Centre for Security, Communications and Network Research

Plymouth University

Insider Threat Specification

Techniques for system level detection and prediction of insider threats

George Magklaras PhD

Center for Security Communications and Networks Research - CSCAN

University of Plymouth, UK

http://www.cscan.org

Universitetet i Oslo - AF Security Seminars - 2012

Agenda

- Who is an "insider"?
- Are insider threats a problem?
- Insider Threat Specification for threat mitigation.
- Logging for Insider Threat Specification (LUARM)
- A model for insider threats
- A DSL approach for specifying Insider Threats (ITPSL)
- Current research issues: Forensics, scalability and privacy issues

Insiders (visually)

Definition of an insider

"An insider is a person that has been legitimately empowered with the right to access, represent, or decide about one or more assets of the organization´s infrastructure"

http://www.dagstuhl.de/08302

Insider cases and the press

SCIENTIFIC AMERICAN™

Winner of the 2011 National Magazine Award for General Excellence

Search ScientificAmerican.com

Subscribe | News & Features | Blogs | Multimedia | Education | Citizen Science | Topics

Home » News »

News | Technology

Tweet 0 Like 17

WikiLeaks Breach Highlights Insider Security Threat

Even the toughest security systems sometimes have a soft center that can be exploited by someone who has passed rigorous screening

By Larry Greenemeier and Charles Q. Choi | December 1, 2010 | 4

Share Email Print

The ongoing *WikiLeaks* exposé not only circulated hundreds of thousands of secretive government documents, it has also swiftly prompted changes to the system designed to share access to them. On Tuesday, the U.S. State Department cut off a military computer network's access to its files, dramatically curtailing data sharing intended to help thwart future disasters like the September 11 terrorist attacks.

ENEMY WITHIN: The U.S. government's post-9/11 efforts to increase information sharing among agencies may have left it vulnerable to *WikiLeaks*.
Image: COURTESY OF DAVID MARCHAL, VIA ISTOCKPHOTO.COM

In response to the leaks, the State Department announced it would cut access to its database of embassy cables via the U.S. Defense Department's Secret Internet Protocol Router Network (SIPRNet), a system of dedicated and

THE WALL STREET JOURNAL. | NEW YORK

Europe Edition Home | Today's Paper | Video | Blogs | Emails | Journal Community | Mobile | Tablet

World | Europe | U.K. | U.S. | Business | Markets | Market Data | Tech | Life & Styl

TOP STORIES IN New York

1 of 12

Selling Tudor City Treasure

2 of

New Tensions Involving Police, Occupy Wall Street Protesters

NEW YORK | February 12, 2011

Data Are Stolen From Hospitals

Article | Comments (5)

Email Print Save Like 11 +1 0 Tweet 0 A A

By JOSEPH DE AVILA

The confidential personal health data of about 1.7 million New York City patients, hospital staffers and others were stolen in December, the city's Health and Hospitals Corp. reported Friday.

The medical files, which were stored on magnetic data tapes and extend back as long as 20 years, were stolen on Dec. 23 from an unlocked vehicle belonging to GRM Information Management Services, the city's medical-records vendor based in Jersey City, N.J.

The medical files also included the confidential information of hospital employees, vendors and contractors at Jacobi Medical Center, North Central Bronx Hospital, Tremont Health Center and Gunhill Health Center.

There is no indication that the data have been misused, according to officials with HHC. Accessing the files would require specialized technical expertise, officials said.

"The loss of this data occurred through the negligence of a contracted firm that specializes in the secure transport and storage of sensitive data," Alan D. Aviles, the president of HHC, said in a statement. "HHC is taking responsibility for providing information and credit monitoring services to any affected individual who may be worried about the possibility of identity theft."

