A Brief History of Time

Igor Muttik - McAfee Labs™

CARO'2010
Helsinki

100 WEEKS ON THE NEW YORK TIMES BESTSELLER LIST
THE MILLION-COPY HARDCOVER BESTSELLER

A BRIEF HISTORY OF TIME

FROM THE BIG BANG TO BLACK HOLES

STEPHEN HAWKING

WITH AN INTRODUCTION BY CARL SAGAN

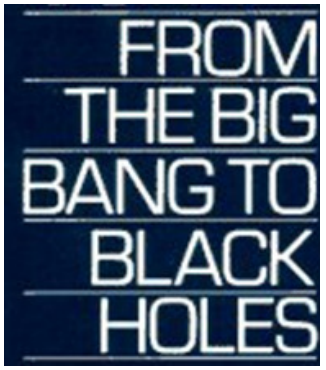
Agenda

McAfee

- Contemporary malware attacks
- Current state of testing
- Proposed metric
  - Multiple attacks
  - Practical implementation
- Conclusions and questions

Contemporary attacks

Contemporary attacks and protection

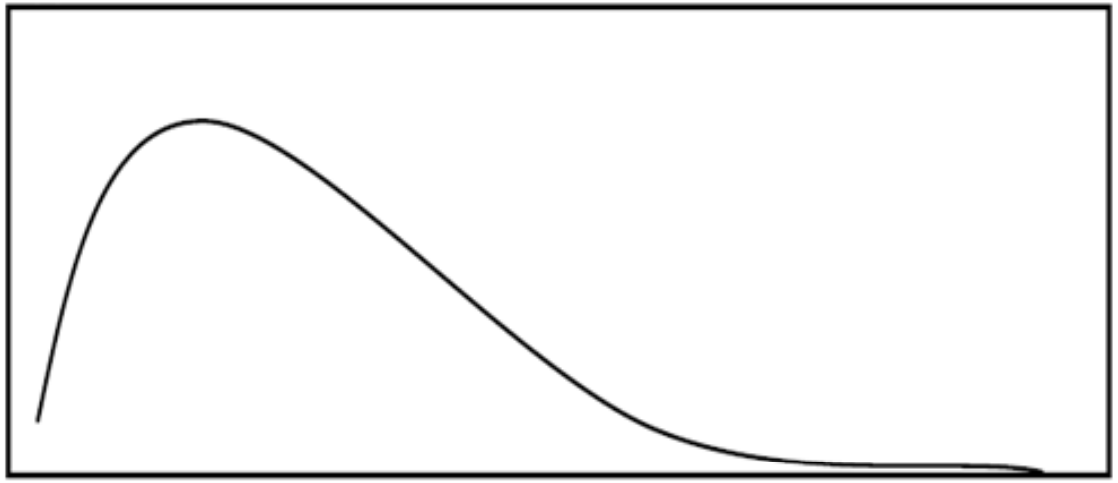


- Short-lived threats
  - When return on investment (ROI) drops – new malware is released
  - Attacks come in waves to sustain ROI
  - Replicating malware (viruses) live longer
  - As protection gets quicker - attacks become shorter

- After a handful of attacks global response protects all users – is it reactive or is it proactive?
  - Applies to quick updating
  - Applies to cloud-based products
  - Applies to pure behavioral technologies too

