WOMBAT: a Worldwide Observatory of Malicious Behaviors and Attack Threats

Corrado Leita

Symantec Research Labs

A Worldwide Observatory of Malicious Behaviors and Attack Threats

EURECOM
Sophia Antipolis

Hispasec Sistemas
Seguridad y Tecnologías de la Información

NASK

symantec.

A*STAR
Institute for Infocomm Research

orange™

WOMBAT

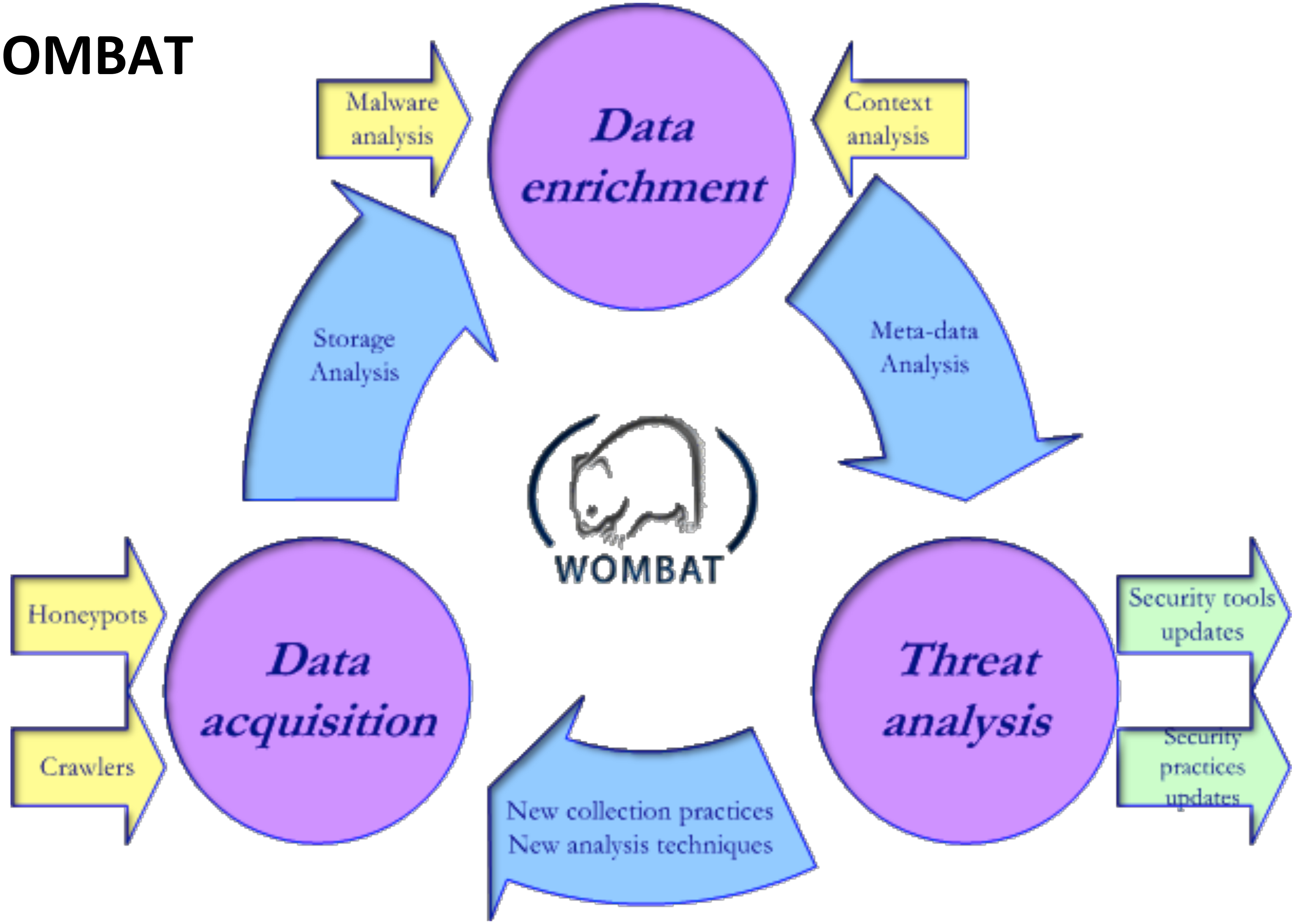
TU VIENNA
Technische Universität Wien
Vienna University of Technology