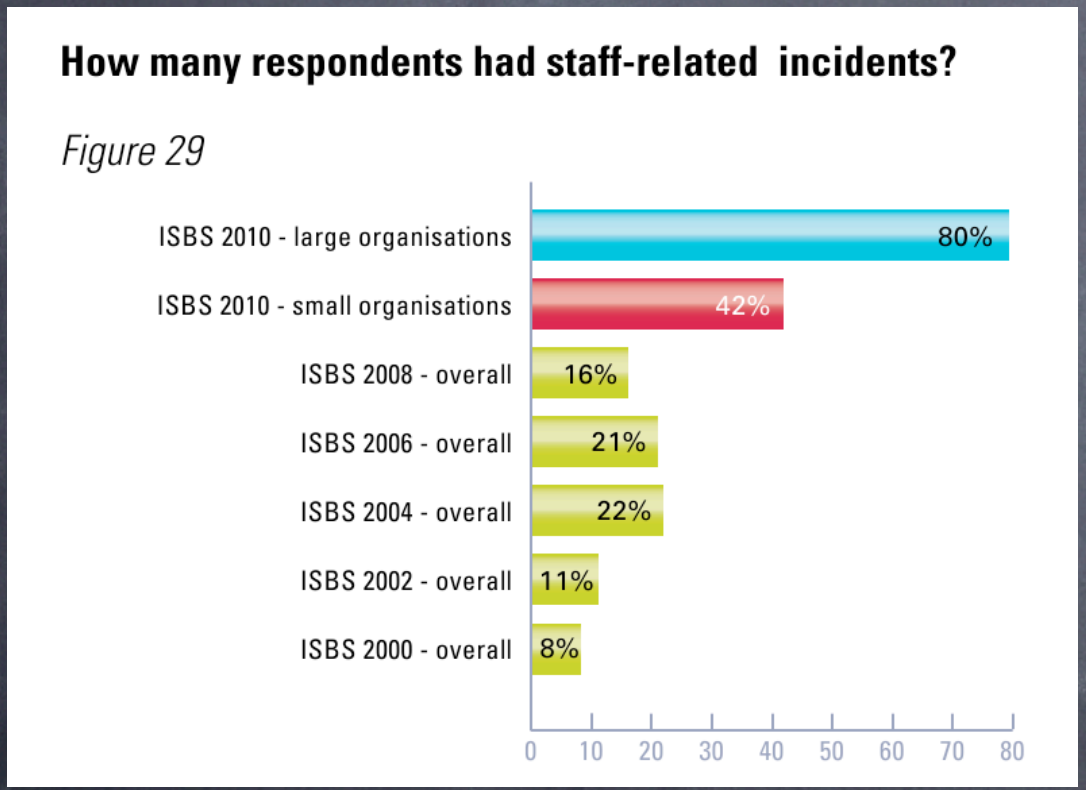
Insider cases in information security surveys

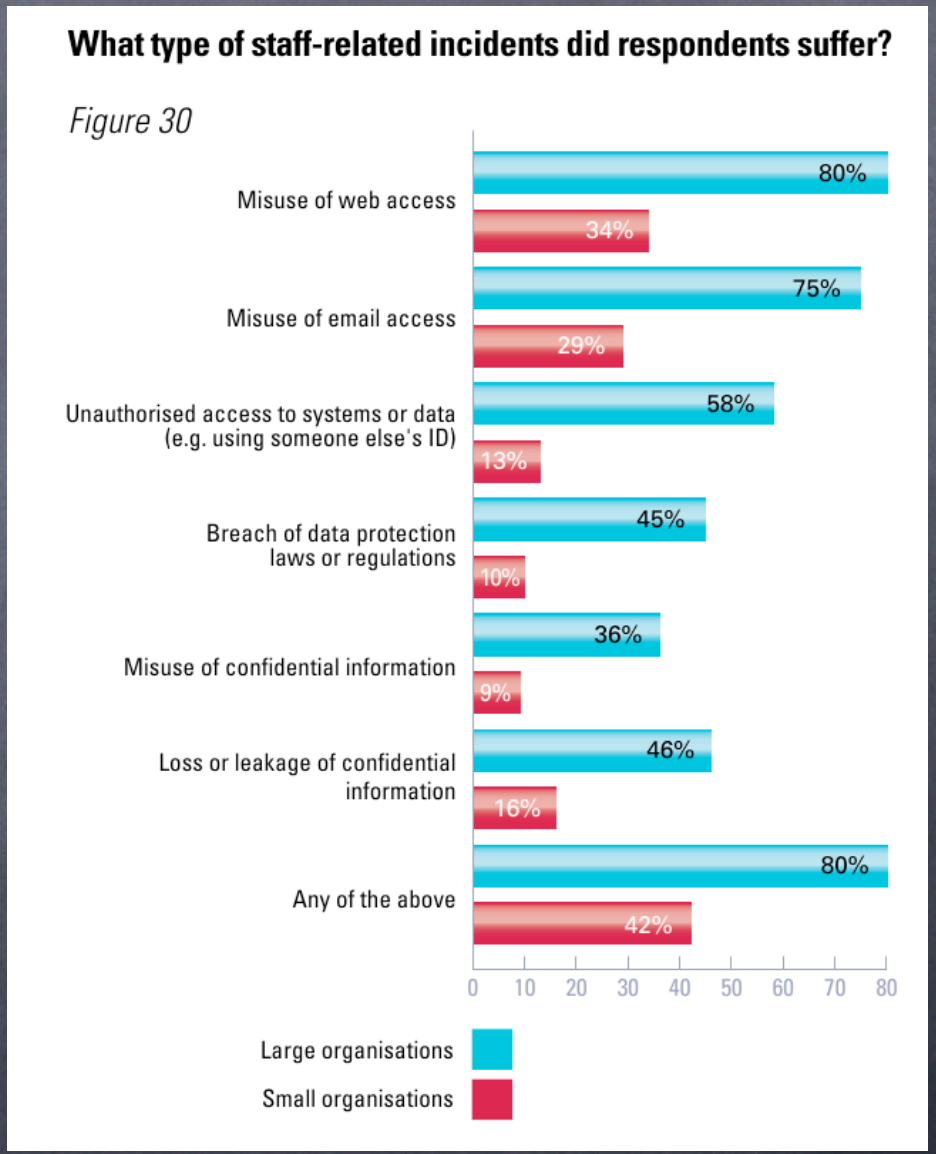
Source:
http://www.infosec.co.uk/files/isbs_2010_technical_report_single_pages.pdf

