Metrics, Methods and Tools to Measure Security and Trustworthiness

Henrique Madeira,
University of Coimbra, Portugal
Measuring trustworthiness

• Trustworthy ICT should be:
  - Dependable
  - Resilient
  - Secure

Require different metrics, methods and tools

to attacks, operational faults and changes

Quite different types of “animals”...
Measuring trustworthiness

- Trustworthy ICT should be:
  - Dependable
  - Resilient
  - Secure

Measuring, assessing, and benchmarking dependability and resilience is not solved...

Measuring, assessing and benchmarking security is even harder to attacks, operational faults and changes
Dependability and resilience metrics

• Dependability
  - Availability - readiness for correct service
  - Reliability - continuity of correct service
  - Safety - absence of catastrophic consequences
  - Integrity - absence of improper system alteration
  - Maintainability - ability to undergo modifications and repairs
  (Performance, measured as response time & throughput, is part of correct service)

Fine!
But just for faults!

• Resilience
  The persistence of delivery of justifiably trusted service in the presence of evolutionary changes. The changes can be planned, predictable, or totally unforeseen

(ReSIST - Resilience for Survivability in IST, www.resist-noe.org)
Security metrics

Metrics Definitions

No consensus exists on what security metrics should be used for measuring security effectiveness. This page documents commentary on metrics definitions from external sources.

Metrics Definitions

The Robert Frances Group recently reported in CSO magazine that the companies it surveyed used these metrics definitions:

<table>
<thead>
<tr>
<th>Metric</th>
<th>% using</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viruses detected in user files</td>
<td>92.3%</td>
</tr>
<tr>
<td>Viruses detected in e-mail messages</td>
<td>92.3%</td>
</tr>
<tr>
<td>Invalid logins (talled password)</td>
<td>64.6%</td>
</tr>
<tr>
<td>Intrusion attempts</td>
<td>64.6%</td>
</tr>
<tr>
<td>Spam detected/filtered</td>
<td>76.9%</td>
</tr>
<tr>
<td>Unauthorized website access (content filtering)</td>
<td>69.2%</td>
</tr>
<tr>
<td>Invalid logins (talled username)</td>
<td>69.2%</td>
</tr>
<tr>
<td>Viruses detected on websites</td>
<td>61.5%</td>
</tr>
<tr>
<td>Unauthorized access attempts (internal)</td>
<td>61.5%</td>
</tr>
<tr>
<td>Admin violations (unauthorized changes)</td>
<td>61.5%</td>
</tr>
<tr>
<td>Intrusion successes</td>
<td>53.6%</td>
</tr>
<tr>
<td>Unauthorized information disclosures</td>
<td>38.5%</td>
</tr>
<tr>
<td>Spam not detected (missed)</td>
<td>38.5%</td>
</tr>
<tr>
<td>Spam false positives</td>
<td>30.8%</td>
</tr>
</tbody>
</table>
Security metrics

• The problem is not the lack of security metrics

• Different groups of metrics
  ◆ Organizational Security Metrics
    – Measure the effectiveness of organizational programs and processes
    – Related to best practices
    – Specify levels of maturity
  ◆ Technical Security Metrics
    – Access security attributes of ICT objects
    – Many ad hoc metrics… addressing quite specific aspects
  ◆ Operational Security Metrics
    – Measure security risks based on metrics produced as a part of normal operations
    – Operational readiness
    – Security posture (e.g., measure antivirus protection, password strength,…)
Measuring trustworthiness

- Trustworthy ICT should be:
  - Dependable
  - Resilient
  - Secure

Measuring, assessing, and benchmarking dependability and resilience is not solved...

Measuring, assessing and benchmarking security is even harder to attacks, operational faults and changes

A single word is not enough...

**Measuring, assessing and benchmarking** trustworthiness
Measuring

The act of obtaining a proper measurement for a parameter or metric. It relies on a quantitative with well-known scale/reference.

Measurement uncertainty can make the difference between good and bad measurements.

Quality of measurements is all about uncertainty evaluation.
Assessing

The act of classifying something with respect to its worth. It can just be qualitative.
Benchmarking

Agreement/contract and well-identified properties to ensure fairness in comparison (DBench project).
What we have in mind when we talk about security measuring?

