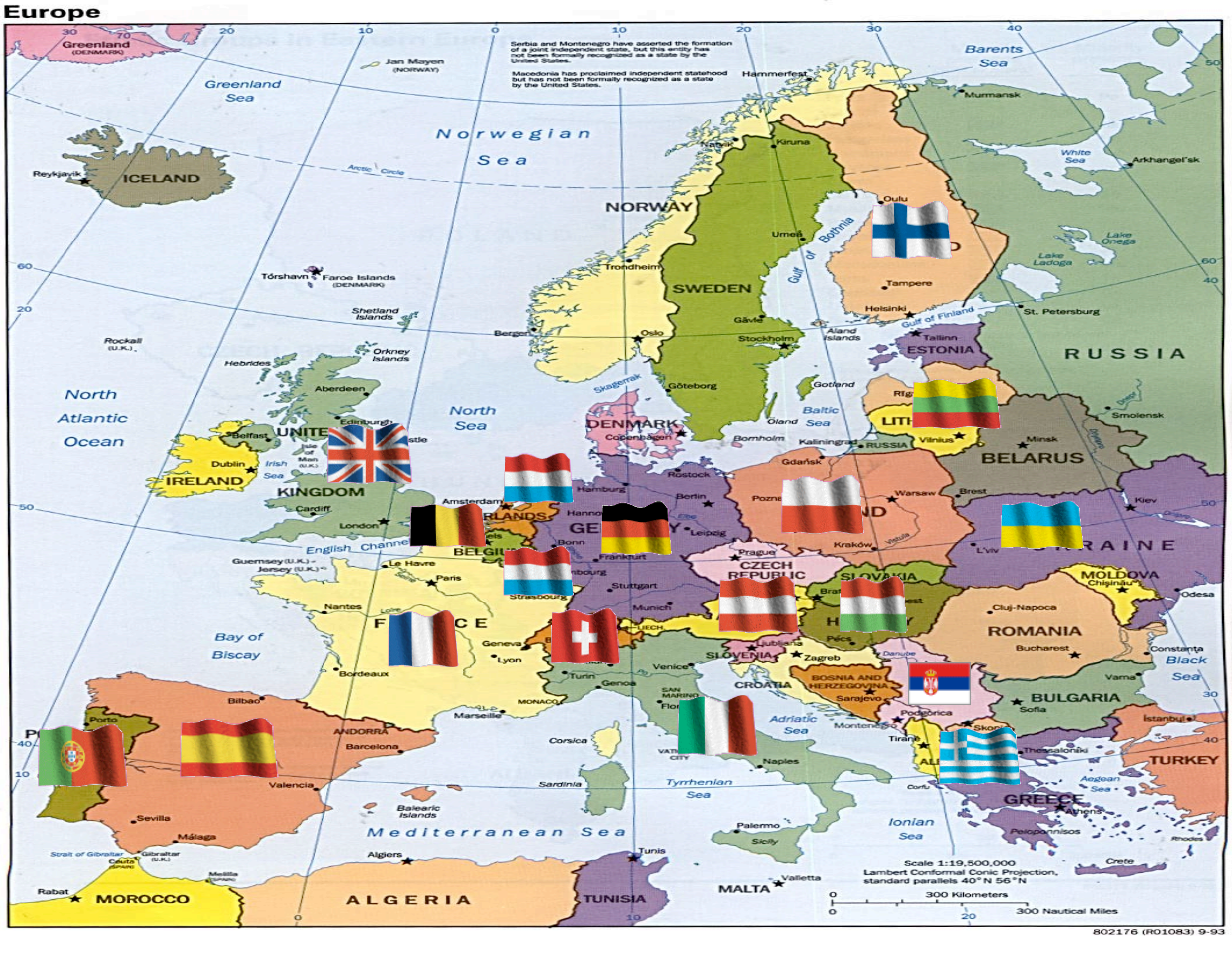


Collaborative approach

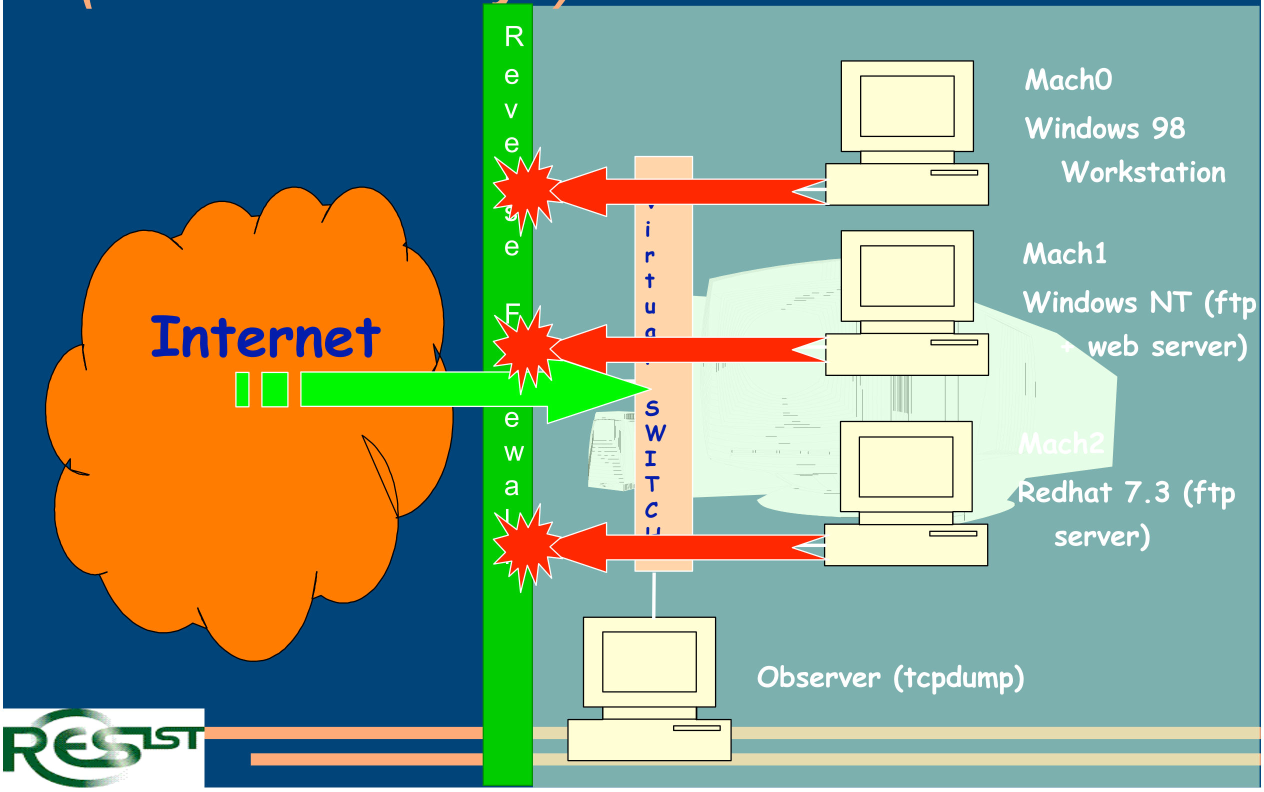
- Leurré.com framework used as a common umbrella to carry out joint research in this thema.
- Some partners bring also on the table the expertise gained with their own proprietary dataset (eg. IBM with its internal Billy Goat project).

50 partners in 30 countries covering the 5 continents





Experimental Set Up
(based on honeyd)



Reverse Firewall

Internet

virtual SWITCH

Mach0
Windows 98
Workstation

Mach1
Windows NT (ftp + web server)

Mach2
Redhat 7.3 (ftp server)

Observer (tcpdump)