- "Staff at a London educational institution replied to a phishing email. This resulted in spammers sending over 100,000 emails from the compromised accounts, and to the organization's mail servers being blacklisted around the world."
- "A charity infringed data protection laws when it disposed of an old computer without wiping the hard drive. The staff member concerned was blasé, saying he had deleted the files and trusted the person to whom he had sold the computer."

Quantifying insider misuse manifestation

Source:
http://www.infosec.co.uk/files/isbs_2010_technical_report_single_pages.pdf





Quantifying insider misuse manifestation (2)

Source:
15th Annual Computer Crime and Security Survey
http://gocsi.com/Survey_2010

	None	Up to 20%	21 to 40%	41 to 60%	61 to 80%	81 to 100%
Malicious insider actions	59.1%	28.0%	5.3%	0.8%	3.8%	3.0%
Non-malicious insider actions	39.5%	26.6%	6.5%	8.9%	4.0%	14.5%

-**Intentional misuse**: Insiders with malicious intentions (for example, theft of proprietary information)

-**Accidental misuse**: Insiders that do not intend to do harm (loss of company laptop)

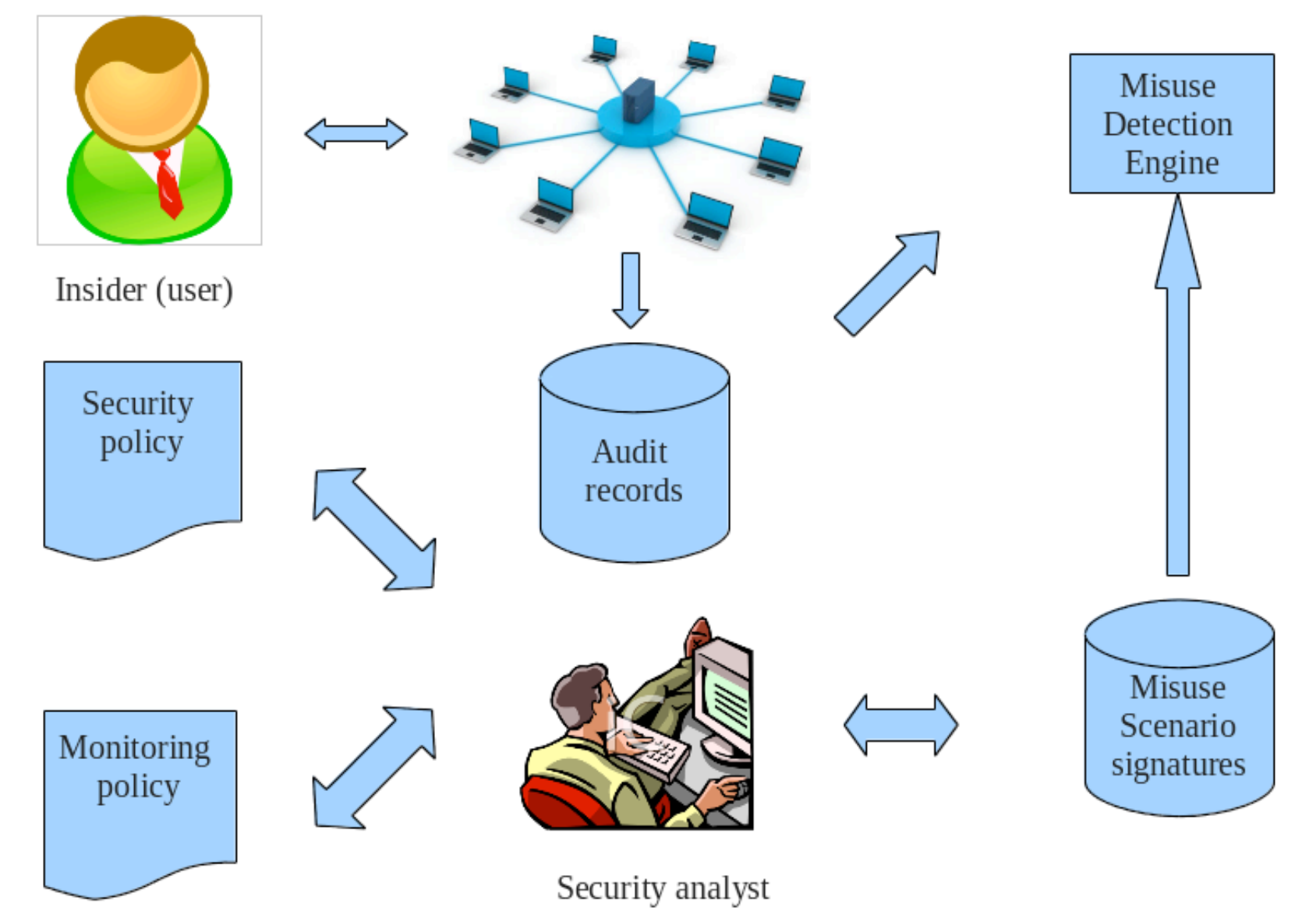
Is the insider threat a real problem?