• Qualitative and quantitative measurements

• Quantitative measurement in two flavors:
  - Relative sense (benchmarking), to choose among alternatives
  - Absolute sense, to give guarantees to users

• Predictive value

• At different levels
  - Organizational measurements: measure effectiveness of organizational processes
  - Technical measurements: assess/compare technical objects
  - Operational measurements: produced as a part of normal operations

• Throughout systems lifecycle
Limitations of existing security measuring methods?

• In general, are focused on specific aspects of security.

• Lots of useful methods that help improving security, but none of them can quantify security.
  ◆ Methods based on process guidelines and best practices
  ◆ Red Teams, that may be effective in finding problems
  ◆ Formal methods fail proving absolute security in realistic scenarios, although useful to find problems
  ◆ Vulnerability detection approaches
  ◆ Etc…

• Lack of predictive metrics and methods.

• Organizational-level and technical security metrics are not integrated to provide a comprehensive view
Measuring security; judging on trust

• Security is a feature of a system or a service
  (taking system/service in both the technical and organizational levels and including people as elements of the system)

• Trust is a relationship between two entities
  ◆ Linked to confidence on something (a person, a machine, the related environment)
  ◆ Dependent on the context
  ◆ Varies along the time
  ◆ Non-symmetric and non-transitive

• Security measurements could provide support/evidences for the establishment of trust
Challenges

• **Challenge 1:** Define the right elements
  - Metrics
  - Evaluation methods
  - Tools

• **Challenge 2:** Validation

• **Challenge 3:** Impact and relevance
Challenge 1

Metrics, methods, and tools

- Appropriate metrics
- Methods for estimating metrics
- Effective evaluation tools
Challenge 1
Metrics, methods, and tools

• Appropriate metrics

• Methods for estimating metrics

• Effective evaluation tools

• Which metrics? Just a small number? Specific of a given context? Metrics purpose (related to requirements)?

• Metrics at different levels: component, system, infrastructure, organization,…

• Product oriented and process oriented

• Applied throughout the system lifecycle

• Metric composition, namely how to relate organizational and technical metrics.
Challenge 1
Metrics, methods, and tools

• Appropriate metrics

• Methods for estimating metrics

• Effective evaluation tools
  • Which methods? Formal, experimental, risk assessment, threat & vulnerability assessment, ...
  • Should be both model and experimentally based
Challenge 1

Metrics, methods, and tools

- Appropriate metrics?

- Methods for estimating metrics

- Effective evaluation tools
  - Effective tools that can be used by practitioners
  - Integration of different tools in a toolset, including tools for organizational and technical metrics
Challenge 1

Metrics, methods, and tools

Example of benchmarking vulnerability scanners tools

- Appropriate metrics?
- Methods for estimating metrics
- Effective evaluation tools
- Effective tools that can be used by practitioners
- Integration of different tools in a toolset, including tools for organizational and technical metrics

Scanners used
- Acunetix Web Vulnerability Scanner 4
- Watchfire AppScan 7 (acquired by IBM)
- Spi Dynamics WebInspect 6.32 (acquired by HP)
Challenge 2
Validation

• Properties
  ◆ Representativeness
  ◆ Repeatability
  ◆ Reproducibility
  ◆ Portability
  ◆ Scalability
  ◆ …

• Generalization

• Just being pragmatic is enough?
Challenge 3
Impact and relevance

• How to select the challenges and research problems that maximize the impact of research outcomes?

• Metrics that promote security improvement
  ◆ User view
  ◆ Industry
  ◆ Metrics that show the return on investment

• Regulation policy

• Benchmarking
How can we give some contribution? (back to 5 years ago)

Apply methods used in the dependability and resilience evaluation to the security field

- Fault injection
- Robustness testing
- Dependability benchmarking

Vulnerability injection
- Attack injection
- Security benchmarking
What is fault injection?
(and a little bit of history…)

Deliberate insertion of upsets (faults or errors) in computer systems to evaluate its behavior in the presence of faults or validate specific fault tolerance mechanisms in computers.
Examples of fault injection approaches

- Pin-level fault injection (e.g., RIFLE)
  
  Reproduce pin-level fault injection by software

- Software Implemented Fault Injection (e.g., Xception)
  Injection of hardware faults only!
Examples of fault injection approaches

- Pin-level fault injection (e.g., RIFLE)

- Software Implemented Fault Injection (e.g., Xception)
  Reproduce pin-level fault injection by software

Injection of hardware faults only!

