



Overview

- Introduction
 - Evolution of the threats landscape
 - New threats vs. new challenges
 - WINE
- Conclusions

Symantec in one slide (!)

- Founded in 1982
- Facilities in more than 40 countries, more than 17,500 people in the world.
- Ranked #353 on the Fortune 500 in 2010
- Total revenues of \$6.01 billion in FY10
- 69% enterprise revenue, 31% consumer revenue (FY10Q4)
- 50% of revenues from outside the US: EMEA: 30% - APAC: 15% - Americas (incl. Latin America and Canada): 55%
- More than 120 million active consumer users and 110 million enterprise customers
- Over 40,000 resellers

M. Dacier – HCSS, Annapolis MD, May 5 2011



CARD: Collaborative Advanced Research Dept.

- CARD is part of Symantec Research Labs, led by Joe Pasqua (VP, Research) who reports to Mark Bregman (CTO).
- CARD members are engaged in long term, exploratory, collaborative research projects
- CARD members are located in the USA (Herndon and Culver City) as well as in Europe (Ireland and France).
- All projects, but one (WINE), are partially supported by various external funding agencies.

M. Dacier – HCSS, Annapolis MD, May 5 2011



Collaboration ?

- We participate to research projects only if we see the added value our active collaboration will bring.
- Win-Win:
 - partners bring some complementary skills on the table, in domains where we are interested in potentially growing in the future.
- The output of the project must have some potential for technology transfer, even if not immediately.

M. Dacier – HCSS, Annapolis MD, May 5 2011



5

WINE

- WINE aims at fostering collaboration in a simple, flexible and scalable way.
- WINE aims at improving rigorous and sound experimental research within the computer security domain (to start with).
- WINE aims at providing academic researchers with an environment suitable to validate experimentally their research on the ever changing threat landscape.

M. Dacier – HCSS, Annapolis MD, May 5 2011



6

Overview

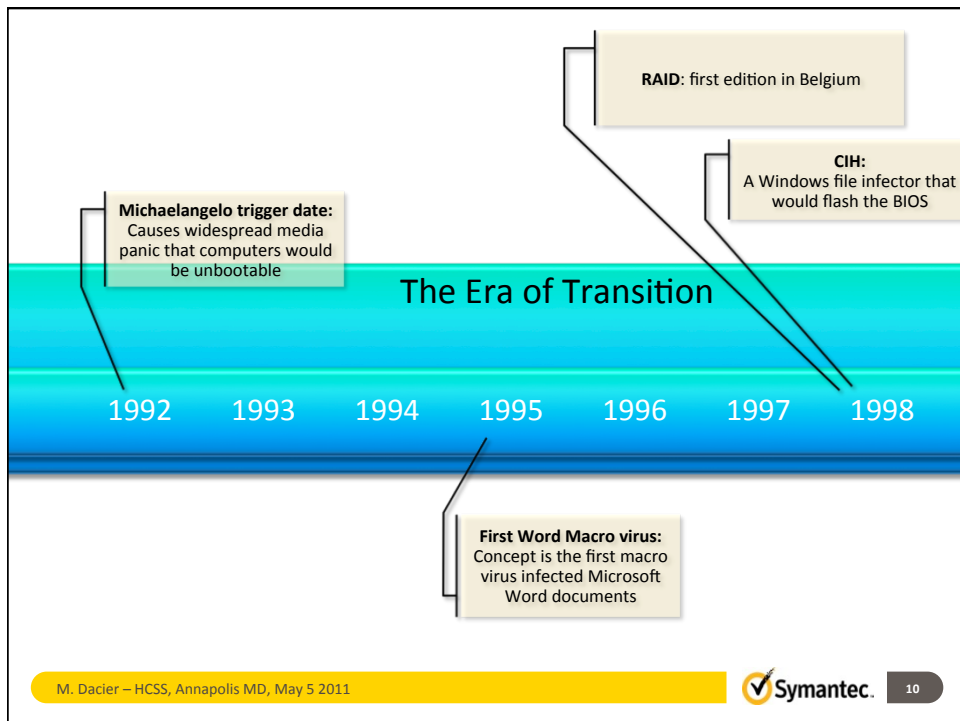
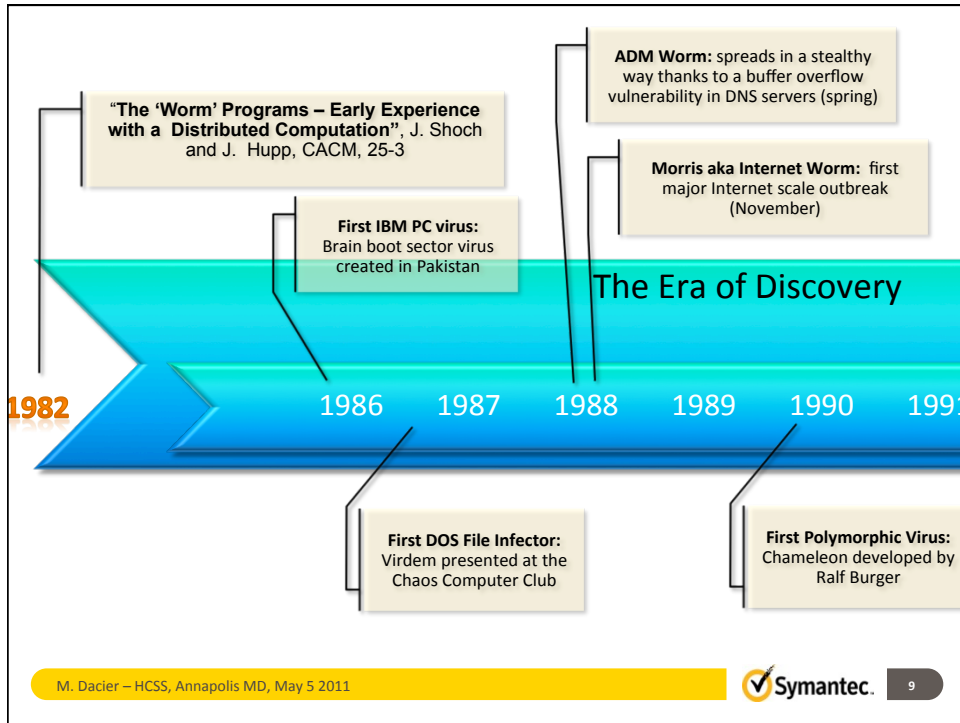
- Introduction
 - Evolution of the threats landscape
 - New threats vs. new challenges
 - WINE
- Conclusions