- Yes certainly.
- Not because the press and the surveys document it.
- Because it is a complex problem.
- Because the infrastructure/tools to systemically collect information about it does not exist.

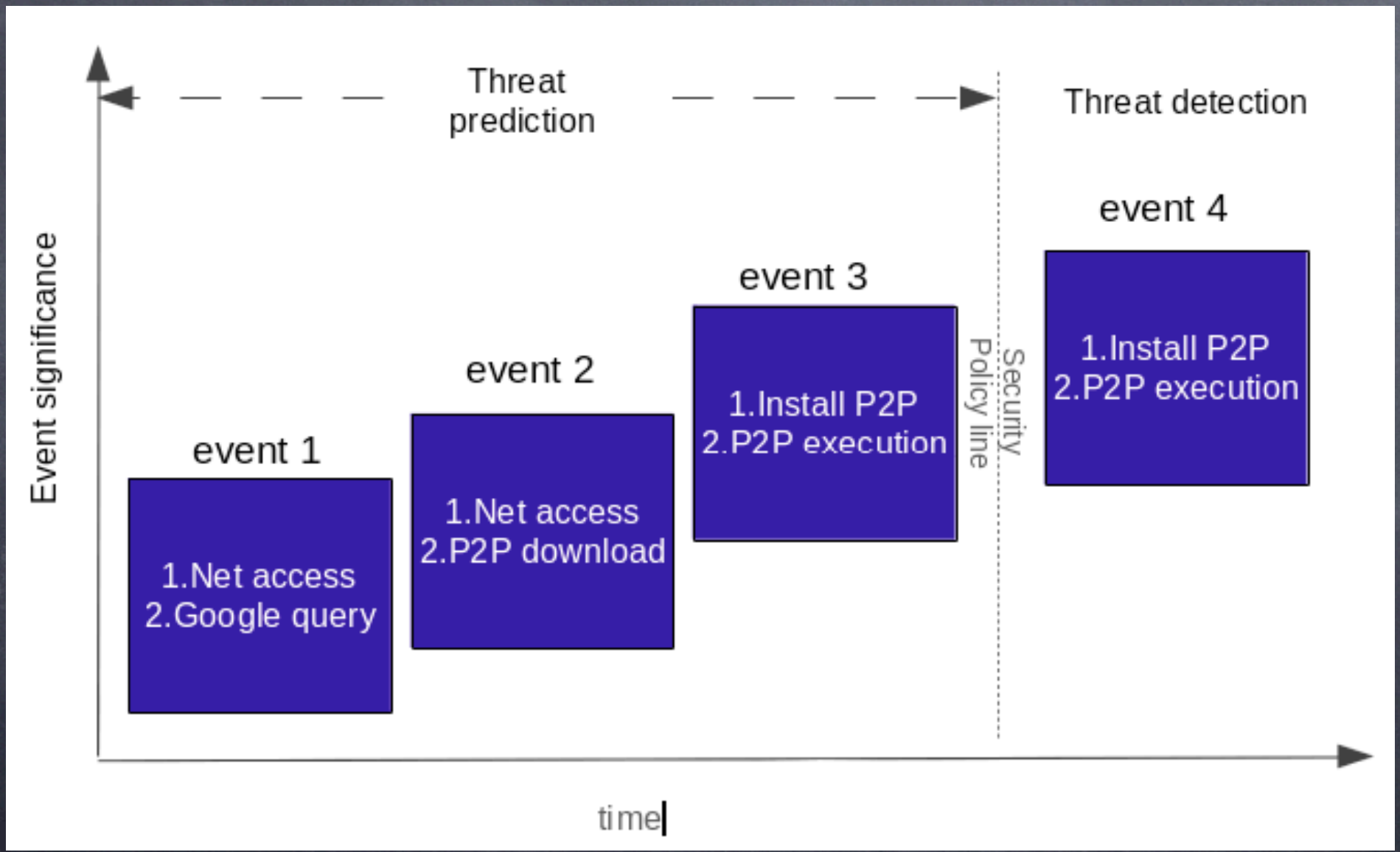
Defining Insider Threat Specification

