Politecnico di Milano
Dip. Elettronica e Informazione
Milano, Italy

WOMBAT: Building a Worldwide Observatory of Malicious Behavior and Attacks Threat

Stefano Zanero, Ph.D.
Assistant Professor

SecureIT, Los Angeles, 5/03/2009

Knowledge: granting success, since ~500 b.C.

- Knowing your enemy is the key to success
  - *"He will win who knows when to fight and when not to fight... He will win who, prepared himself, waits to take the enemy unprepared. Hence the saying: If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle."* [Sun-Tsu]
- Perhaps the most often quoted, and less often practiced, sentence in history
- Understanding is the key to (re)acting *sensibly*, and we are failing in a lot of fields, notably anti-terrorism controls in the airports

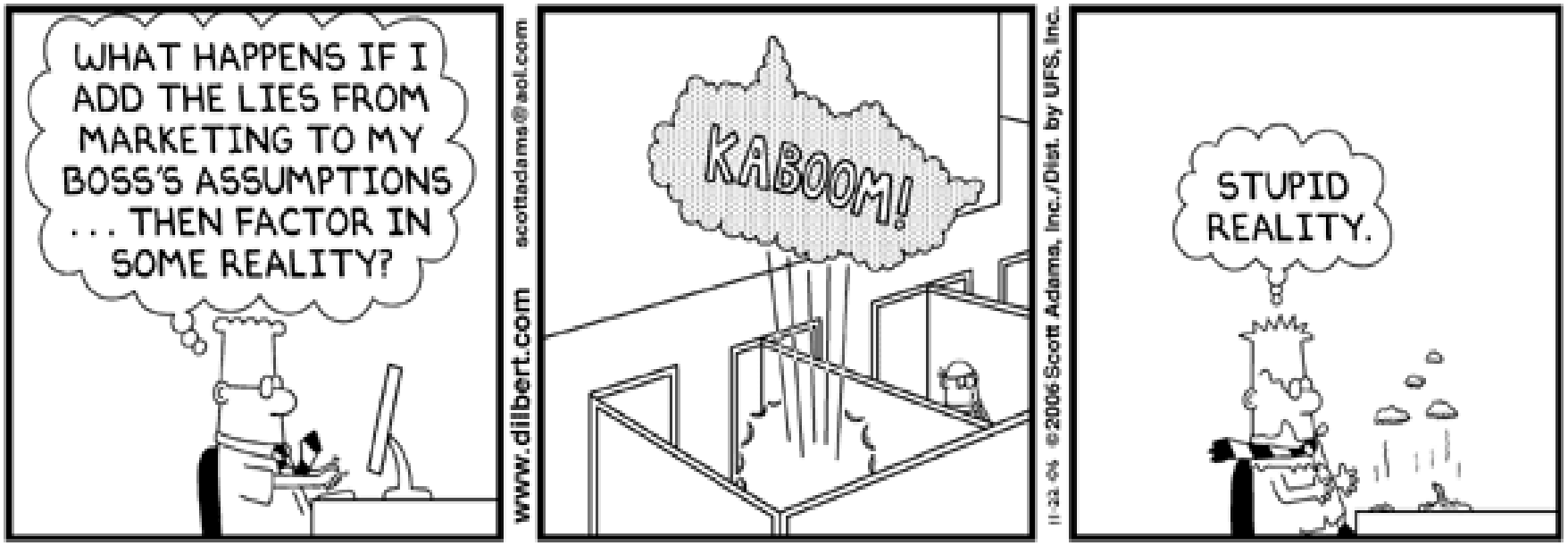
True or false, or somewhere in between ?

- "Asymmetric warfare potential of cyberspace will lead to an increase in electronic warfare and cyberterrorism". True or False ?
  - Repeated countless times, since 9/11/01 (at least)
  - "If we ever manage to get real-world terrorists to blow up computers instead of airplanes, it will be at our advantage, as computers have backups and humans don't" (R. Power, CSI)
  - No public data which confirms or disconfirms cyberterrorism activities, also because there's no or little distinctive features of cyberterrorism from common cyberattacks
  - Some doubts were recently cast even on the Estonian cyberattack, which was seen by many as the first publicly confirmed cyberwarfare event

Lies, Damn Lies, or Very Good Statistics ?