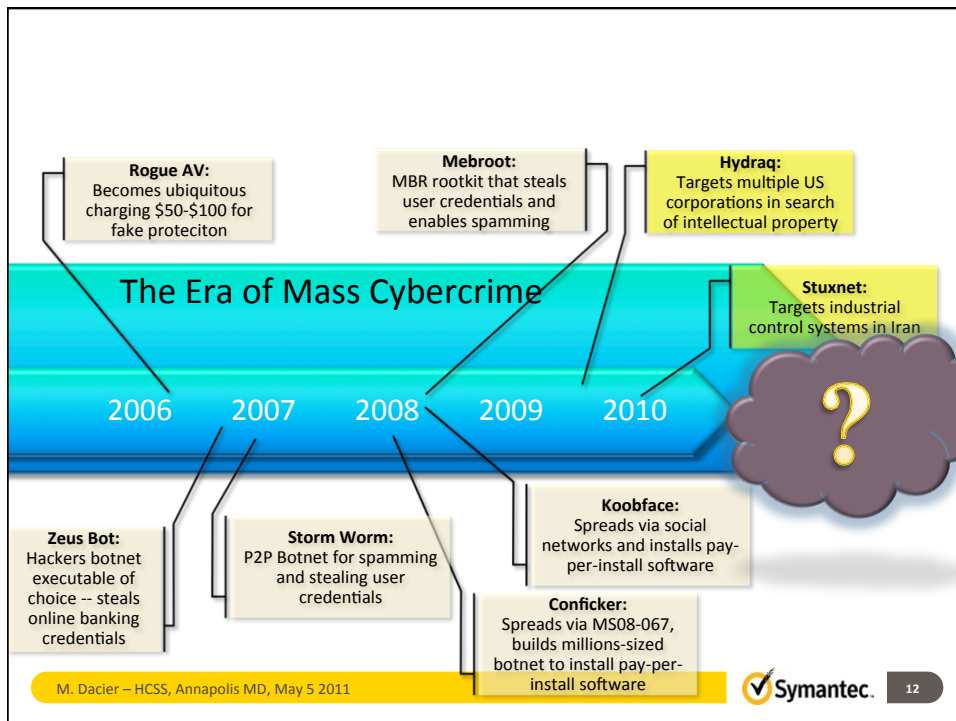
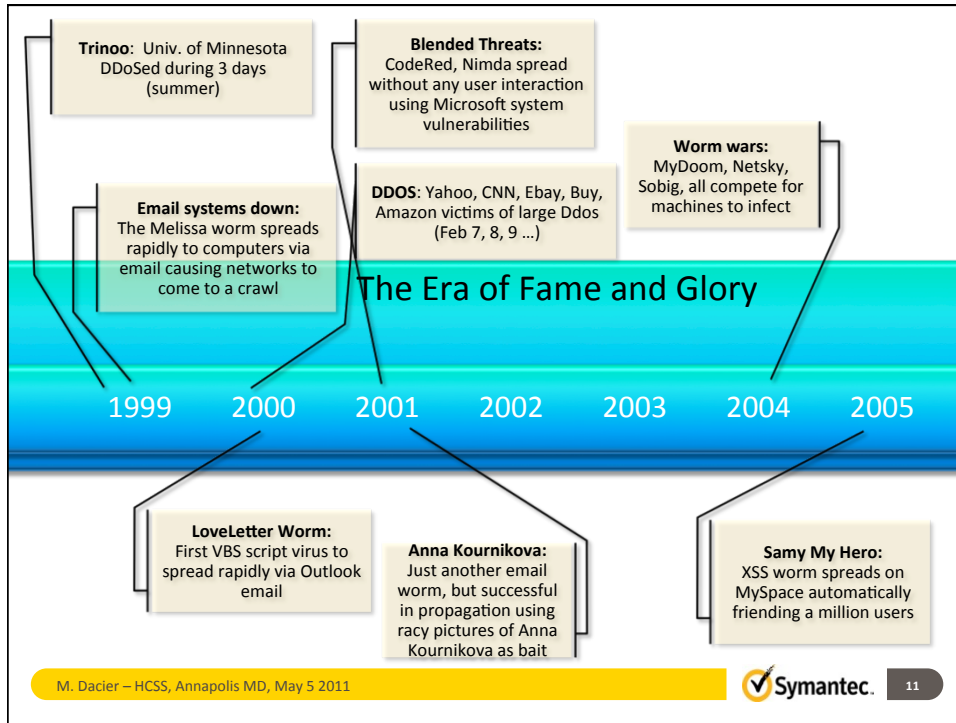
M. Dacier – HCSS, Annapolis MD, May 5 2011