Politecnico Milano

FORTH
Foundation for Research & Technology - Hellas

vrije Universiteit

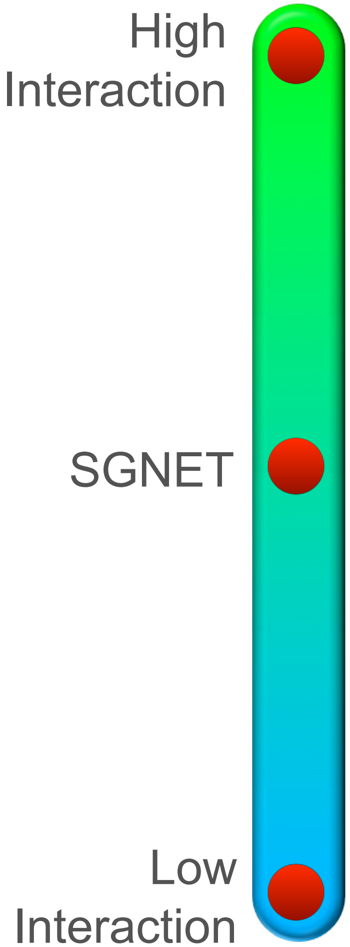
WOMBAT


Malware analysis
Data enrichment
Context analysis
Storage Analysis
Meta-data Analysis
WOMBAT
Honeypots
Crawlers
Data acquisition
Threat analysis
Security tools updates
Security practices updates
New collection practices
New analysis techniques

WOMBAT datasets

- Data sources
  - SGNET *(Symantec)*: server side threats
  - HARMUR *(Symantec)*: client side threats
  - HoneySpider *(NASK/CERT Polska)*: hybrid crawler
  - Shelia *(VU Amsterdam)*: memory tainting-based analysis of exploits against Internet Explorer
  - NoAh *(FORTH)*: honeypot deployment
  - Bluebat *(Politecnico di Milano)*: Bluetooth honeypots
- Data enrichment
  - Anubis Sandbox *(iseclab.org)*: malware behavioral analysis
  - VirusTotal *(Hispasec Sistemas)*: malware detection rates/static analysis
- Aggregators
  - www.maliciousnetworks.org : AS reputation based on the output of the different datasets