- FBI – CSI report: “croce e delizia”
- There is always a "rising wave of Internet crime"
- Reports of losses usually out of thin air
- Reports based on respondent's honesty and knowledge (“I have no intrusion detection process”, so how do you know?)
- Q: Why reported incident losses fall every year ?
- A: Because the numbers are not statistically solid
- From the CSI Alert Newsletter (quoted by A. Chuvakin)
- 5,000 members of CSI surveyed (they are not a representative set). Response rate 12% (616 of 5000). We do not know any statistics on these 12% and their dissimilarity to the others.

What happens if you add...



- "*** Report: Surge in Viruses and Worms Targeting Mobile Devices, Satellite Communications Anticipated in 2005"
- ... hell-looooooooo ? It's 2009... where are youuuuu ? :)

Tales of a death foretold (way too early)

- Prediction anonymized and mixed up to protect the innocent and clueless analysts out there
- "In July 2001, Code Red spread to $HUGE_INT systems within $SMALL_INT hours; the worldwide economic impact was estimated to be $INSANE_FIGURE billions. SQL Slammer was even faster.
- "We'll see an even greater increase in the speed and destructive capabilities of threats."
  - Warhol Worms, Flash worms, etc
  - Extremely good academic papers, but never incarnated

And by the way… where are the worms ?!

- We *all* thought that the Internet would get wormier
  - Don't try to deny it: I am sure you have **at least** one slide where **you** said that!
- The trend was clear:
  - 2001: Li0n, Code Red, Nimda
  - 2002: Slapper, Klez
  - 2003: SQL Slammer, Blaster, SoBig
  - 2004: Sober, MyDoom, Witty, Sasser
  - I have even an iDefense t-shirt with this list on it!
- Since then, silence on the wires. No new "major" worm outbreaks
  - Weaponizable vulns were there, we even collectively braced for impact a couple of times
  - Did we get *so better* at defending networks? I bet "not"

Rise of the Bots

- Bots, bots everywhere
  - When I was a youngster, bots were IRC warriors' stuff (~1999-2000)
  - We used to call remote control trojans “zombies”, and they were usually DDoS tools (2000-2)
- Today's bots are different
  - Intelligent, evolving, with complex C&C infrastructures, difficult to remove as well
  - Larger botnets (10k common, 1M+ seen)
  - Phishing, spamming and pharming bots... more difficult to track than DDoS events
- How do we track them? How do we analyze them?
  - Worm explosive propagation vs. bot slow and steady diffusion: there's no network telescope that can see them
  - Exemplar case: Conficker/Downadup

Open wormy questions: example

- Why no worm has ever targeted the infrastructure?
  - (possible exception of Witty, targeting firewalls)
- Possible explanation: routers and the like are a difficult vector to exploit
  - Not really true anymore, see FX's and Michael Lynn's works
  - Can use a traditional worm for propagation + a specialized payload for infrastructure damage
  - Windows of opportunity were there:
    - June 2003: MS03-026, RPC-DCOM Vulnerability (Blaster) + Cisco IOS Interface Blocked by IPv4 Packets
    - April 2004: MS04-011, LSASS Vulnerability (Sasser) + TCP Vulnerabilities in Multiple IOS-Based Cisco Products (resets)
- So why, oh why, the /bin/ladens of the world were not there, grinning and reaping?

He who knows not the enemy, nor himself

- Summary of the worm rise and fall:
  - Most folks and consultants were clueless about worms in 2000 (lost preparing for the 2-digits-years cataclism)
  - Since 2004 lots of money and consultant-speak in the direction of fighting "the dreadful and impending Big One of the flash worms"
  - The era of the worms was actually almost over already
- The result
  - Not the disappearance of worms
  - Nor an improved resilience to them (infrastructure is just as exposed to a flash worm today as it was in 2004)
  - A mass distraction of resources from the real, impending threats (endpoint security and prevention of client-side attacks and botnets)
- "...every battle is a certain risk"

Observing attacks != Knowing attackers

- Various questions about the attackers
  - Attribution (tipically for law enforcement)
  - Characterization aka profiling
- Usually observation of attacks is not enough to answer such questions
  - In particular, characterization of attackers is still in its infancy
  - There are also various hacker profiling projects, but in most cases they are linked either to criminal case review or to dissemination of questionaires
  - The efficacy is highly debatable, to be honest
- Example of a good methodology: the work by T.J. Holt and Max Kilger, UNCC