M. Dacier – HCSS, Annapolis MD, May 5 2011







Overview

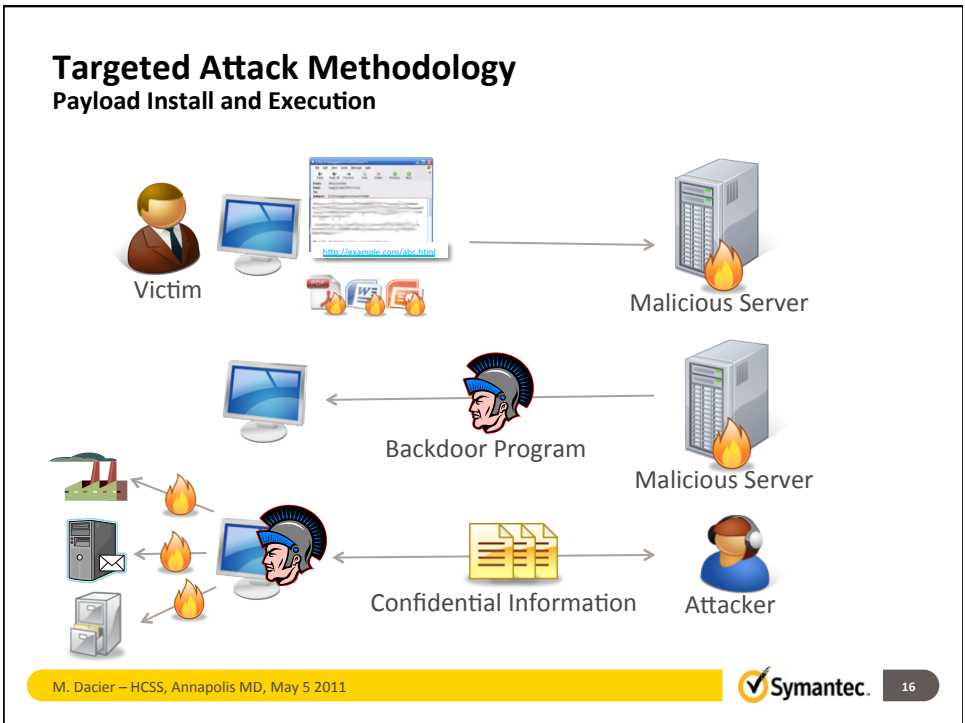
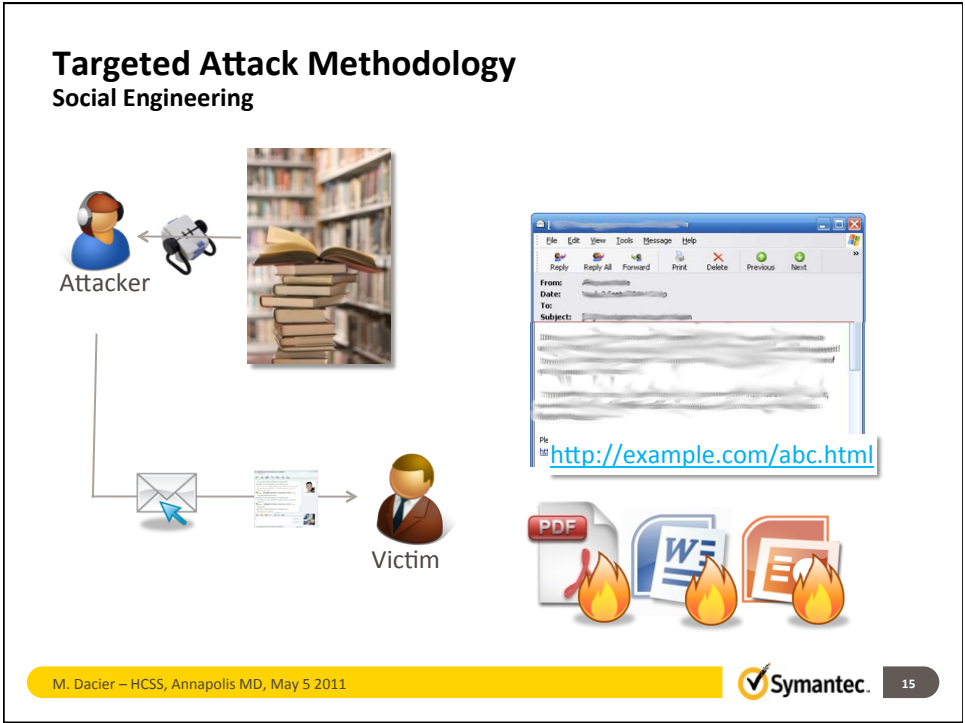
- Introduction
 - Evolution of the threats landscape
 - New threats vs. new challenges
 - WINE
- Conclusions