Honeypots and protocol emulation: the problem

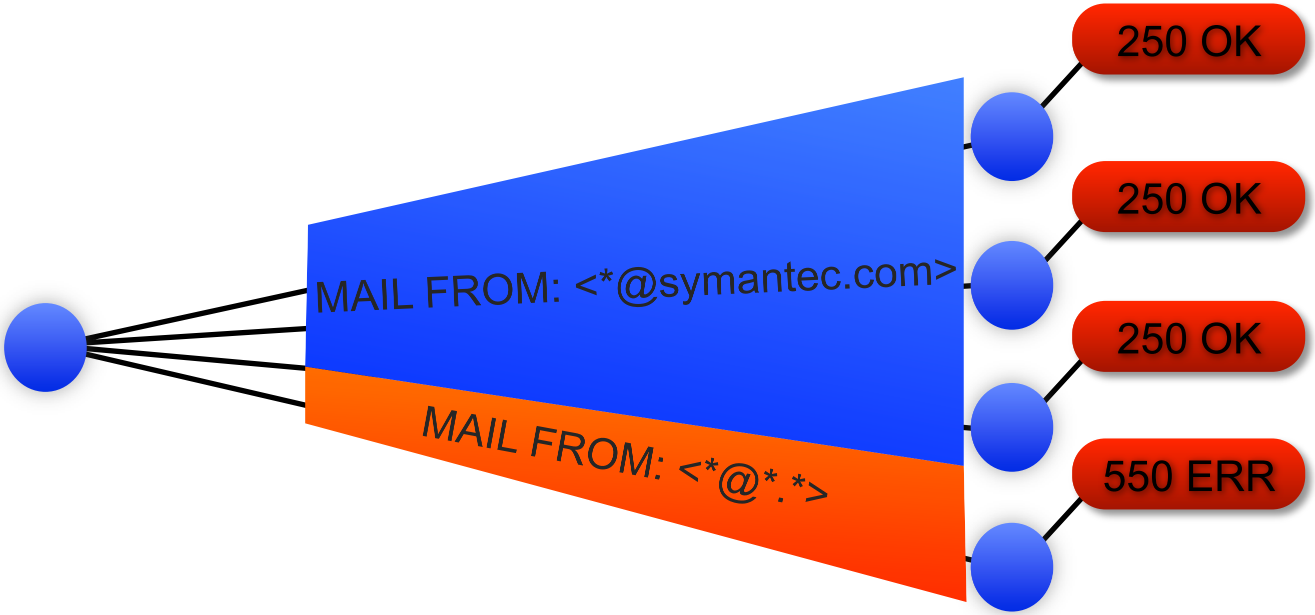


- Need to increase level of interaction
  - Required to retrieve information on the root cause of the observed activities

- Need to minimize the cost of the sensors
  - Implicit requirement of a distributed deployment of sensors hosted by volunteering partners