Insider Threat Specification is the process of using a **standardized vocabulary** to describe in an abstract way how the **aspects** and **behavior** of an insider relate to a security policy defined misuse scenario.

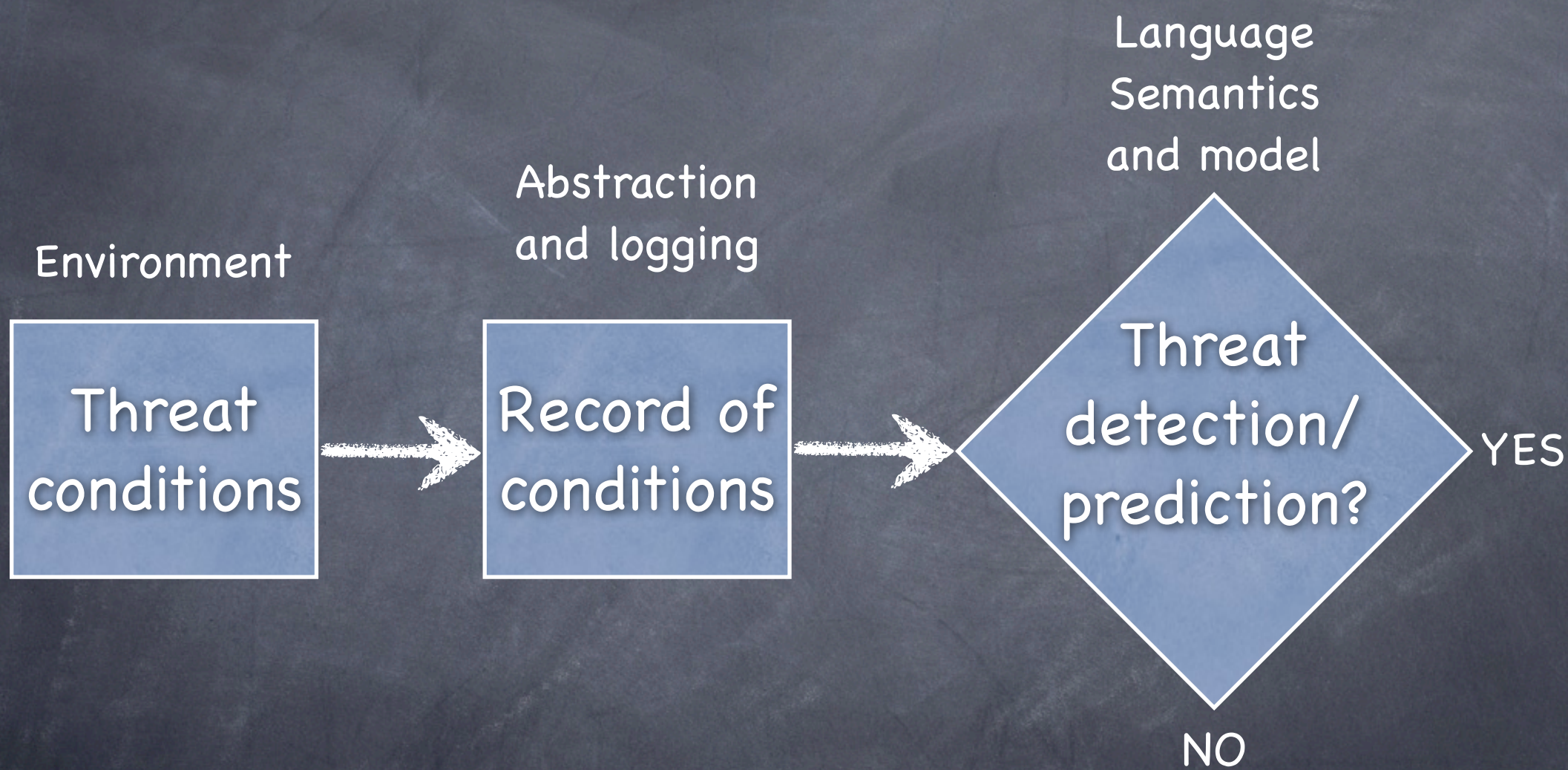
Insider Misuse Detection Information flow