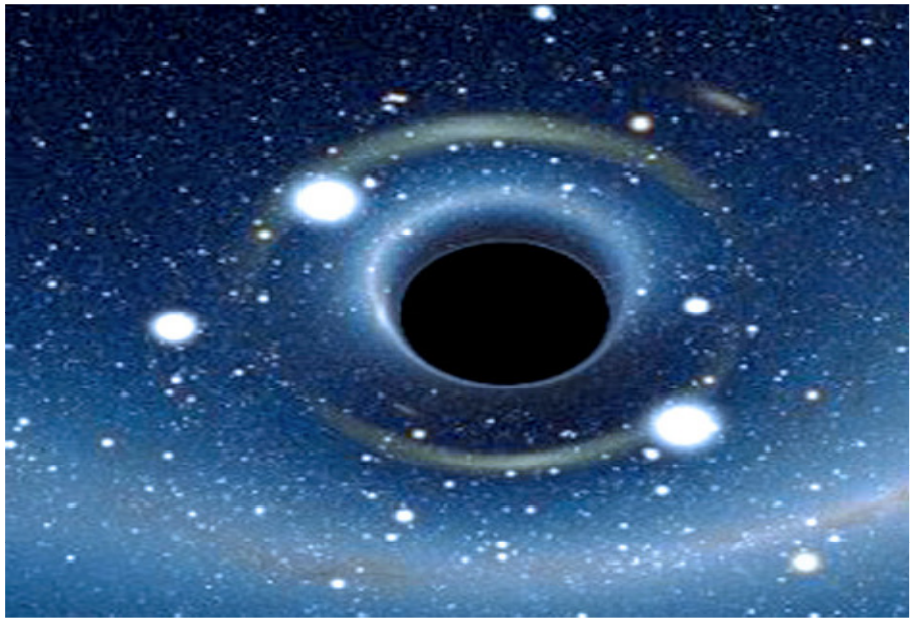
Attack pattern

McAfee



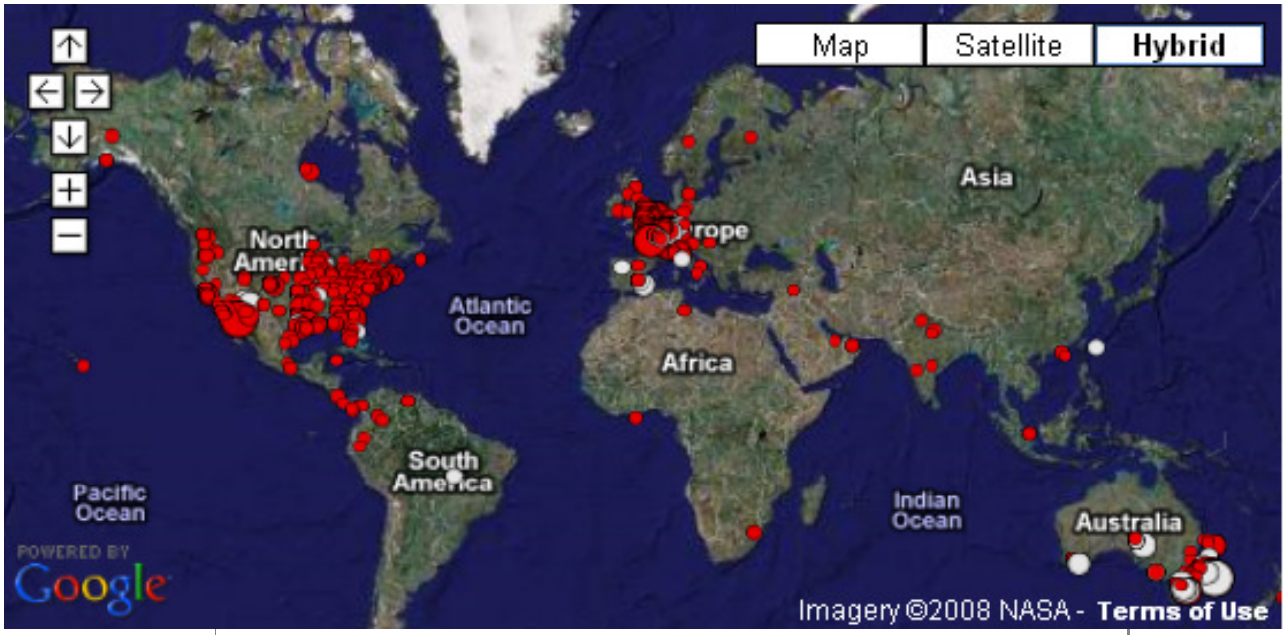
BIG BANG

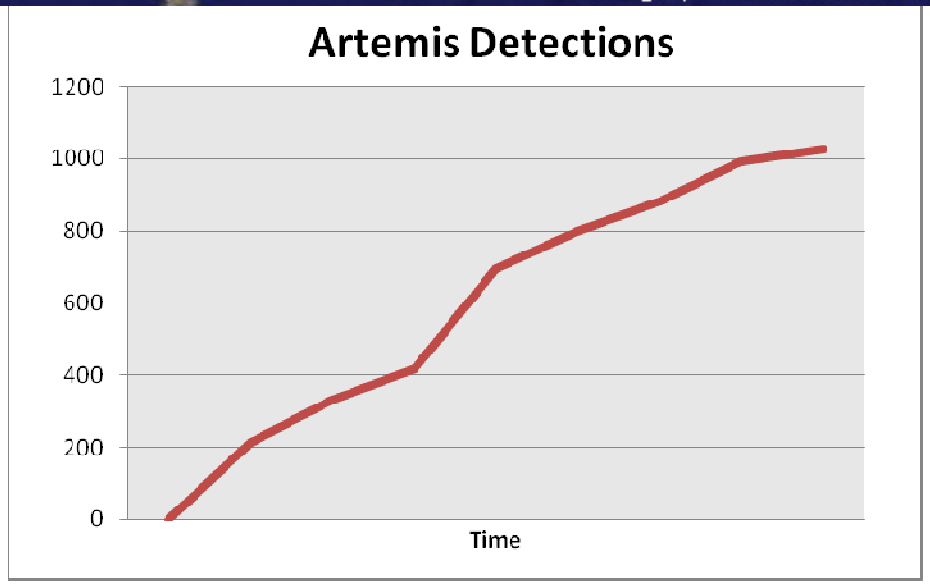
BLACK HOLE



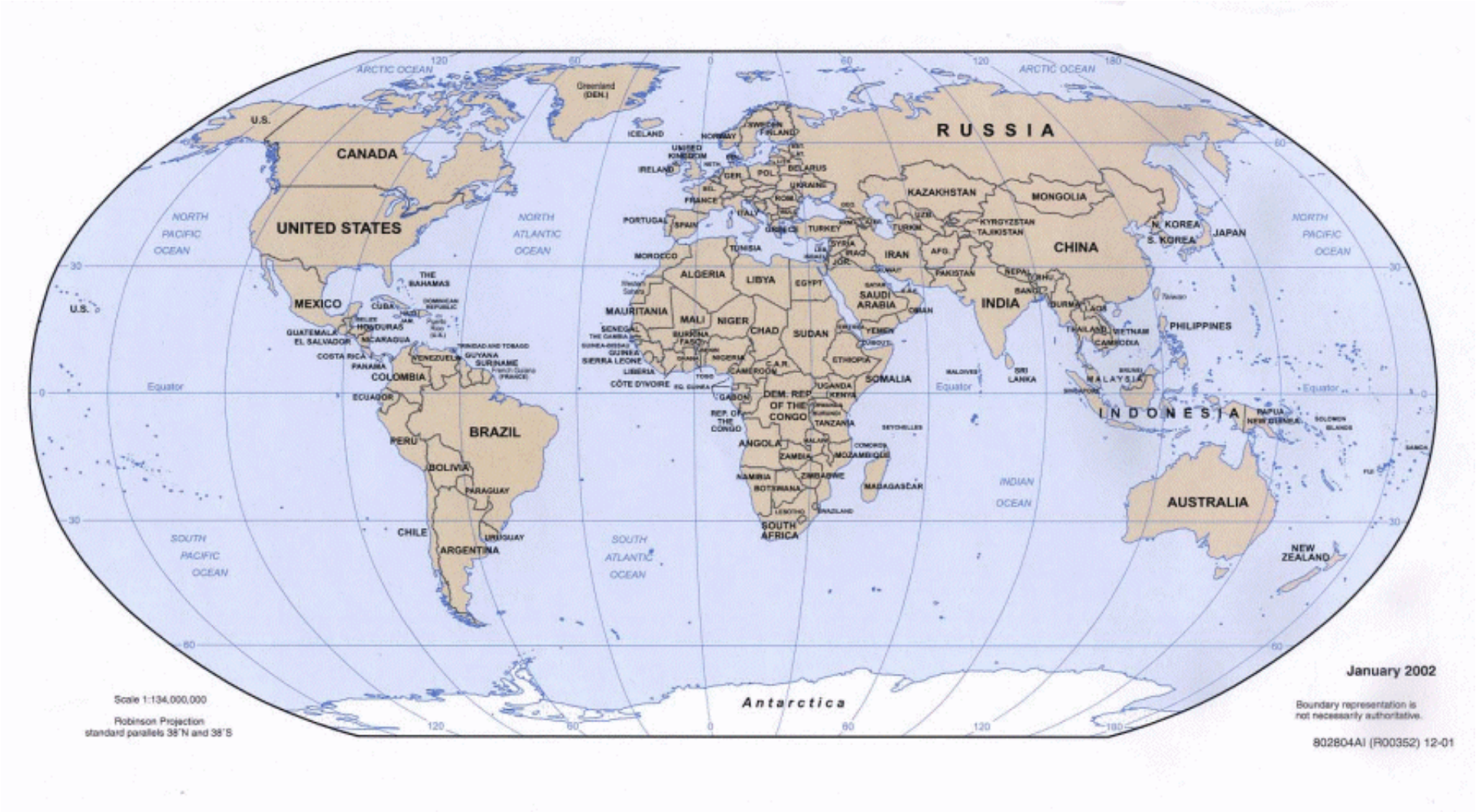
Field telemetry example: Spy-Agent.bw

McAfee





Simulation



- Source: D.Gryaznov “Taking Down the Internet”, proceedings of Virus Bulletin conference, 2003