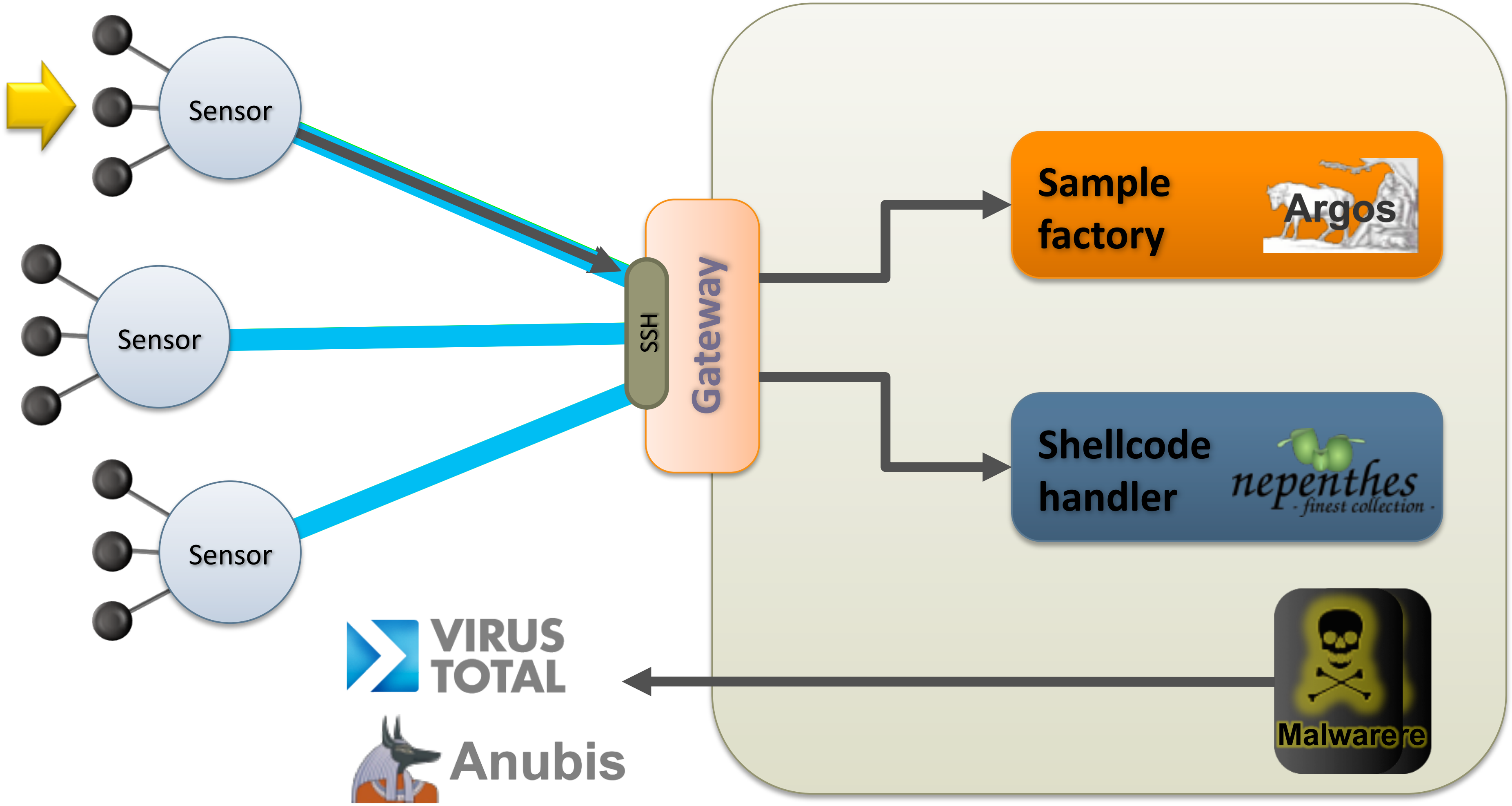
ScriptGen

- Protocol-agnostic algorithm
- Observe conversation samples between a client and a real server
- Infer semantics using bioinformatics algorithms
- Proved good results in handling deterministic exploit scripts