The basis for Insider Threat prediction



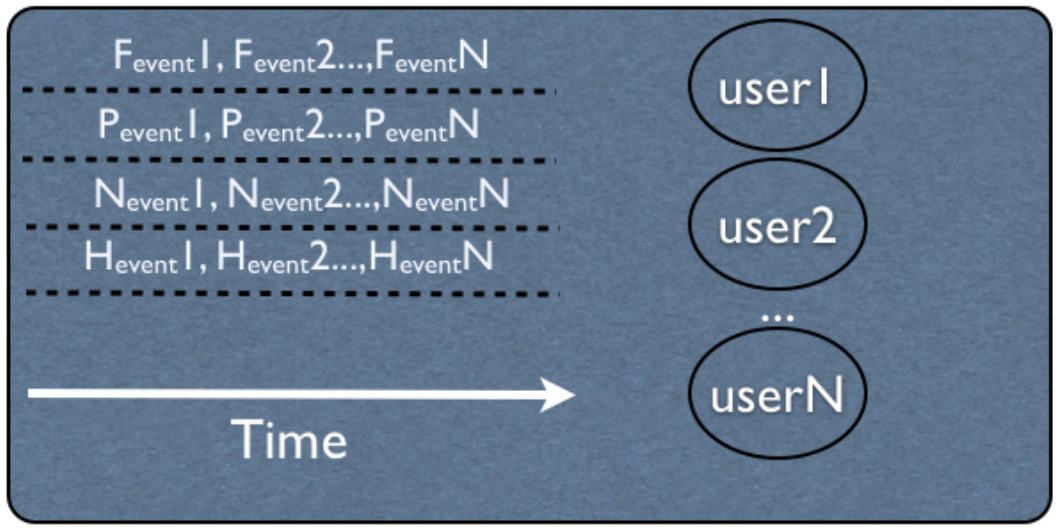
Conceptual Insider Threat mitigation flow



Logging requirements for Insider Threat Specification

- OS agnostic
- Correct timing of records
- Integrity and availability of log data: The "observer effect"
- Provide user entity accountability
- Accommodate static and dynamic (volatile) data

Logging requirements for Insider Threat Specification (2)



Computer system

F_{event} = File Event (read,access,copy,move, erase)

P_{event} = Process execution event (process start/finish)

N_{event} = Network endpoint and route event (creation, deletion)

H_{event} = Hardware device event (attachment, detachment)

"User x was able to launch process b at 16:48:32 which resulted in two connections to websites A and B and as a result left file loic.pro at 16:52:21 in

user's x Document area"

Logging requirements for Insider Threat Specification (3)

Sample of existing logging/audit engines:
-Syslogd, WinSyslog, RFC 5424
-OpenXDAS, Cisco MARS
-Event Data Warehouse, Arc Sight Logger 4

Most of these solutions are geared towards network and application security events and/or data audit compliance.

They do not meet all of the previous requirements.