Combined Telemetry

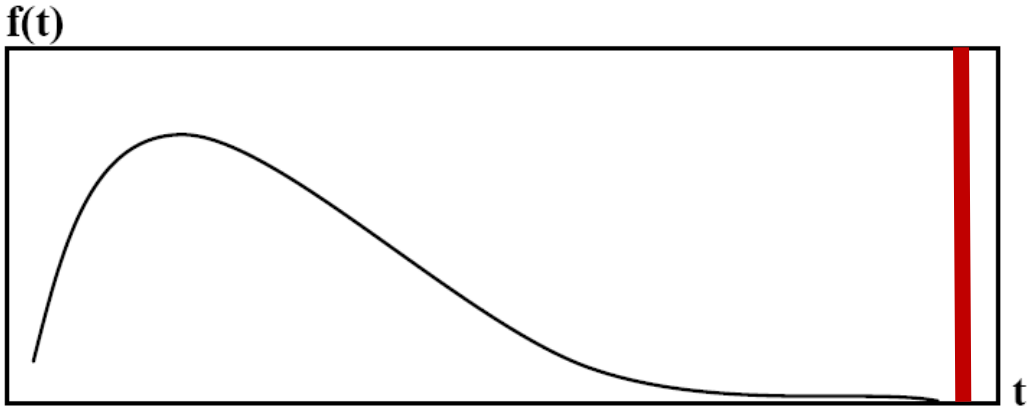
McAfee

Current state of testing – reactive and proactive

Problems with “detected/missed” approach

McAfee

- Detection rate over a collection ≠ probability of protection
- “Reactive” scanners’ tests give detection rates >95%
  - Assumes protection was available
  - Do not work well - easy to manipulate