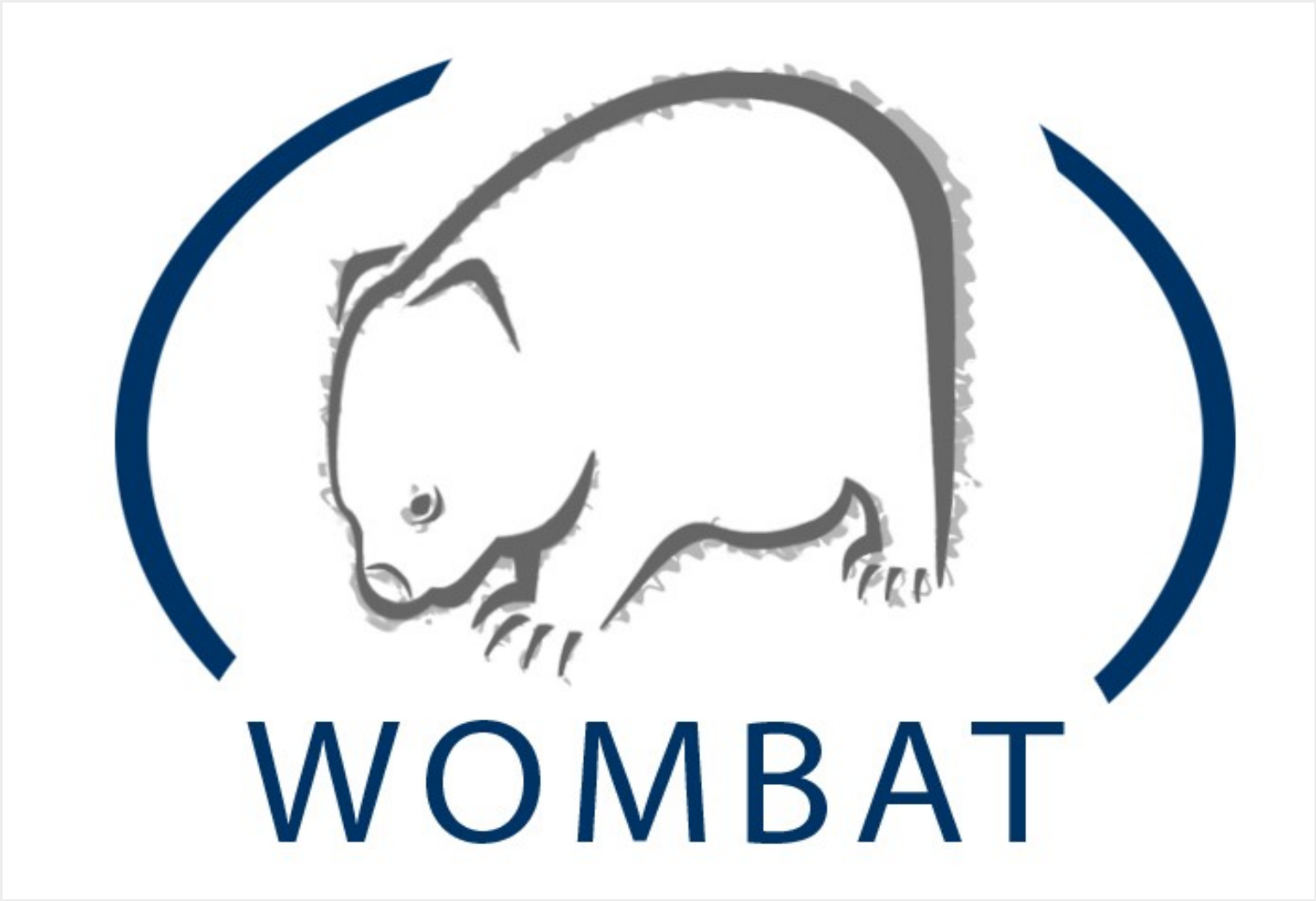
The need is felt also at political levels

- EU Commissioner Vivianne Reding stressed how difficult it is for decision-makers to create appropriate policies for fighting cybercrime without reliable data, models and theories on the root causes and the underlying generative processes of the tidal wave
- Testimonies in front of the U.S. House Committee on Homeland Security: better sharing and analysis mechanisms needed
- DHS investments in Information Sharing & Analysis Centers (ISACs)
- National Strategy to Secure Cyberspace (NSSC) has 3 out of 8 action items related to log sharing