Logging requirements for Insider Threat Specification (4)

```
File  Edit  View  Search  Terminal  Help
[georgios@slartibartfast Volatility-1.3_Beta]$ python volatility connections -f xp-laptop-2005-07-04-1430.img
Local Address             Remote Address            Pid
127.0.0.1:1037            127.0.0.1:1038            3276
127.0.0.1:1038            127.0.0.1:1037            3276
[georgios@slartibartfast Volatility-1.3_Beta]$ python volatility pslist -f xp-laptop-2005-07-04-1430.img | grep 3276
firefox.exe          3276   2392   7      189    Mon Jul 04 18:21:11 2005
[georgios@slartibartfast Volatility-1.3_Beta]$ python volatility files -f xp-laptop-2005-07-04-1430.img | grep -5 3276
**********************************************************************
Pid: 3256
File   \dd\UnicodeRelease
File   \WINDOWS\WinSxS\x86_Microsoft.Windows.Common-Controls_6595b64144ccf1df_6.0.2600.2180_x-ww_a84f1ff9
**********************************************************************
Pid: 3276
File   \WINDOWS\WinSxS\x86_Microsoft.Windows.Common-Controls_6595b64144ccf1df_6.0.2600.2180_x-ww_a84f1ff9
File   \WINDOWS\WinSxS\x86_Microsoft.Windows.Common-Controls_6595b64144ccf1df_6.0.2600.2180_x-ww_a84f1ff9
File   \Documents and Settings\Sarah\Application Data\Mozilla\Firefox\Profiles\z5vogzjr.default\parent.lock
File   \Endpoint
File   \AsyncConnectHlp
[georgios@slartibartfast Volatility-1.3_Beta]$
```

Volatile data versus a collection of time-ordered volatile data.

Insider Threat Specification Logging

http://luarm.sourceforge.net

- Logging User Actions in Relational Mode - LUARM
- Prototype Insider Threat Specification logging engine to:
  - Satisfy the previously mentioned requirements.
  - Allow researchers to replay/study insider incidents
  - Insider logging forensic capability

LUARM publication

LUARM: An Audit Engine for Insider Misuse Detection

View Sample

0

Author(s): G. Magklaras (University of Plymouth, UK), S. Furnell (University of Plymouth, UK) and M. Papadaki (University of Plymouth, UK)

Volume: 3
Issue: 3
Pages: 13
Source title: International Journal of Digital Crime and Forensics (IJDCF)
Editor(s)-in-Chief: Chang-Tsun Li (University of Warwick, UK) and Anthony T.S. Ho (University of Surrey, UK)
DOI: 10.4018/jdcf.2011070103
ISSN: 1941-6210
EISSN: 1941-6229
Keywords: Digital Crime & Forensics / Information Science Reference / IT Security/Ethics / Security Technologies, Ethics & Law