M. Dacier – HCSS, Annapolis MD, May 5 2011



M. Dacier – HCSS, Annapolis MD, May 5 2011








Targeted Attacks:
Certainly more sophisticated and prevalent
but not that novel


M. Dacier – HCSS, Annapolis MD, May 5 2011  17

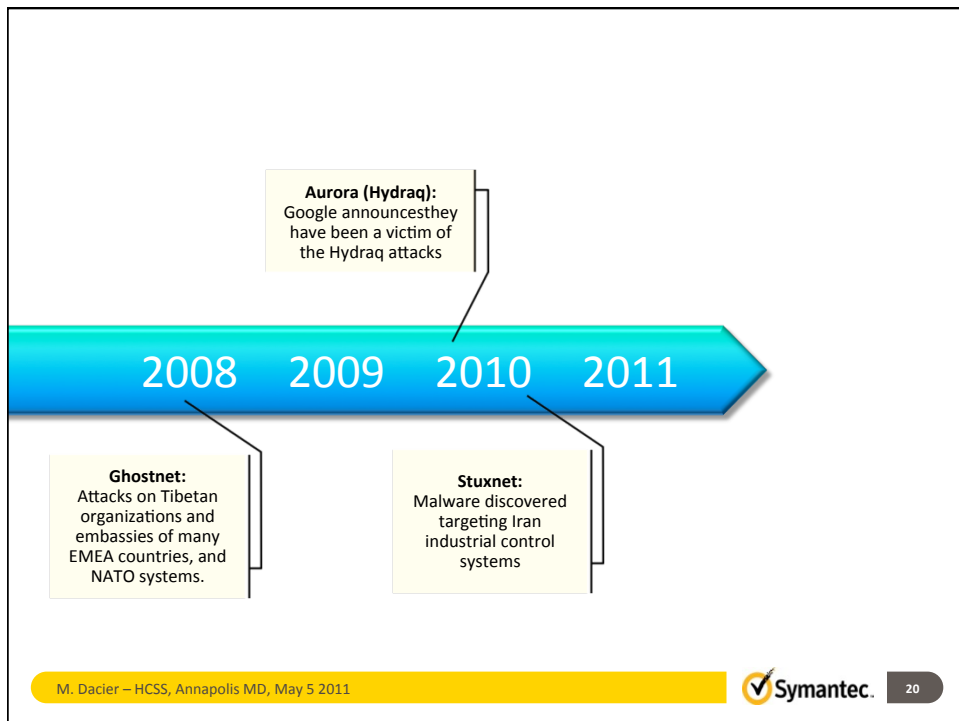
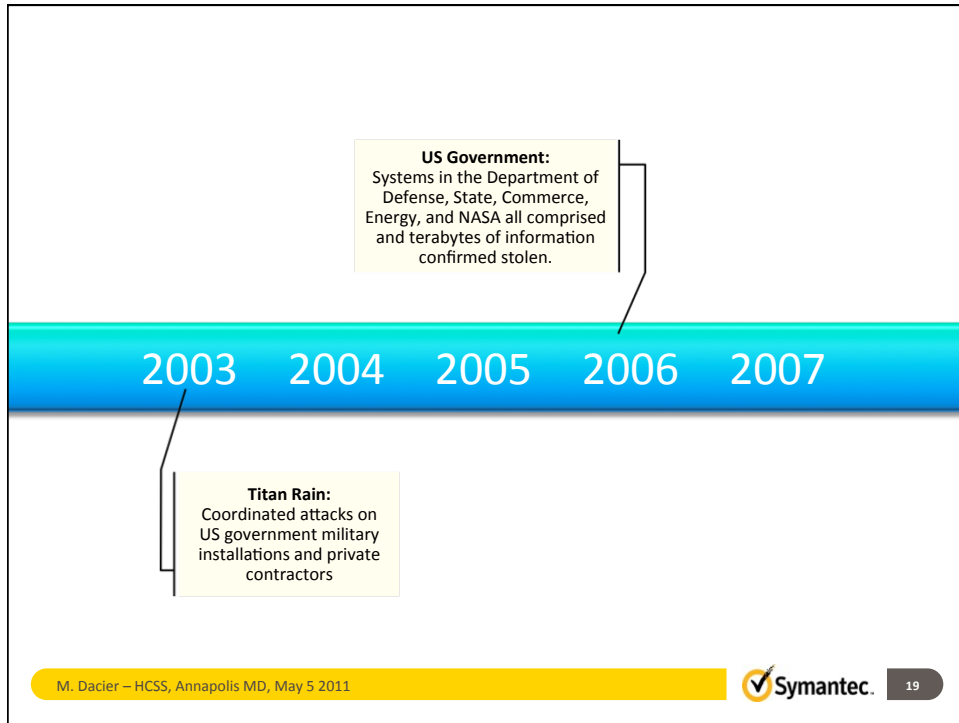