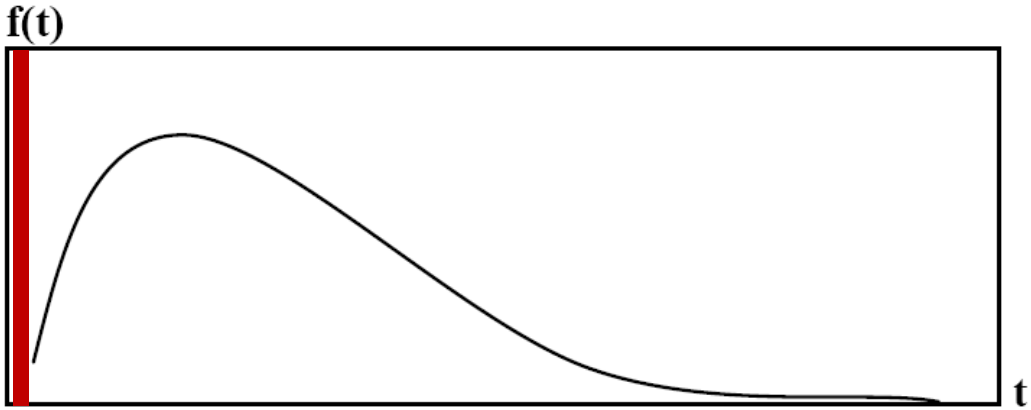


A graph of field sightings vs time

“Reactive” and “proactive” both look rather silly…

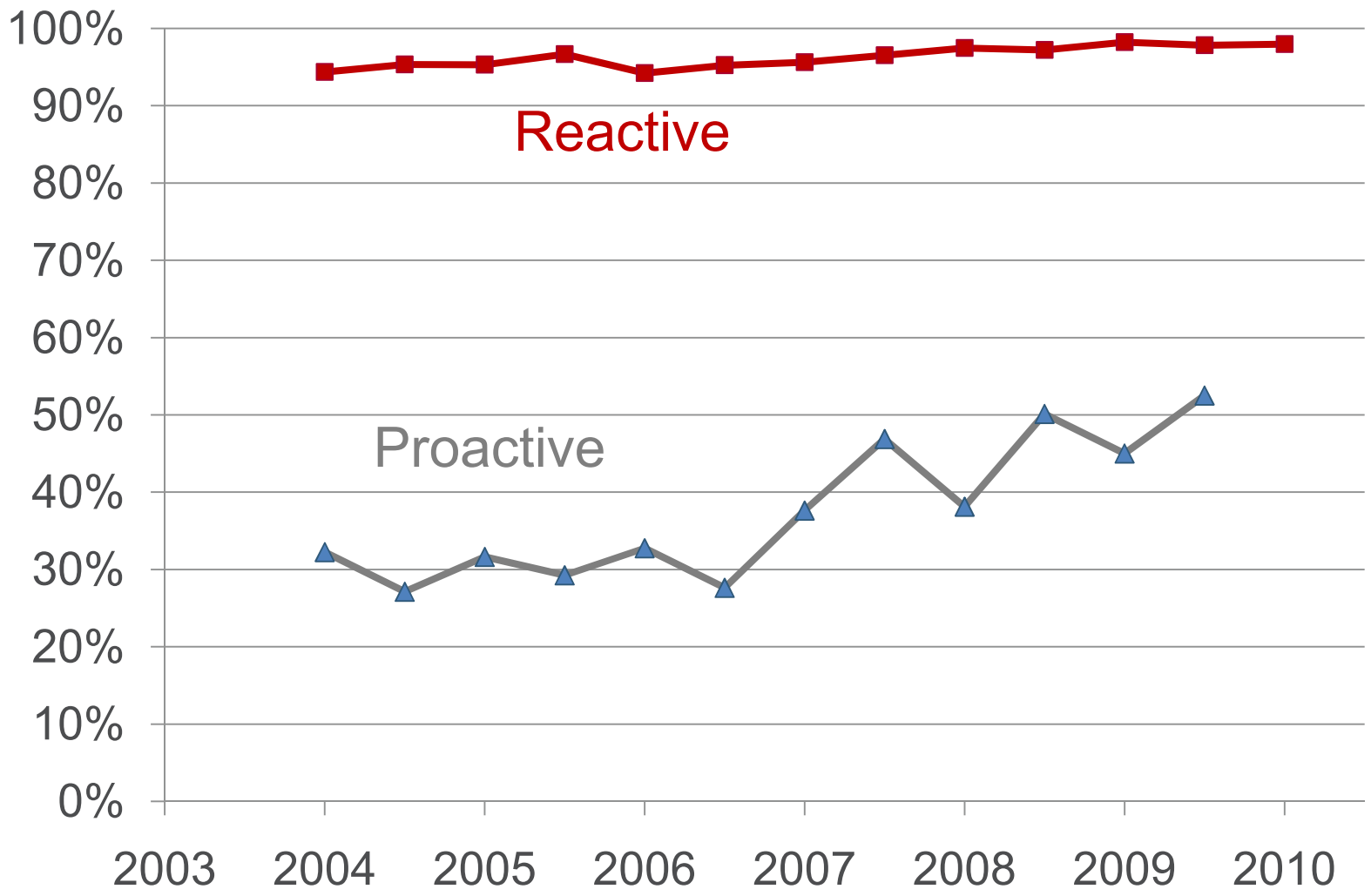
McAfee

- Proactive (a.k.a. “retrospective”) tests show 30-50%
  - Do not work well with quick updating (e.g. streaming anti-spam rules)
  - Do not work at all for cloud-based products
  - Assumes protection is reliable



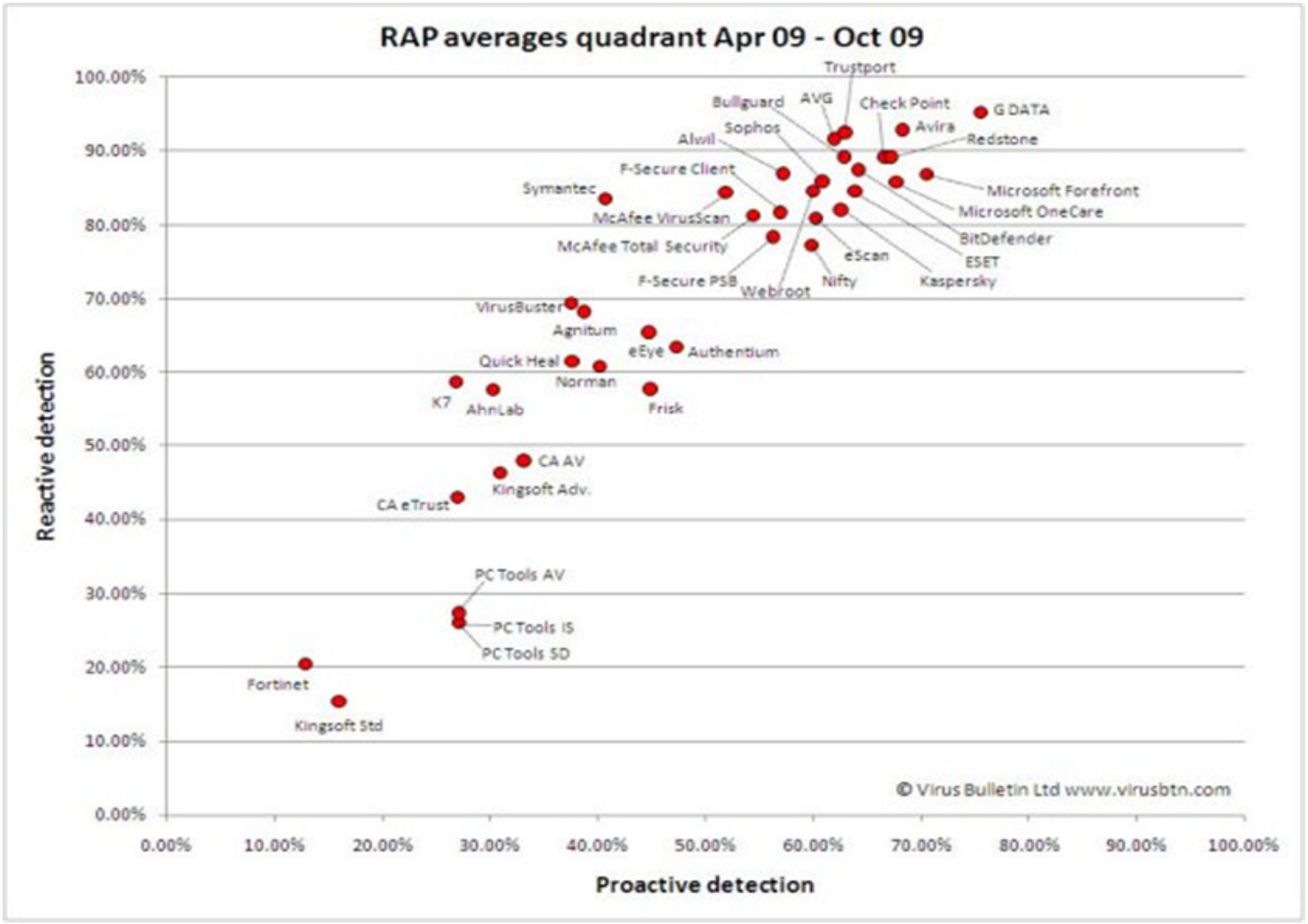
A graph of field sightings vs time

AV industry in proactive and reactive tests



Source: average rates for 6 common AV scanners based on the data from www.AV-Comparatives.org

Mixing proactive and reactive tests



- Source: http://www.virusbtn.com/news/2009/10_09.xml?rss

McAfee

Proposed metric – based on timing and coverage

Timing of protection is the key

McAfee

- Majority of existing tests are too simplistic
  - That's why AMTSO (www.amtso.org) exists
  - Check protection at the wrong time
  - Do not track protection reliability

- For a user the "proactive" and "reactive" are meaningless. They want to know:
  - Am I protected now from current threats?
  - What is the best security product going forward?

How can we measure “the best”?