Today's observation points

- Efforts by vendors
  - ATLAS (Arbor)
  - DeepSight (Symantec, formerly SecurityFocus)
- Community and no-profit efforts
  - Dshield and the Internet Storm Center (SANS)
  - Network Telescope
  - The HoneyNet project
  - NoAH and Leurrecom projects

Worldwide Observatory of Malicious Behavior and Attack Tools



Basic facts on WOMBAT

- A project funded by the EU (and partner countries) and several non-EU partner institutions in the 7th Framework Programme
- Started at the beginning of 2008
- Total cost: 4,422,746 €
- EC Contribution: 2,890,796 €
- All of the project's contributions are available at:

- www.wombat-project.eu

The WOMBAT Consortium

EURECOM
Sophia Antipolis

orange™

symantec™

TU WIEN
Technische Universität Wien
Vienna University of Technology

HISPASEC SISTEMAS

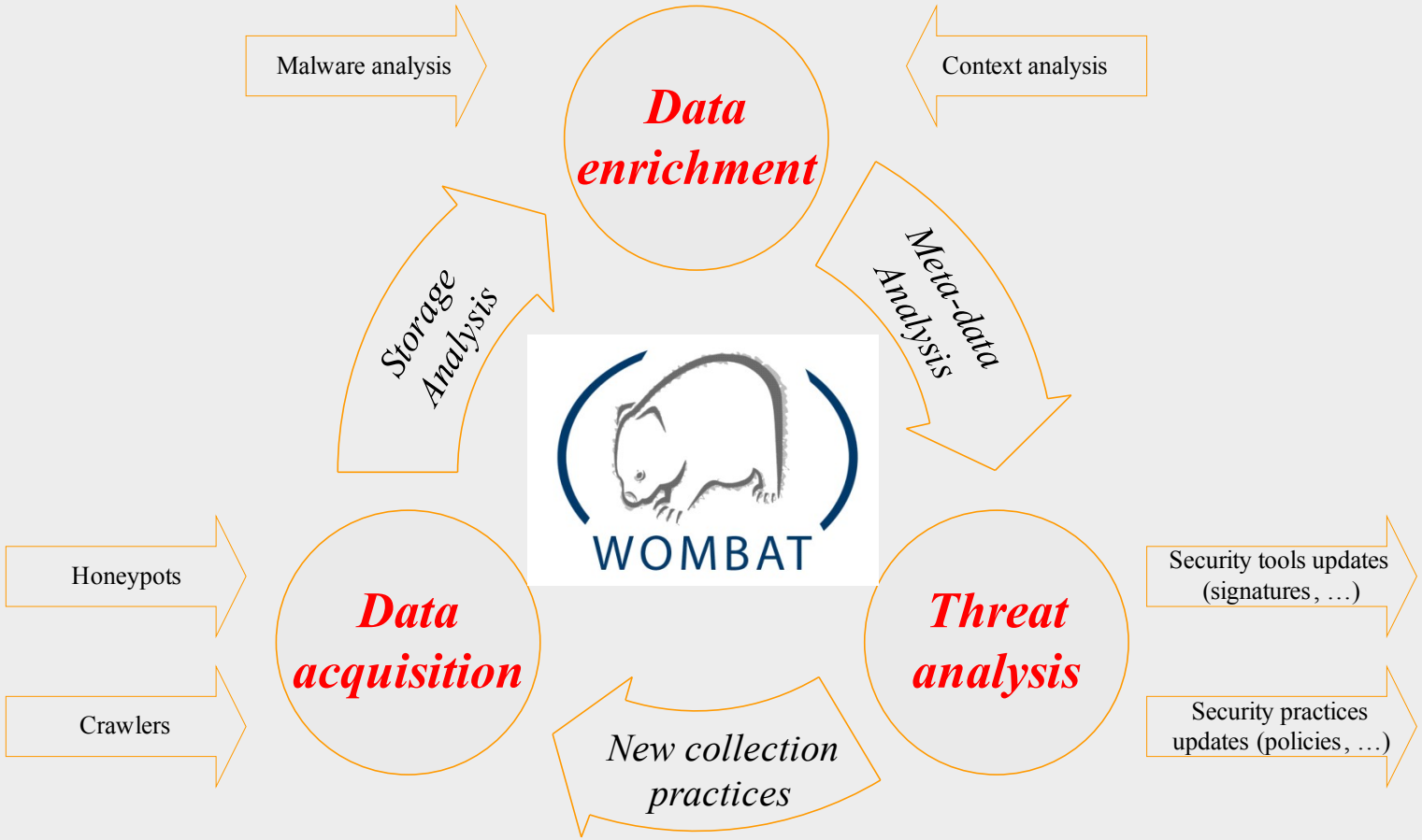
FORTH

POLITECNICO DI MILANO

vrije Universiteit

NASK

Main objectives and principles



Data Acquisition

- Creation of an infrastructure for storage, access and analysis
  - Creation of a set of standard API (WOMBAT API) and an infrastructure for data sharing and integrated query
- We sought and seek international collaboration
  - Invitation-based Workshop on Internet Security Threats Data Collection and Sharing (WISTDCS), held in Amsterdam, 21-22/04/08, Proc. by IEEE Computer Society Press, available on IEEEXplore
  - Already integrated data from SANS ISC, SORBS DNSBL, Abuse.ch FastFlux Tracker, HTTPBL (Project Honeypot), Kyoto University, Spamhaus.org, OffensiveComputing.net and we are integrating SRI Malware Threat Center and Honeynet Project data
  - Get in touch if interested to participate

Also, some new and improved sensors…

- HARMUR
  - Historical Archive of Malicious URLs
- Shelia
  - Client-side Windows-based IDS
- Paranoid Android
  - Smartphone sensor
- HoneySpider Network
  - Client-side honeypot (focuses on drive-by download attacks)
- BlueBat
  - Bluetooth honeypots
- Wireless honeypots

Leurré.com/SGNet

- We extended Leurré.com (www.leurrecom.org), an existing project operated by Eurécom into SGNet, a scalable network of LIH, MIH and HIH
- Broad network of honeypots covering more than 30 countries
- V1.0
  - Architecture of distributed **low-interaction** honeypots
  - All traces captured on each platform are uploaded on a daily basis into a centralized relational database
- V2.0 – SGNet (WOMBAT component)
  - Not just following known attack, get to the point of malware deployment and STILL avoid using HIH