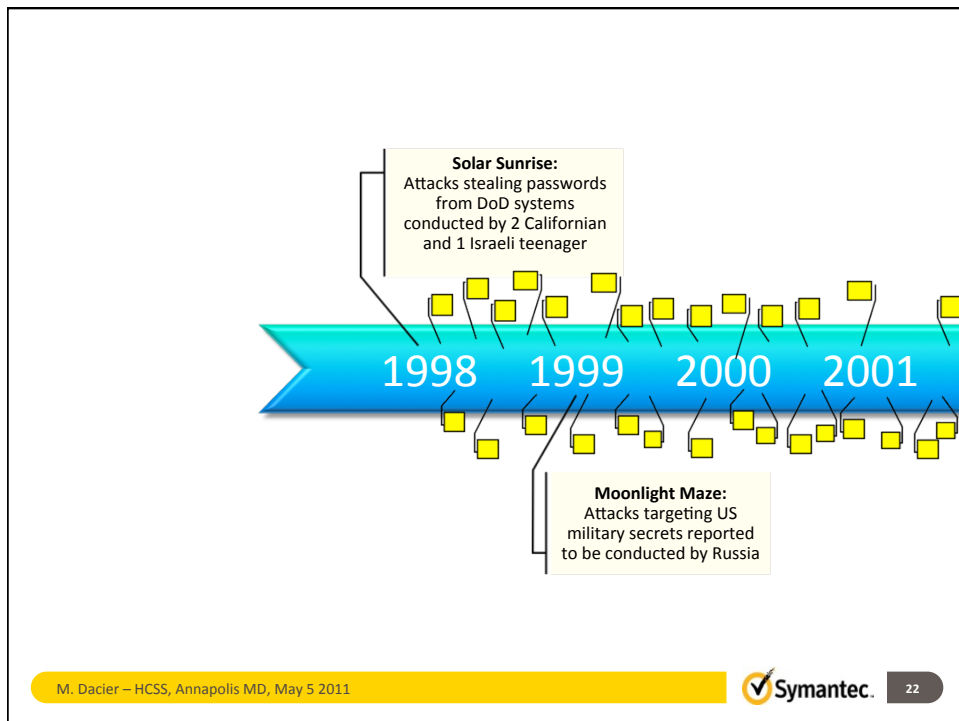
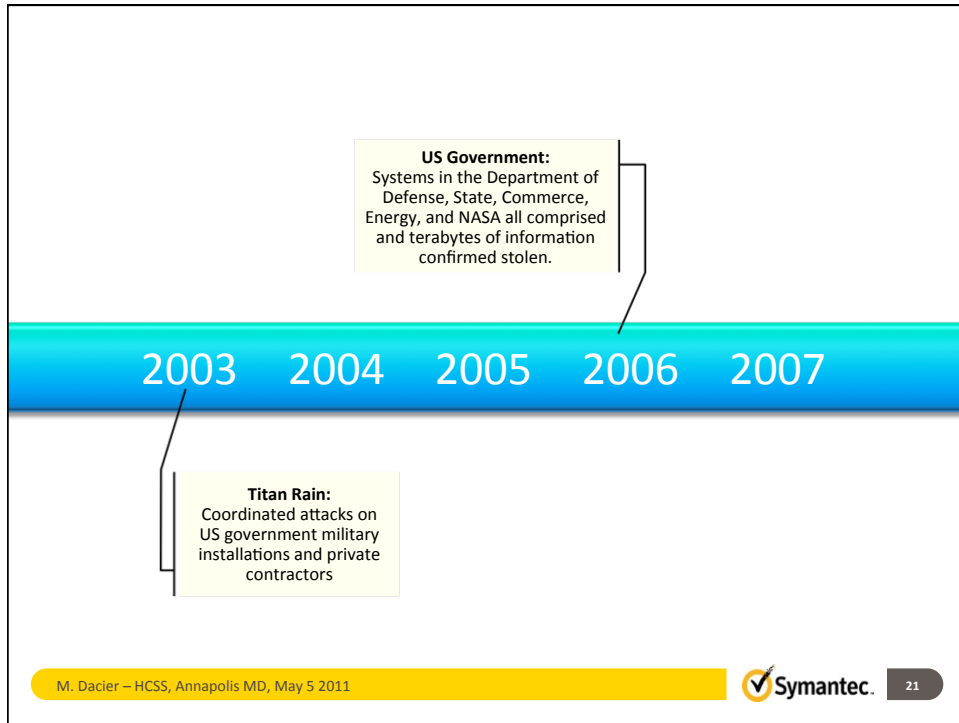
Solar Sunrise:
Attacks stealing passwords
from DoD systems
conducted by 2 Californian
and 1 Israeli teenager

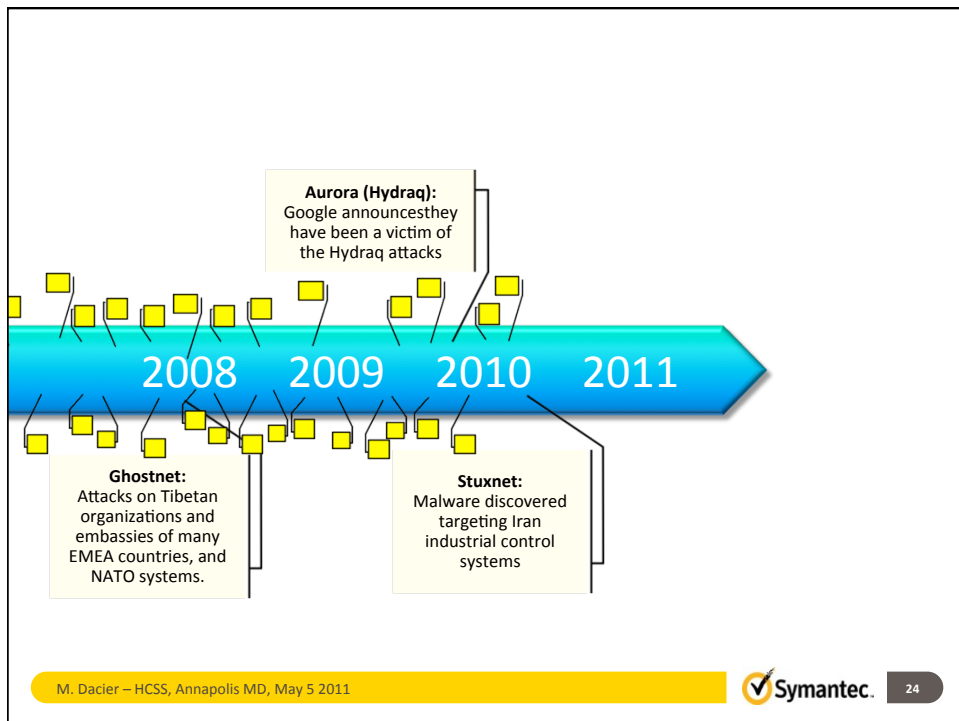
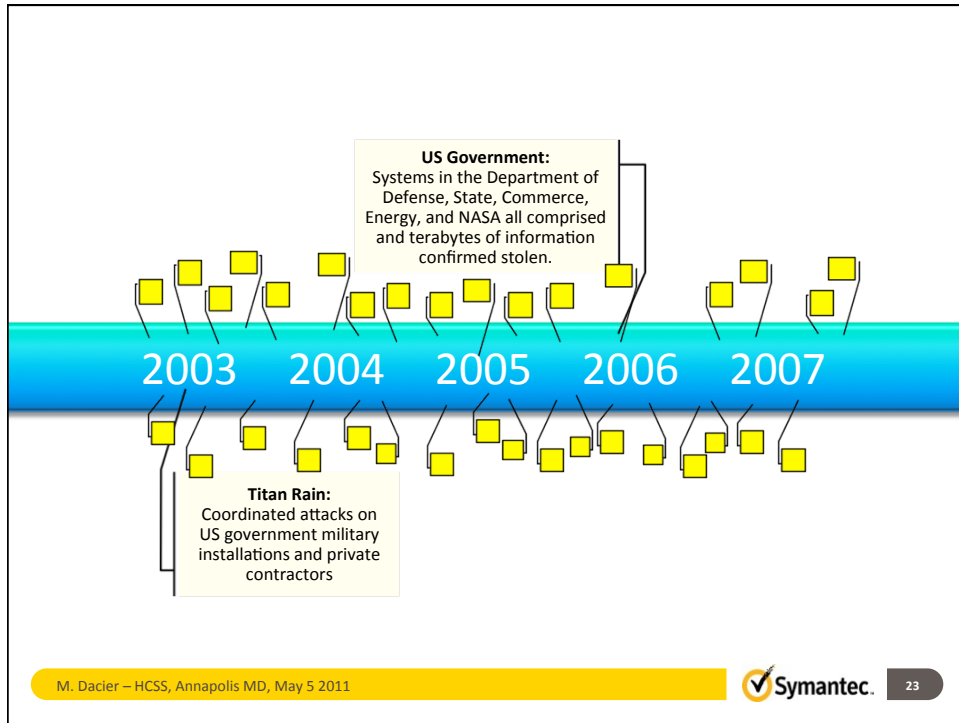
1998 1999 2000 2001