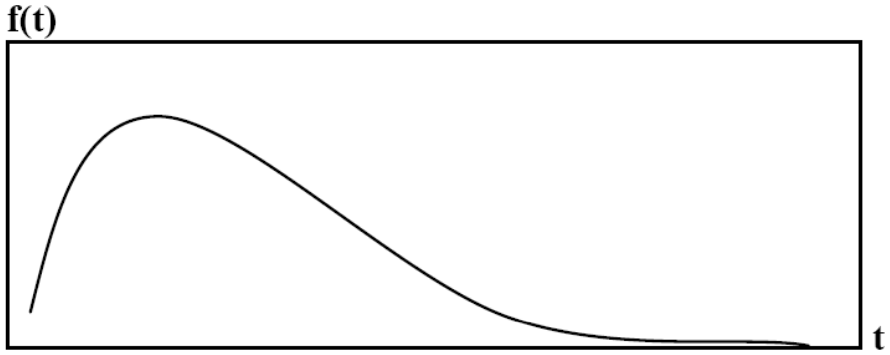
McAfee

- We need to monitor the attack (timing and intensity)
- We need to monitor the security response when there is any field attack
- (We record the security response but may not know yet if it is correct or not. So we collect the data but will process it later.)

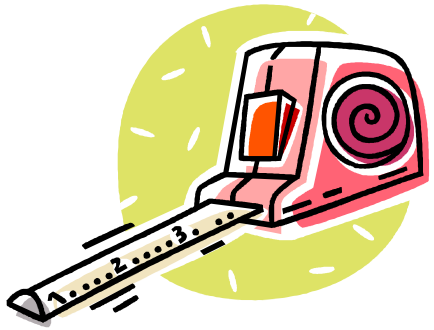
Sampling of the attack

McAfee

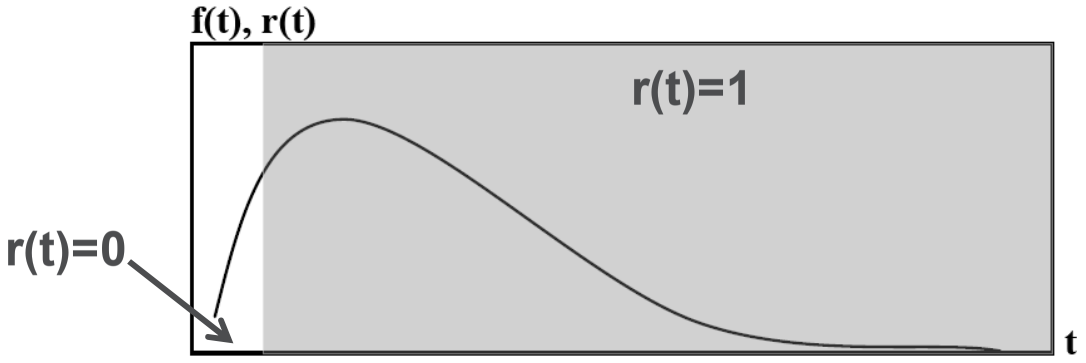
- Attack frequency throughout the whole attack



A graph of field sightings vs time



- Security response sampling

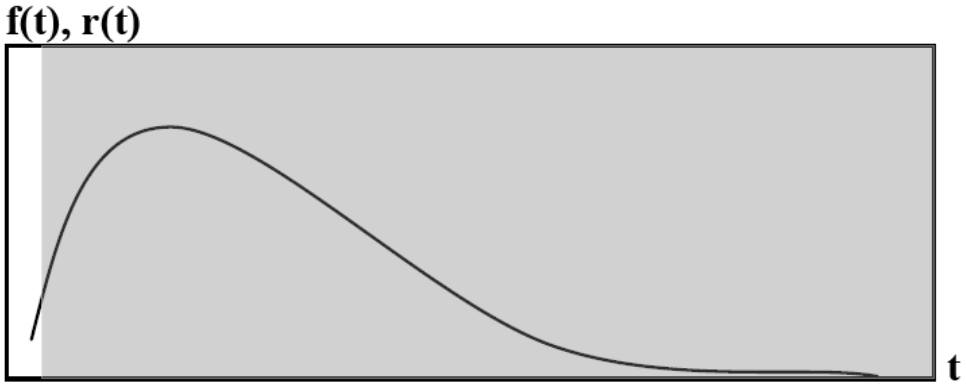


Field sightings f(t) and security reaction r(t) in gray

When “missed” - matters

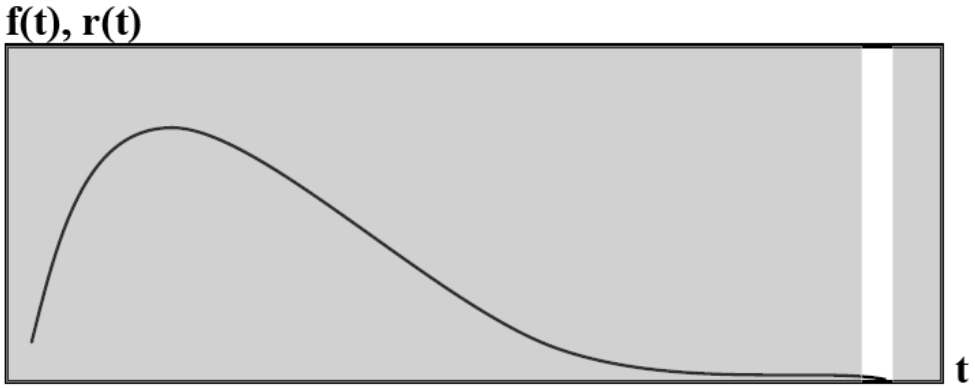
McAfee

- Start:



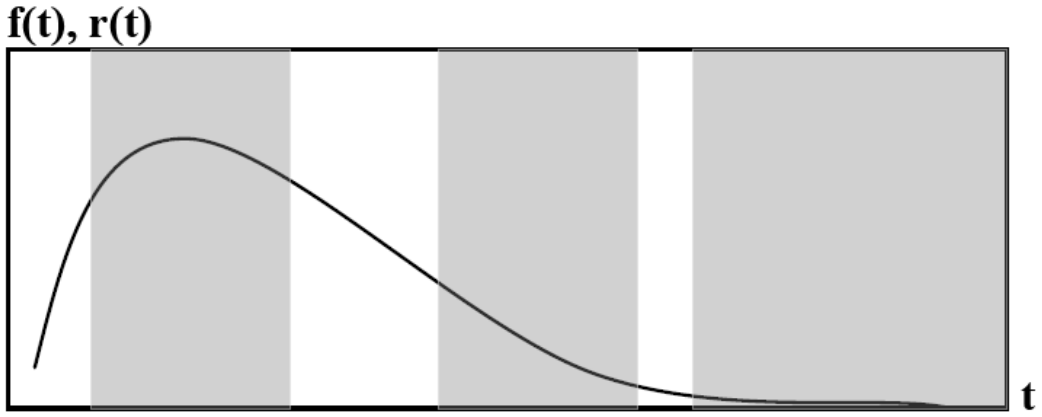
Security reaction r(t) missing attack at start