Win-Win Partnership

- The interested partner provides …
  - One old PC (pentiumII, 128M RAM, 233 MHz…),
  - 4 routable IP addresses,
- The project offers …
  - Installation CD Rom
  - Remote logs collection and integrity check.
  - Access to the whole SQL database by means of a secure GUI and a wiki (over https).

D12 - Appendices

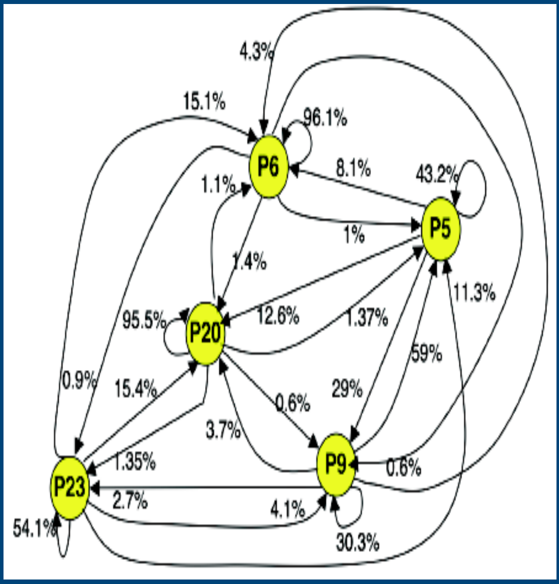
- [Alata et al. 2006] E. Alata, V. Nicomette, M. Kaaniche and M. Dacier, “Lessons learned from the deployment of a high-interaction honeypot”, Proc. Sixth European Dependable Computing Conference (EDCC-6), Coimbra, Portugal, October 18-20, 2006