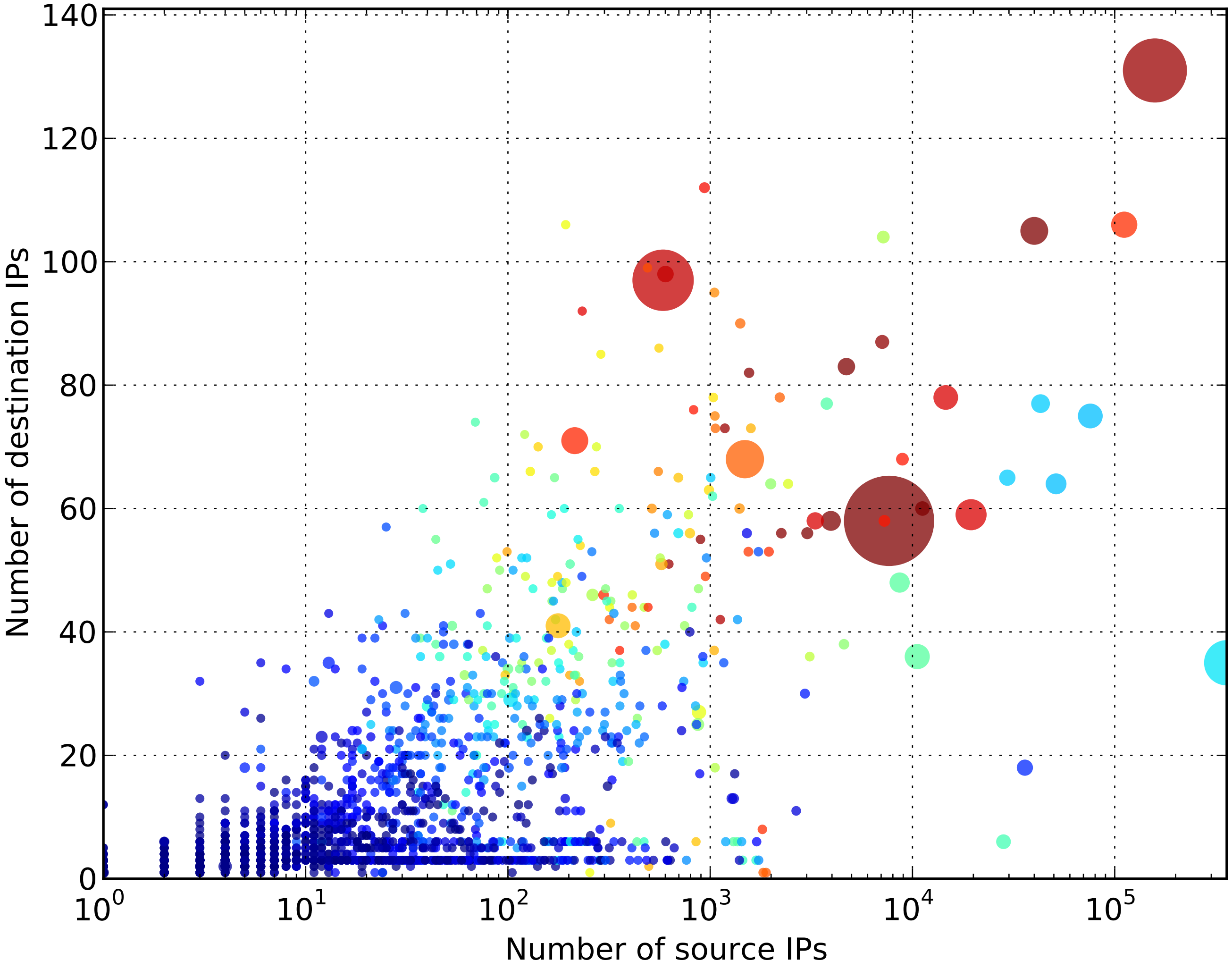


SGNET operation

- Normal operation
- New exploit encountered
- Global update of the FSM knowledge
- Submission of a shellcode sample
- Analyze new malware sample