What about injecting software faults?
Two possible injection points (software faults)

1. Injection of interface faults in software components (classical robustness testing)

2. Injection of realistic software faults inside software components (new approach)
Example of results from interface faults (robustness testing)

Robustness failure in RTEMS 4.5.0

Excerpt of application code:

```c
requestedSize1 = 4294967295;
returnStatus = rtems_region_get_segment (regionId,
             requestedSize1,
             option,
             timeout,
             ptsegment1);
```

Result:

Memory exception at ffffffff (illegal address)
Unexpected trap (0x09) at address 0x0200aaac
Data access exception at 0xffffffffc

This software fault was discovered automatically by the Xception robustness testing tool!
Example of results obtained with injection of software faults

What happens if a software bug in the disk device driver becomes active?

1. Software faults are injected in the disk device driver using the G-SWFIT technique.

2. The device is heavily used by programs.
Example of results obtained with injection of software faults

What happens if a software bug in the disk device driver becomes active?

**Availability**

**Feedback**

**Stability**

![Graphs comparing Windows NT, Windows 2000, and Windows XP](image-url)
Why injection or real software faults?

- Error propagation through non conventional channels is a reality.
- Faults injected inside components are more representative.
Which are the most representative SW faults?

- Field data on real software errors is the most reliable information source on which faults should be injected.

- Typically, this information is not made public.

- Open source projects provide information on past (discovered) software faults.
## Open source field data survey

<table>
<thead>
<tr>
<th>Programs</th>
<th>Description</th>
<th># faults</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDEX</td>
<td>CD Digital audio data extractor.</td>
<td>11</td>
</tr>
<tr>
<td>Vim</td>
<td>Improved version of the UNIX vi editor.</td>
<td>249</td>
</tr>
<tr>
<td>FreeCiv</td>
<td>Multiplayer strategy game.</td>
<td>53</td>
</tr>
<tr>
<td>pdf2h</td>
<td>pdf to html format translator.</td>
<td>20</td>
</tr>
<tr>
<td>GAIM</td>
<td>All-in-one multi-protocol IM client.</td>
<td>23</td>
</tr>
<tr>
<td>Joe</td>
<td>Text editor similar to Wordstar®</td>
<td>78</td>
</tr>
<tr>
<td>ZSNES</td>
<td>SNES/Super Famicom emulator for x86.</td>
<td>3</td>
</tr>
<tr>
<td>Bash</td>
<td>GNU Project's Bourne Again SHell.</td>
<td>2</td>
</tr>
<tr>
<td>LKernel</td>
<td>Linux kernels 2.0.39 and 2.2.22</td>
<td>93</td>
</tr>
<tr>
<td><strong>Total faults collected</strong></td>
<td></td>
<td><strong>532</strong></td>
</tr>
</tbody>
</table>
The “Top-N” software faults