- Tail:

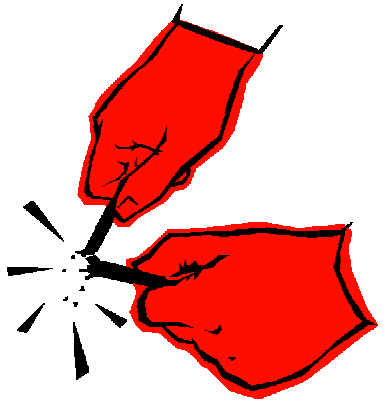


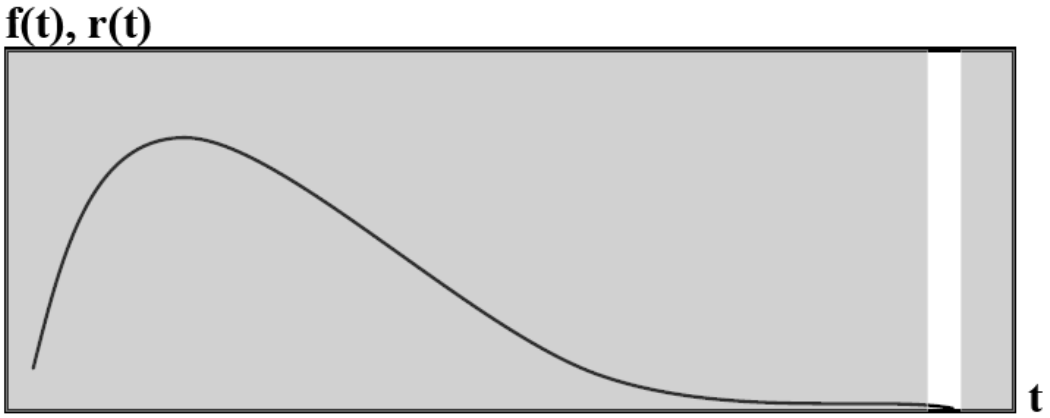
Security reaction r(t) missing attack at the end

Unreliable security response



Unreliable security reaction r(t)

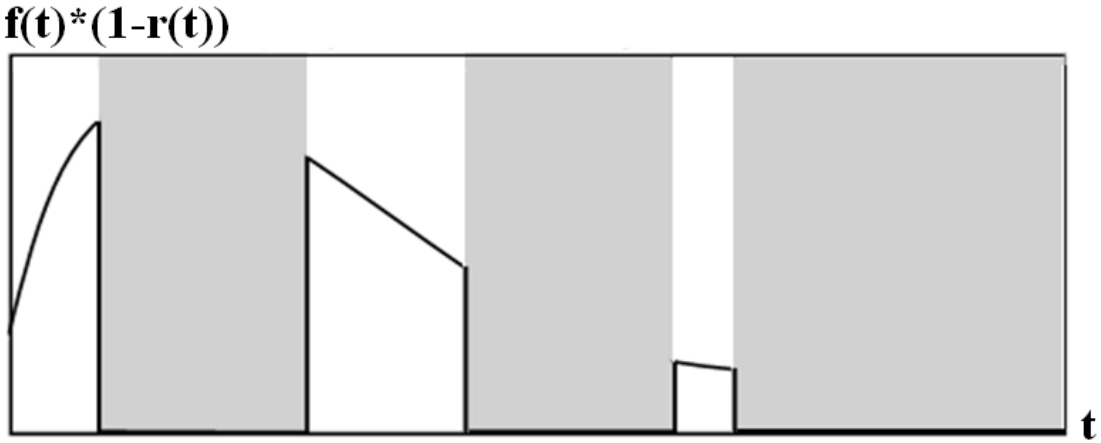




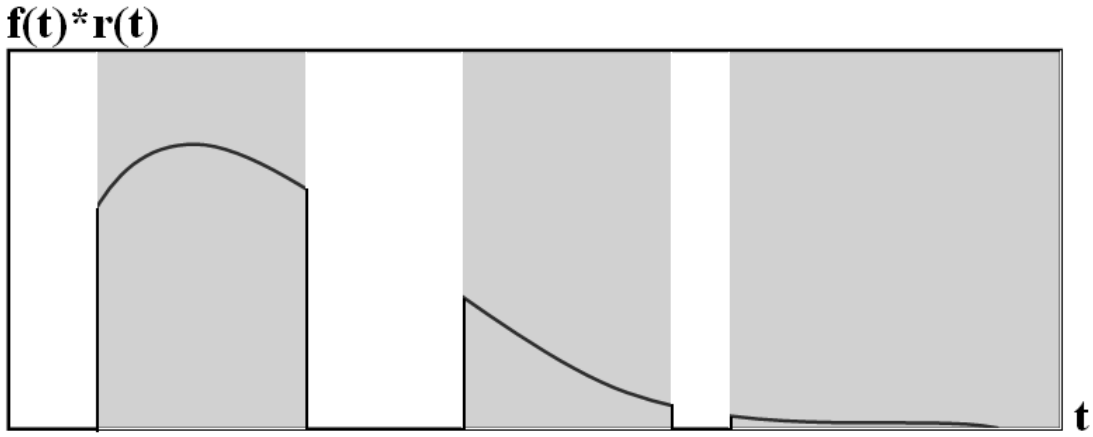
Security reaction r(t) missing attack at the end