Moonlight Maze:
Attacks targeting US
military secrets reported
to be conducted by Russia

M. Dacier – HCSS, Annapolis MD, May 5 2011  18







New challenges

- Targeted attacks highlight the difficulty to prevent bad things from happening.
- New application domains (social networks, cloud computing, BGP, SCADA, etc.) are also ill suited for a purely preventive approach.
- Securing system requires several families of means:
 - Do your best to build secure things in the first place
 - Be ready to patch vulnerabilities as they are discovered, exploited
 - Have fallback plans available and means to trigger them
 - Measure the risks that systems (to be) deployed are facing

New (?) vocabulary

- Targeted attacks highlight the difficulty to prevent bad things from happening.
- Securing system requires several families of means:
 - Fault Prevention
 - Fault Removal
 - Fault Tolerance
 - Fault Forecasting

Fault Tolerance

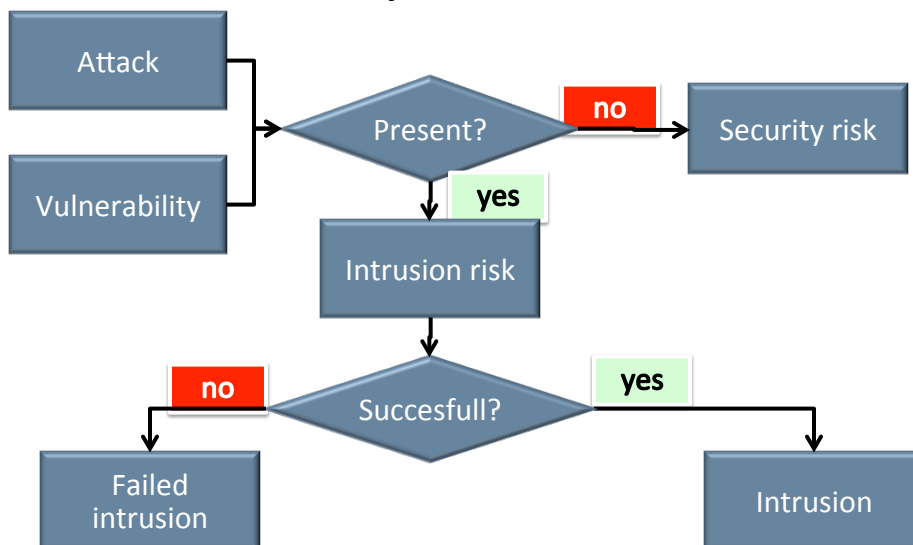
- Error Processing:
 - Error Detection
 - Is something going wrong?
 - Error diagnosis
 - What is really going wrong?
 - Error Recovery
 - How can I fix the situation?
- Fault Treatment:
 - Fault Diagnosis
 - What is the cause of this error?
 - Fault Passivation
 - Can I prevent it from happening again?