- [Kaâniche et al. 2006] M. Kaâniche, E. Alata, V. Nicomette, Y.Deswarte, M. Dacier, “Empirical analysis and statistical modelling of attack processes based on honeypots”, Proc. of WEEDS 2006 - workshop on empirical evaluation of dependability and security, Philadelphia (USA), June 25 - 28, 2006.

[Alata et al. 2006]

- High interaction honeypots are not that rapidly detected.
- They help in identifying groups of attackers and their strategies.
- They are complementary to low interaction ones
- Very difficult to use to collect long term datasets.

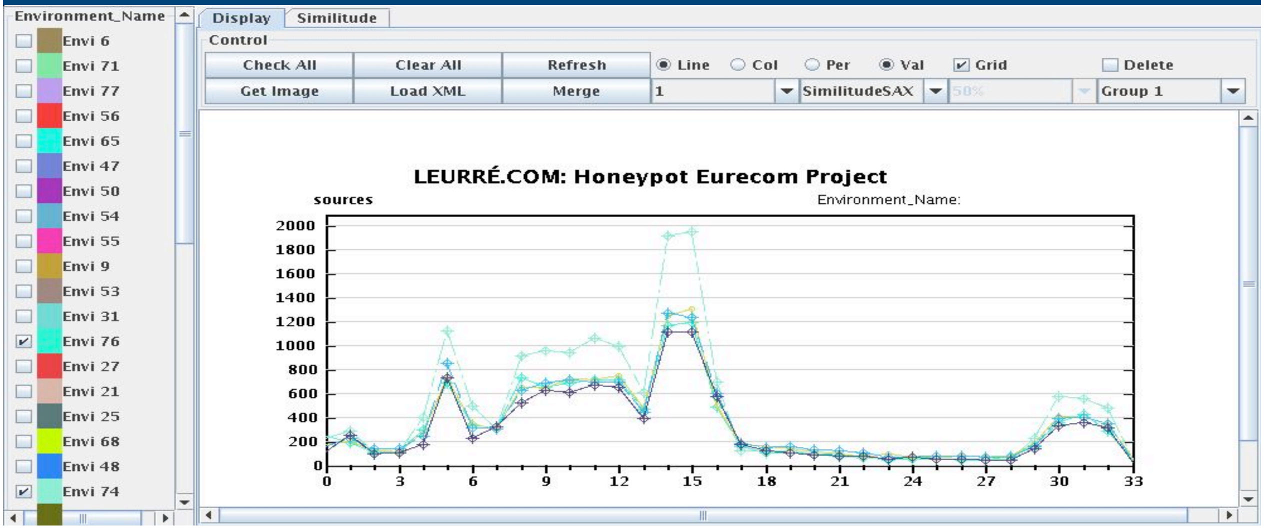
[Kaâniche et al. 2006]



- Propagation graphs open the way to predictive models for *some attacks*

[Kaâniche et al. 2006]

- Patterns of attacks common to several platforms open the way to predictive models for *some platforms* (20/12/06 - 31/1/07)



Environment_Name

□ Envi 6
□ Envi 71
□ Envi 77
□ Envi 56
□ Envi 65
□ Envi 47
□ Envi 50
□ Envi 54
□ Envi 55
□ Envi 9
□ Envi 53
□ Envi 31
☑ Envi 76
□ Envi 27
□ Envi 21
□ Envi 25
□ Envi 68
□ Envi 48
☑ Envi 74

LEURRÉ.COM: Honeypot Eurecom Project

Overview

- Introduction
- *State of Knowledge*
- *Contributions of ReSIST Partners*
- *Conclusions*

Conclusions

- First results demonstrate the usefulness of such datasets with respect to the proposed objectives.
- Honeypots with higher degree of interaction would be welcome.
- Models must be formalized and validated.