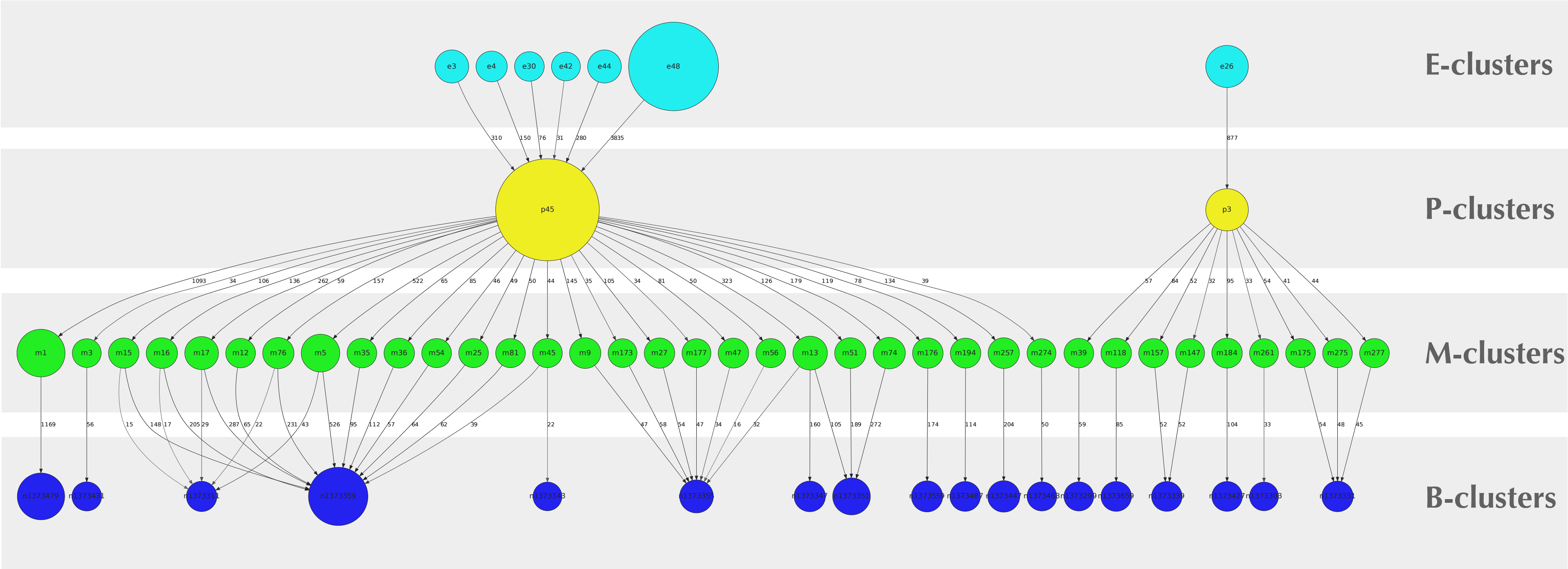
Seeing "things"…



Corrado Leita, Ulrich Bayer, Engin Kirda, “Exploiting diverse observation perspectives to get insights on the malware landscape”, DSN PDS 2010

…and dissecting them

- High visibility activities
  - Few exploits/payload combinations for a large number of M-clusters
  - Same shellcode reused across lots of different variants
  - M-clusters are more than B-clusters