  - SGNET = Scriptgen (Eurecom) + Argos (VU Amsterdam) + Nepenthes (TU Manheim/MWCollect All.) + Anubis (TU Wien) + Virustotal (Hispasec)

ScriptGen

- Autogenerate scripts that emulate a service
  - Impossible, a reverse engineer's wet dream :)
- Autogenerate scripts that emulate the answers of a service to a deterministic script (the exploit)
  - Far simpler
- Three steps approach
  - A real machine answers are recorder
  - If the machine gets compromised, usual cleanup
  - Messages are analyzed and a state machine in python is derived, representing requests and replies
    - Using bioinformatics techniques from http://www.insidiae.com/PI
- Details published at NOMS08, EDDC08

Data Enrichment

- Commonly acquired data have proven not to be sufficient to reveal root cause(s)
  - Collecting thousands of malware: easy
  - Identify and classify them automagically: more difficult
  - Figuring out who's developed them and why: priceless
- Examples of the types of analysis we are integrating:
  - code behavior characterization
  - Malware clustering based on behavior
  - structure of the malicious code and philogeny
  - attack contextual information (how it was performed; scanning activities; type of deployed payload; subsequent actions)

Threat analysis

- Final goal:
  - Find out the root causes of the observed attacks
  - Build upon this acquired knowledge in order to better predict upcoming threats.
- Tools
  - Data and metadata correlation (very different from correlating alerts for intrusion detection purposes)
  - Statistical analysis
- Expected results:
  - Strategic (as opposed to tactical) early warning capabilities
  - Security investments and policy making decisions support

Conclusions & Future Work

- Conclusions:
  - We need to be able to observe, understand and infer
  - Before WOMBAT we were partially able to observe, to understand (but generally late), and not to infer
  - We are improving collection practices (a little bit), data analysis and enrichment (a lot), in order to devise automatic inference mechanisms for root cause analysis
- WOMBAT:
  - Funded global initiative for studying attacks and threats
  - Trying to make good use of the excellent work that has already been done in this area
  - Aiming to coordinate, rather than compete, with other large initiatives
  - You can join **SGNET** and get access to data already
  - Seeking US-based partners for joint work

Thank you!

Any question?

I would greatly appreciate your feedback !

Stefano Zanero
zanero@elet.polimi.it
http://home.dei.polimi.it/zanero/eng