WOMBAT: a Worldwide Observatory of Malicious Behaviors and Attack Threats

Corrado Leita
Symantec Research Labs
A Worldwide Observatory of Malicious Behaviors and Attack Threats
WOMBAT

Data enrichment

Data acquisition

Threat analysis

Malware analysis

Context analysis

Storage Analysis

Meta-data Analysis

Honeypots

Crawlers

New collection practices
New analysis techniques

Security tools updates

Security practices updates

WOMBAT
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 (*Hispacec Sistemas*): malware detection rates/static analysis

• Aggregators
  – [www.maliciousnetworks.org](http://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”...
...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
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:

• 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