M. Dacier – HCSS, Annapolis MD, May 5 2011



27

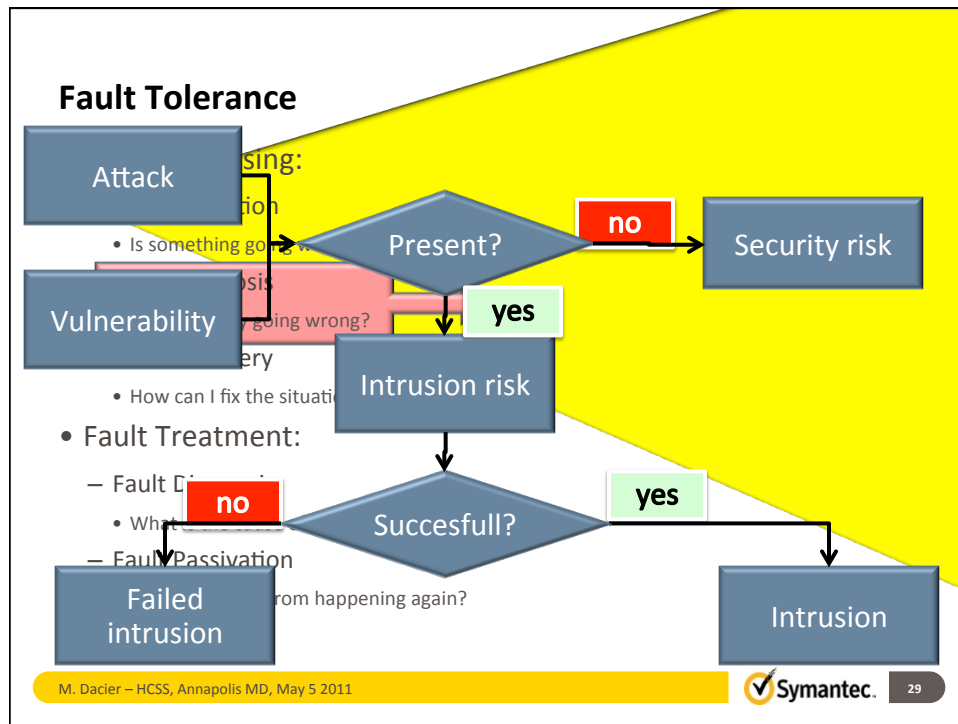
Intrusion, Vulnerability, Attack



M. Dacier – HCSS, Annapolis MD, May 5 2011



28



Side note

- Intrusion detection systems do not, in general, solely detect intrusions
- False positives and False negatives do not exist but error diagnosis tools are missing

Targeted attack vs. intrusion detection

- Prevention is likely to fail
- Once inside, the attacker does not necessarily run attacks anymore.
- Detecting targeted attacks comes down to detecting masqueraders.

Detecting masqueraders: a new problem?

- In 1980, James P. Anderson proposes

«... changes to computer audit mechanisms to provide information for use by computer security personnel when tracking problems ... »

- He Introduces the notion of « *audit reduction* » ...
- .. and the use of « *some sort of statistical analysis of user behavior [...that...]* might represent a way of detecting **MASQUERADERS** »

Old problem, solution found ?

- A large body of knowledge exists in the area of detecting users misusing their privileges
 - Knowledge based approach
 - Behavior based approach
- Almost all possible scientific domains have contributed to a solution.
- The problem is unfortunately still there.

Nice theoretical results

- A number of interesting theoretical results have been published but
 - Experimental validation with real world datasets is almost always lacking
 - The adversarial nature of the mischief imposes a constant monitoring of the effectiveness of the solution.
 - Reproducing experiments, comparing results is not part of the security community culture, to a large extent.

Scientific method

- In general, computer security research (ours included) is not always carried out according to sound scientific methods:
 - *“principles and procedures for the systematic pursuit of knowledge involving the recognition and formulation of a problem, the collection of data through observation and experiment, and the formulation and testing of hypotheses”*
(Merriam-Webster dictionary)
- The lack of experimental validation leads, sometimes to Dahuian research results.

M. Dacier – HCSS, Annapolis MD, May 5 2011



35

Dahu: Definition

- “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,
www.vidonne.com/html/dahu-reignier.htm

M. Dacier – HCSS, Annapolis MD, May 5 2011





Food for thoughts

- Dahus are rare, bizarre, stimulating from an intellectual point of view but ...
 - Does it justify the existence of *Dahusian research*?
 - How can we make sure we are not building tools against *Dahusian hackers*?
 - How can we avoid inventing *Dahusian solutions*?

Overview

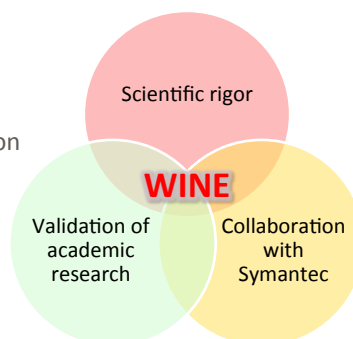
- Introduction
 - Evolution of the threats landscape
 - New threats vs. new challenges
 - WINE
- Conclusions

M. Dacier – HCSS, Annapolis MD, May 5 2011



Goals of WINE Project

- Enable sound experimentation for computer security
 - Real world representative datasets
 - Repeatable experiments
- Promote good science
 - Enable independent verification
 - Ensure statistical relevance



<http://www.symantec.com/WINE>

M. Dacier – HCSS, Annapolis MD, May 5 2011



The questions I will try to answer in the rest of this talk

- What
 - ... data sets does WINE contain?
- Why ...
 - ... are we doing WINE?
- How ...
 - ... can you get access to WINE?
- Who
 - ... can get access to WINE?
- Where
 - ... will WINE be available?
- When
 - ... can you start enjoying WINE?

M. Dacier – HCSS, Annapolis MD, May 5 2011



Representation vs. Reality

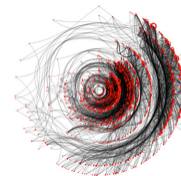


René Magritte (1898-1967)

M. Dacier – HCSS, Annapolis MD, May 5 2011



What datasets does WINE contain?



- Binary Reputation
- A/V Telemetry
- Email Spam
- URL Reputation
- Malware Samples

And much more: Symantec Management Platform (Altiris), Endpoint Virtualization Control Compliance, Data Loss Prevention, Backup Exec, Veritas Storage Foundation, Deepsight, PC Tools, PGP, Verisign, etc...

M. Dacier – HCSS, Annapolis MD, May 5 2011



43

What datasets does WINE contain? (ctd.)

Binary Reputation

- Sources: 3.5 million machines
- Information on unknown binaries—i.e., files for which an A/V signature has not yet been created—that are downloaded by users who opt in for Symantec's reputation-based security program.
- This data can indicate for how long a particular threat has existed in the wild before it was first detected.
- Each record includes the submission timestamp, as well as the cryptographic hash and the download URL of the binary.