Exposure and protection functions

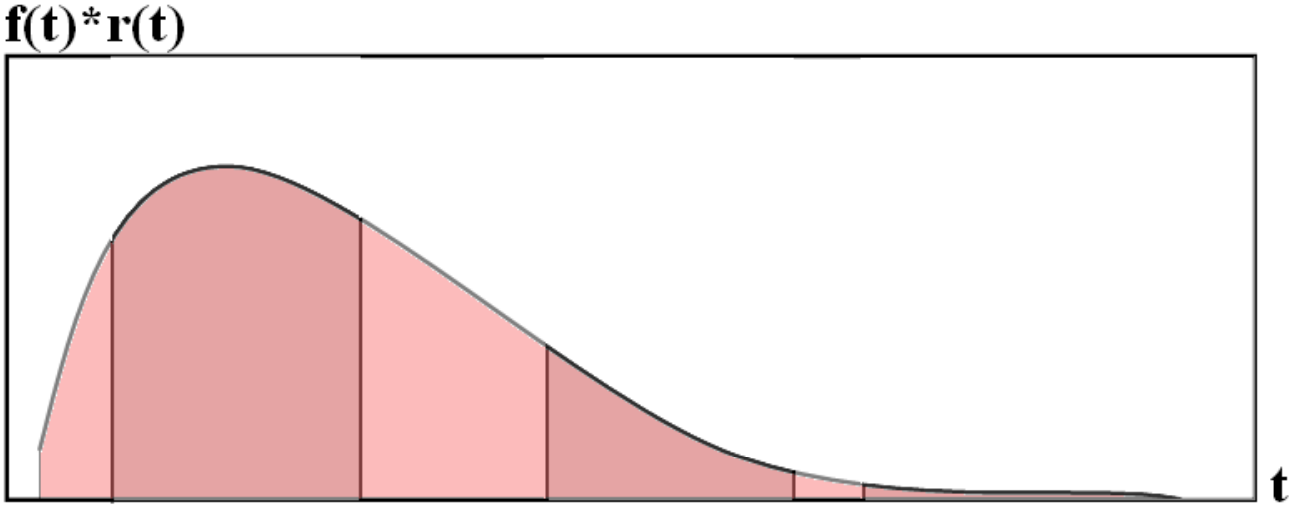
- Exposure:



- Protection:

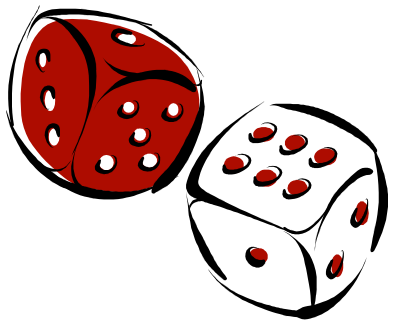


Probability of the protection metric



- We integrate the protection function and normalize:

$$p = \int f(t)*r(t)dt \Big/ \int f(t)dt$$



- $p \in [0...1] = [0\ldots100\%]$

For multiple attacks

McAfee

- We just add them up:

$$\mathbf{p} = \sum_{i=1..N} \left(\int f_i(t) * r_i(t) dt \right) / \sum_{i=1..N} \left(\int f_i(t) dt \right)$$

- But if we do not wish to differentiate attacks – we can treat all of them as one continuous attack
  - Selecting “start” and “finish” becomes subjective
  - Could be “continuous” monitoring
  - But more field sightings makes results more statistically viable

Practical considerations

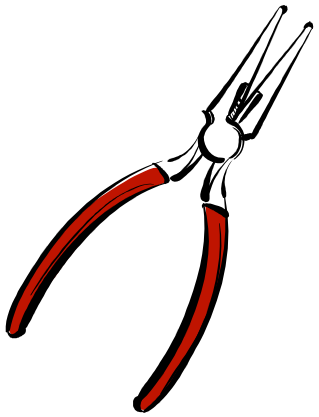
McAfee

- We will have discrete data so our probability will be:

$$p = \sum_{i=1..N} \left(f(t_i) * r(t_i) \right) \Big/ \sum_{i=1..N} f(t_i)$$

- Integrating by time also covers geographical distribution

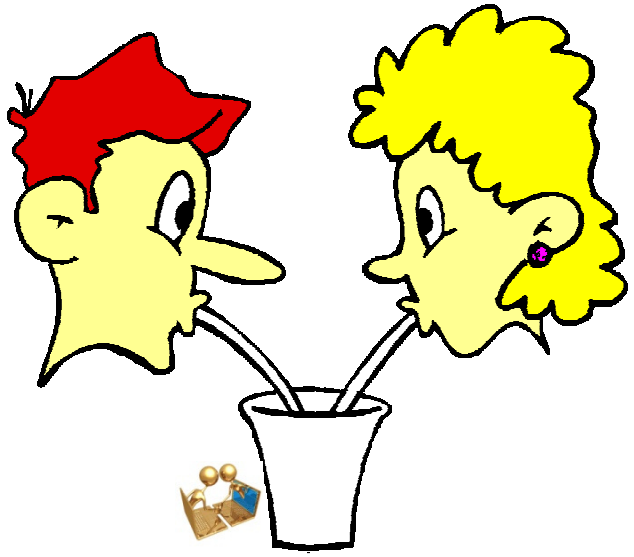
- Security response is only important when there are field sightings



IEEE XML meta-data sharing

McAfee

- Open and free standard of sharing security data
- Pre-IEEE: Norman, Sunbelt, VirusTotal
- IEEE XML standard becomes common
  - 3 companies in July 2009 (AVG, McAfee, Microsoft)
  - 8 companies in May 2010 (+ Symantec, Sophos, Trend, Panda, Eset)
  - 2 more are working on it (at least)



http://grouper.ieee.org/groups/malware/malwg/Schema1.1/

IEEE XML (simple)

McAfee

```
<?xml version="1.0" encoding="UTF-8"?>
<malwareMetaData xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://xml/metadataSharing.xsd metadataSharing.xsd"
    xmlns="http://xml/metadataSharing.xsd" version="1.0" id="1234">
    <company>McAfee</company>
    <author>Raiden</author>
    <comment>This is minimal - just some files</comment>
    <timeStamp>2008-11-25T21:34:56</timeStamp>
    <objects>
        <!-- files -->
        <file id="2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599">
            <!--<attribute type="filename">116.exe</attribute>-->
            <md5>8b31da6402d850ce94e7c19bc97effe1</md5>
            <sha1>850e5b037c799f86f04ee63da786f9ee139ebf57</sha1>
            <sha256>2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599</sha256>
            <size>32769</size>
            <crc32>34efdbca</crc32>
        </file>
        <file id="3a437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599">
            <!--<attribute type="filename">116.exe</attribute>-->
            <md5>aa31da6402d850ce94e7c19bc97effe1</md5>
            <sha1>990e5b037c799f86f04ee63da786f9ee139ebf57</sha1>
            <sha256>22437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599</sha256>
        </file>
    </objects>
</malwareMetaData>
```