<table>
<thead>
<tr>
<th>Fault types</th>
<th>Perc. Observed in field study</th>
<th>ODC classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing &quot;If (cond) { statement(s) }&quot;</td>
<td>9.96 %</td>
<td>Algorithm</td>
</tr>
<tr>
<td>Missing function call</td>
<td>8.64 %</td>
<td>Algorithm</td>
</tr>
<tr>
<td>Missing &quot;AND EXPR&quot; in expression used as branch condition</td>
<td>7.89 %</td>
<td>Checking</td>
</tr>
<tr>
<td>Missing &quot;if (cond)&quot; surrounding statement(s)</td>
<td>4.32 %</td>
<td>Checking</td>
</tr>
<tr>
<td>Missing small and localized part of the algorithm</td>
<td>3.19 %</td>
<td>Algorithm</td>
</tr>
<tr>
<td>Missing variable assignment using an expression</td>
<td>3.00 %</td>
<td>Assignment</td>
</tr>
<tr>
<td>Wrong logical expression used as branch condition</td>
<td>3.00 %</td>
<td>Checking</td>
</tr>
<tr>
<td>Wrong value assigned to a value</td>
<td>2.44 %</td>
<td>Assignment</td>
</tr>
<tr>
<td>Missing variable initialization</td>
<td>2.25 %</td>
<td>Assignment</td>
</tr>
<tr>
<td>Missing variable assignment using a value</td>
<td>2.25 %</td>
<td>Assignment</td>
</tr>
<tr>
<td>Wrong arithmetic expression used in parameter of function call</td>
<td>2.25 %</td>
<td>Interface</td>
</tr>
<tr>
<td>Wrong variable used in parameter of function call</td>
<td>1.50 %</td>
<td>Interface</td>
</tr>
<tr>
<td><strong>Total faults coverage</strong></td>
<td><strong>50.69 %</strong></td>
<td></td>
</tr>
</tbody>
</table>
**G-SWFIT: Generic software fault injection technique**

![Diagram](image)

- **Component executable code**
  - 01011 00010 01001
- **Low-level code fault injection engine**
- **Component with injected faults**
  - 01011 00010 01001
  - 01011 00010 01001
  - 01011 00010 01001
  - 01011 00010 01001
  - 01011 00010 01001

**Library of low level fault injection operators**

**Emulate common programmer mistakes**

The technique can be applied to binary files prior to execution or to in-memory running processes.
Software fault operators

- **Location pattern** – how to locate where to inject a software fault
- **Code change** – what to change in order to inject a software fault
Fault/operator ex 1: Missing and-expression in condition

Target source code (avail. not necessary)

```java
if ( a==3 && b==4 )
{
    do something
}
```

Code with intended fault

```java
if ( a==3 && b==4 )
{
    do something
}
```

Original target code (executable form)

```assembly
cmp dword ptr off_a[ebp],3
jne short ahead
cmp dword ptr off_b[ebp],4
jne short ahead
; ... do something ...
ahead:
...

; remaining prog. code
```

Target code with emulated fault

```assembly
cmp dword ptr off_a[ebp],3
jne short ahead
nop
nop
nop
; ... do something ...
ahead:
...

; remaining prog. code
```

The actual mutation is performed in executable (binary) code. Assembly mnemonics are presented here for readability sake.
Vulnerability Injection (for Web Applications)

- Some possible scenarios:
  - Train security teams
  - Evaluate security teams
  - Estimation of vulnerabilities in the code
  - **Attack Injector**
    - and with an attack injector we can evaluate security mechanisms…
  - …
Field Study on Vulnerabilities

- 6 LAMP web applications
- 655 security vulnerabilities (XSS and SQLi)

Relative distribution

- Missing function call extended
- Wrong variable used in parameter of function call
- Missing if construct plus statements
- Wrong value assigned to variable

Fault types:
- SQL
- XSS
Vulnerability Operators

• **Location pattern** – how to locate where to inject a vulnerability

• **Vulnerability code change** – what to change in order to inject a vulnerability
Vulnerability Injection methodology

1. Analysis of the source code of the web application
2. Search for the locations where a vulnerability may exist
3. Change the code to inject a vulnerability
**Attack Injection Methodology**

- Methodology to test security mechanisms:
  1. Injection of realistic vulnerabilities
  2. Controlled attack of the vulnerabilities

- Stages:
  1. Preparation
  2. Injection of Vulnerabilities
  3. Attack
Simple example

1. Verify the quality of the vulnerabilities and attacks injected

2. Test one IDS for databases
1: Verification of the Injection

- Some vulnerabilities could not be attacked:
  - Multiple protection of the variable
  - Different variables with the same name
The IDS missed five attacks due to a bug in the code of one of its core functions.
Conclusion

• Measuring, assessing and benchmarking security is a big challenge

• Metrics, methods, and tools
  ♦ What is required?
  ♦ Who is measuring and who is going to use the measurements?
    There are a lot of stakeholders…
  ♦ Validation?
  ♦ Trusting on the tools?

• Learn with “old” dependability evaluation techniques could be useful