M. Dacier – HCSS, Annapolis MD, May 5 2011

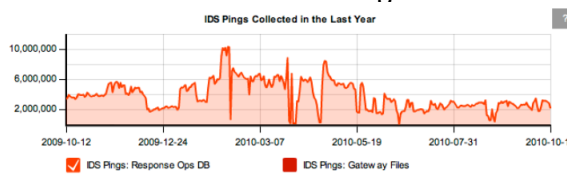


44

What datasets does WINE contain? (ctd.)

A/V Telemetry Data Sets

- Sources: 90 million machines
- Records occurrences of known threats, for which Symantec has created signatures and which can be detected by anti-virus products.
- Includes intrusion- detection telemetry.
- Each record includes the detection timestamp, the signature of the attack, the OS version of the attack's target, the name of the compromised process and the file or URL which originated the attack.



Example: statistics for intrusion-detection telemetry

M. Dacier – HCSS, Annapolis MD, May 5 2011



What datasets does WINE contain? (ctd.)

Email Spam

- Sources: 2.5 million decoy accounts
- Samples of phishing and spam emails, collected by Symantec's enterprise-grade systems for spam filtering.
- This data set includes samples of email spam and statistics on the messages blocked by the spam filters.

M. Dacier – HCSS, Annapolis MD, May 5 2011



What datasets does WINE contain? (ctd.)

URL Reputation

- Sources: 10 million domains
- Website-reputation data, collected by crawling the web and by analyzing malicious URLs
- Each record includes the crawl timestamp, the URL, as well as the name and the type of threat found at that URL.
- A subset of this data was used to analyze the rogue A/V campaigns [Cova et al. 2010].
- A simplified interface for querying this data is available at <http://safeweb.norton.com/>.

[Cova et al. 2010] Cova M., Leita C., Thonnard O., Keromytis A., Dacier M., "An analysis of rogue AV campaigns", *In International Symposium on Recent Advances in Intrusion Detection*, RAID 2010 Ottawa, Canada, 442–463.

What datasets does WINE contain? (ctd.)

Malware samples

- Sources: 200 countries
- A collection of both packed and unpacked malware samples (viruses, worms, bots, etc.), used for creating Symantec's A/V signatures.
- A subset of these samples was used for validating research on automatic malware detection [Griffin et al. 2009].

[Griffin et al. 2009] Griffin K, Schneider S. Hu X., Chiueh T.-C., "Automatic generation of string signatures for malware detection", *In International Symposium on Recent Advances in Intrusion Detection*, RAID 2009, Saint-Malo, France, 101–120.

Why are we doing WINE?

- To promote sound scientific experimentation in the security area
- To have more smart people looking at new ways of using our datasets
- To convince you that Symantec is *the place* to work
- To simplify the establishment of long term relationships and technology transfer.

M. Dacier – HCSS, Annapolis MD, May 5 2011



49

How can you get access to WINE?

- The data do not come to you, you have to come to the data. We grant you access to our premises to run your programs on our data.
- This is the only possibility to offer you access to very large, real world, representative raw data
- You stay typically 3 weeks on site, it could be much more, of course.
- IP remains yours.
- You are free to publish results
- Datasets used will be frozen and offered to any other team willing to replicate your experiments or compare their approach with yours.

M. Dacier – HCSS, Annapolis MD, May 5 2011



50

Intellectual property and usage

- Non-disclosure agreements to protect the confidentiality of the shared data but with a provision for publication
- Symantec receives copies of all research products
- Researchers assume all risks and liabilities from use of data
- All right, title and interest belong to the researchers
 - Unless licensing exception is negotiated beforehand
 - Data set should be acknowledged in publications



Who can get access to WINE?

- This is open to any non profit organization, mostly targeting academics.
- You need to describe before what data you need and what experiments you want to run so that we can double check the feasibility of the experiment.
- If you are funded by NSF, you are allowed to use your budget to pay the lump sum per day we ask you to offset our incremental costs.
- If we receive too many requests, an external scientific board will be in charge of the selection.
- No restriction on the scientific domain you come from (eg machine learning, visualization, economics, social sciences ...)

Operational Model



- Project proposals
 - Researchers in academia request access to data sets
 - NSF support: Trustworthy Computing program
http://www.gtisc.gatech.edu/nsf_workshop10_data.html
- Internal operations
 - Collect data continuously
 - Each proposal's requested data is frozen as reference
 - Experimental environment is hosted on SRL site
- Selection of projects
 - Advisory board: senior researchers (external and internal)

M. Dacier – HCSS, Annapolis MD, May 5 2011



53

Where will WINE be available?

- To start with, we will have two labs offering access to WINE:
 - Culver City (Los Angeles, CA): also home of our STAR team located on the same floor as people working on WINE
 - Herndon (Washington DC): also home of one of our Security Operation Centers

M. Dacier – HCSS, Annapolis MD, May 5 2011



54

When can we start enjoying WINE?

- Right now.
- We have already received in excess of 20 requests from various organizations.
- We expect to have at least 3 (resp. 5) teams to start working on WINE before the end of this academic (vs. calendar) year.

Overview

- Introduction
 - Evolution of the threats landscape
 - New threats vs. new challenges
 - WINE
- Conclusions

Conclusions

- The threat landscape is changing
- We face new challenges (eg social networks, SCADA, etc.)
- We need sound scientific approaches and avoid Dahusian research
- WINE is there for you. Use it!
- We are also, of course, welcoming collaboration opportunities!

M. Dacier – HCSS, Annapolis MD, May 5 2011



57



Thank you!

WINE
Can be used without moderation

www.symantec.com/WINE

marc_dacier@symantec.com

Copyright © 2010 Symantec Corporation. All rights reserved. Symantec and the Symantec Logo are trademarks or registered trademarks of Symantec Corporation or its affiliates in the U.S. and other countries. Other names may be trademarks of their respective owners.

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.

M. Dacier – HCSS, Annapolis MD, May 5 2011

58