IEEE XML (classification)

McAfee

```
<objects>
    <!-- one file -->
    <file id="2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599">
        <!--<attribute type="filename">116.exe</attribute>-->
        <md5>8b31da6402d850ce94e7c19bc97effe1</md5>
        <sha1>850e5b037c799f86f04ee63da786f9ee139ebf57</sha1>
        <sha256>2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599</sha256>
        <size>32768</size>
    </file>

<!-- one classification -->
    <classification id="AVG:Virut.BK" type="dirty">
        <classificationName>Virut.BK</classificationName>
        <companyName>AVG</companyName>
    </classification>
</objects>


<!-- this file is Virut -->
<relationships>
    <relationship type="isClassifiedAs">
        <parents>
            <ref>file[@id = '2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599']</ref>
        </parents>
        <children>
            <ref>classification[@id='AVG:Virut.BK']</ref>
        </children>
    </relationship>
</relationships>
```

Example (field data)

McAfee

```
<!-- this is the prevalence data -->
<fieldData>
    <!-- by file  -->
    <fieldDataEntry>
        <references>
        <ref>file[@id = '2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599']</ref>
        </references>
        <startDate>-1999-11-25T00:00:00</startDate>
        <endDate>2008-11-26T00:00:00 </endDate>
        <origin>user</origin>
        <commonality>8</commonality>
        <location type="countryCodeISO3166-2">US</location>
    </fieldDataEntry>
    <fieldDataEntry>
        <references>
        <ref>file[@id = '2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599']</ref>
        </references>
        <startDate>2008-11-26T00:00:00</startDate>
        <endDate>2008-11-27T00:00:00</endDate>
        <origin>user</origin>
        <commonality>5</commonality>
        <location type="countryCodeISO3166-2">US</location>
    </fieldDataEntry>
    <fieldDataEntry>
        <references>
        <ref>file[@id = '2f437c1c8f73c2d6ffbb6214d3f1ccfe994151b3bd80fe2b3934a1bc89384599']</ref>
        </references>
        <startDate>2008-11-27T00:00:00</startDate>
        <endDate>2008-11-28T00:00:00</endDate>
        <origin>user</origin>
        <commonality>1</commonality>
    </fieldDataEntry>
```

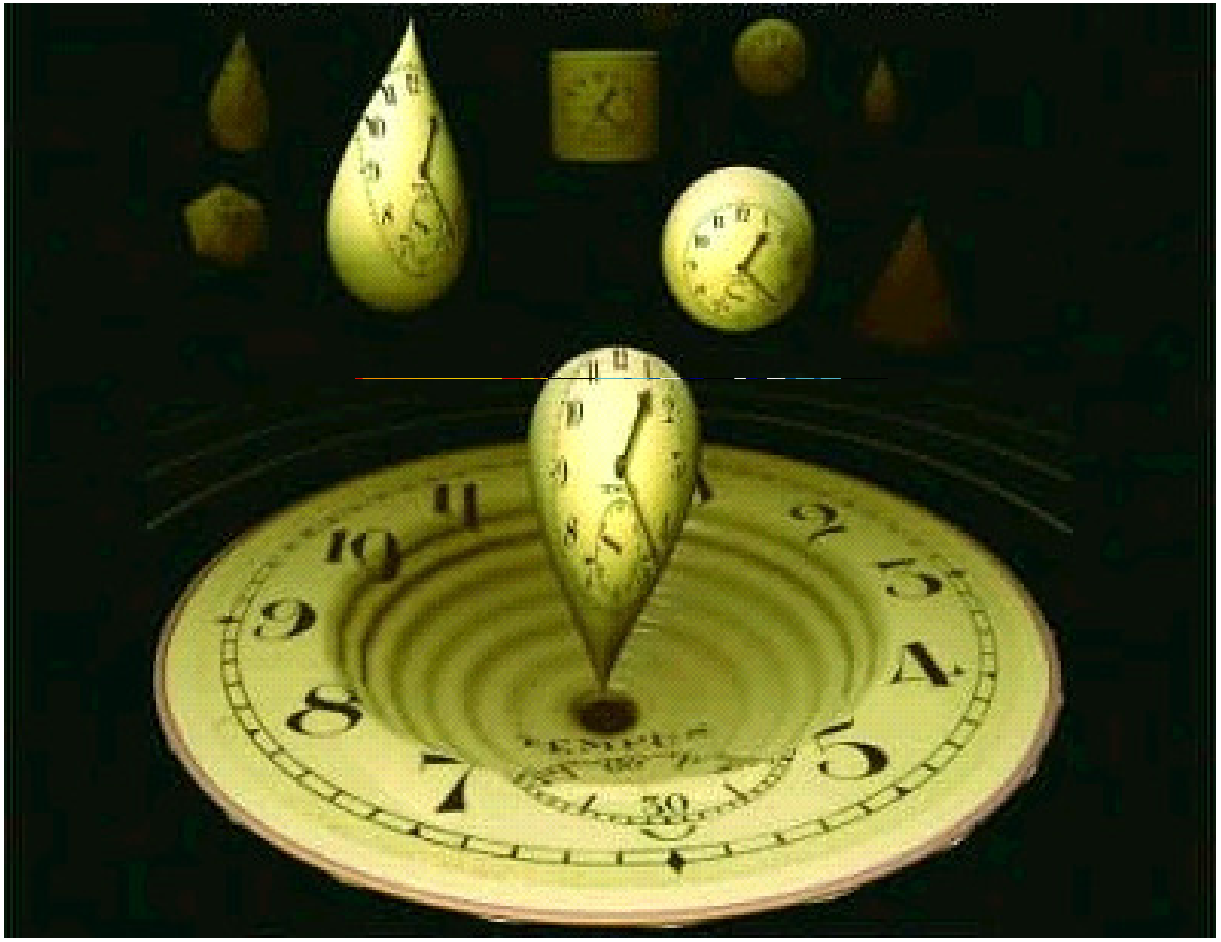
Conclusions

McAfee

- Existing “reactive” and “proactive” tests verify protection at the worst possible times
- Replace “samples” with “attacks”
- Track attacks over time & space, historically

- Our method of computing the probability of protection is generic. It will apply to:
  - Malware protection
  - Anti-spam
  - Vulnerability exposure

Questions



- PDF is at http://www.amtso.org/related-resources.html