Clusters associated to 50+ injection attacks throughout the observation period

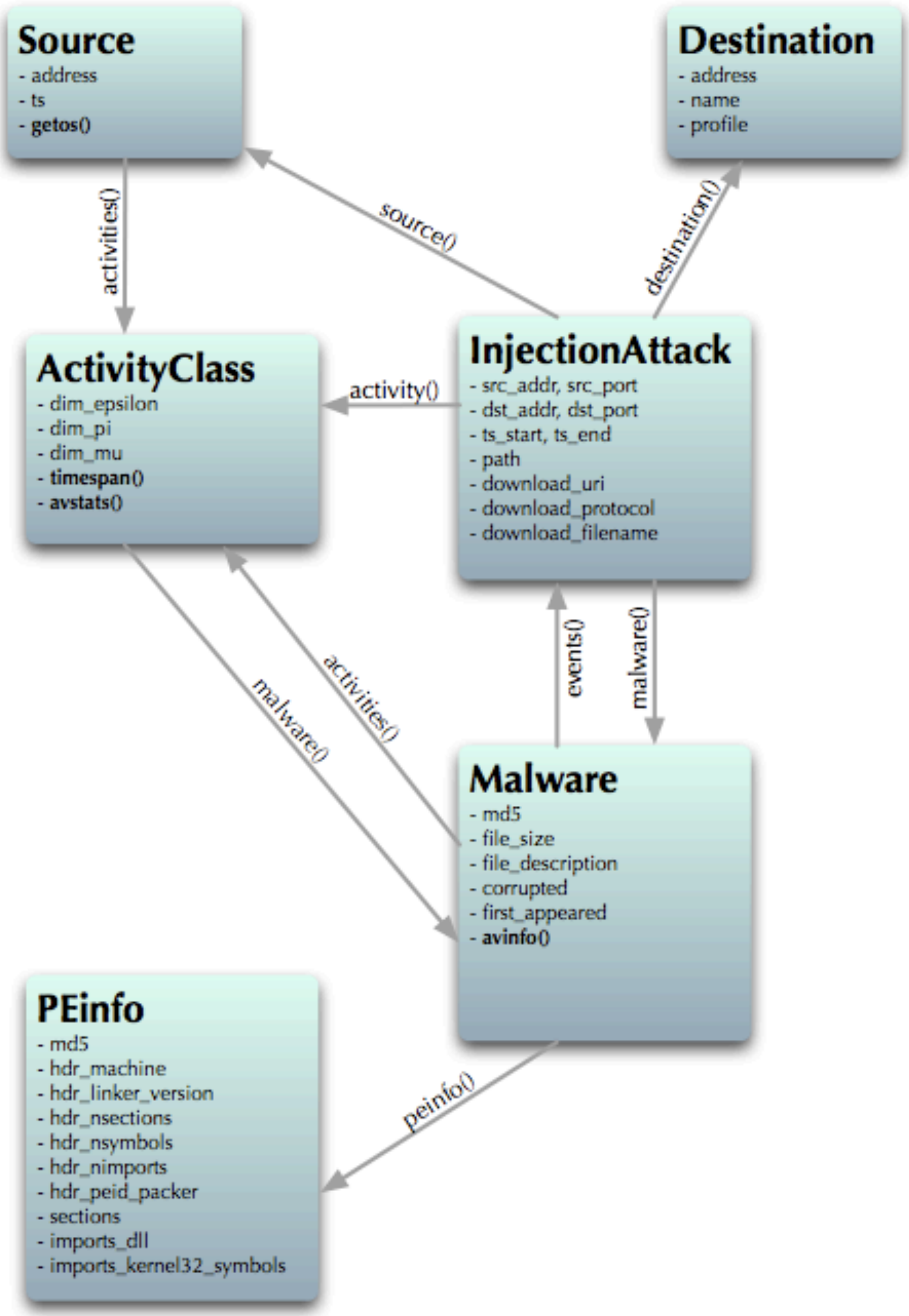
Getting the data: WAPI, the WOMBAT API

- For the **data provider**
  - Control which content to present to the clients, and how
  - Provide the most optimized way
  - Enrich or modify the dataset without needing to modify all the clients

- For the **user**
  - Need for a common "language" to request data from the datasets
    - HTTP submission + XML reports for ANUBIS
    - Email submission + email reports for VirusTotal
    - …
  - Need for programming primitives to easily retrieve information on the fly while performing analysis tasks

WAPI datasets

- Dataset exporable by means of:
  - **Objects:** a specific instance of a given class, e.g. a certain Malware MD5 or a Source IP
  - **Attributes:** basic information on each object, e.g. Malware.file_size
  - **Methods:** more expensive computations on each object, e.g. Malware.avinfo
  - **References:** pointers to related objects
- Reflective API: the client asks at runtime through SOAP primitives what a dataset is currently implementing



Source
- address
- ts
- **getos()**

Destination
- address
- name
- profile

ActivityClass
- dim_epsilon
- dim_pi
- dim_mu
- **timespan()**
- **avstats()**

InjectionAttack
- src_addr, src_port
- dst_addr, dst_port
- ts_start, ts_end
- path
- download_uri
- download_protocol
- download_filename

Malware
- md5
- file_size
- file_description
- corrupted
- first_appeared
- **avinfo()**

PEinfo
- md5
- hdr_machine
- hdr_linker_version
- hdr_nsections
- hdr_nsymbols
- hdr_nimports
- hdr_peid_packer
- sections
- imports_dll
- imports_kernel32_symbols

More info

- WAPI interface implemented on most WOMBAT datasets
  - SGNET, HARMUR, Anubis, VirusTotal, Shelia, …
  - Access managed by dataset maintainers through SSL certificates

- WAPI demo on vimeo:
  - http://www.vimeo.com/15734828 (password: wombat)

- Everybody can join SGNET at low cost
  - Contribute by installing a honeypot (4 routable IPs, old PC is enough)
  - Sign NDA (protect participants identity as well as observed IP addresses)

Thank you!

Corrado Leita

corrado_leita@symantec.com

This document is provided for informational purposes only and is not intended as advertising. All warranties relating to the information in this document, either express or implied, are disclaimed to the maximum extent allowed by law. The information in this document is subject to change without notice.