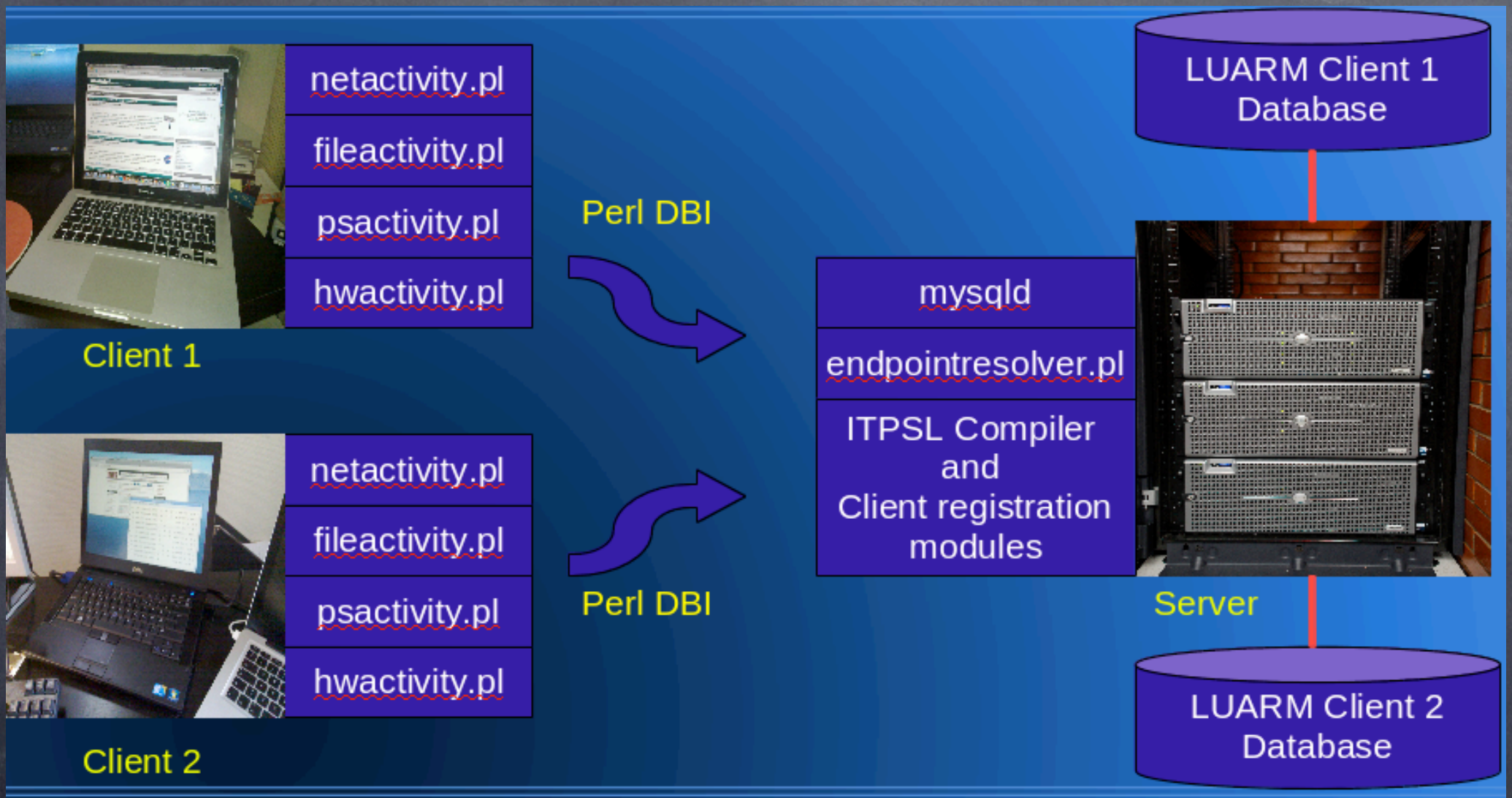
Purchase

View LUARM: An Audit Engine for Insider Misuse Detection on the publisher's website for pricing and purchasing information.

Abstract

Logging User Actions in Relational Mode (LUARM) is an open source audit engine for Linux. It provides a near real-time snapshot of a number of user action data such as file access, program execution and network endpoint user activities, all organized in easily searchable relational tables. LUARM attempts to solve two fundamental problems of the insider IT misuse domain. The first concerns the lack of insider misuse case data repositories that could be used by post-case forensic examiners to aid an incident investigation. The second problem relates to how information security researchers can enhance their ability to specify accurately insider threats at system level. This paper presents LUARM's design perspectives and a 'post mortem' case study of an insider IT misuse incident. The results show that the prototype audit engine has good potential to provide a valuable insight into the way insider IT misuse incidents manifest on IT systems and can be a valuable complement to forensic investigators of IT misuse incidents.

Insider Threat Specification Logging (2)



LUARM architecture

Insider Threat Specification Logging (3)

fileaccessid	bigint
md5sum	text
filename	varchar
location	varchar
username	tinytext
application	text
fd	tinytext
pid	int
size	bigint
cyear	int
cmonth	tinyint
cday	tinyint
chour	tinyint
cmin	tintint
csec	tinytint
dyear	int
dmonth	tinyint
dday	tinyint
dhour	tinyint
dmin	tinyint
dsec	tinyint

fileops

psentity	bigint
md5sum	text
username	tinytext
pid	smallint
ppid	smallint
pcpu	decimal
pmem	decimal
command	text
arguments	mediumtext
cyear	int
cmonth	tinyint
cday	tinyint
chour	tinyint
cmin	tinyint
csec	tinytint
dyear	int
dmonth	tinyint
dday	tinyint
dhour	tinyint
dmin	tinyint
dsec	Tinyint
usermame	tinytext
pid	int

procops

Text

endpointinfo	bigint
md5sum	text
transport	tinytext
sourceip	tinytext
sourcefqdn	tinytext
destip	tinytext
destfqdn	tinytext
sourceport	smallint
destport	smallint
ipversion	smallint
cyear	int
cmonth	tinyint
cday	tinyint
chour	tinyint
cmin	tinyint
csec	tinytint
dyear	int
dmonth	tinyint
dday	tinyint
dhour	tinyint
dmin	tinyint
dsec	Tinyint
usermame	tinytext
pid	int
application	text

netops

hwdevd	bigint
md5sum	text
devbus	tinytext
devstring	text
devvendor	text
application	text
userslogged	text
cyear	int
cmonth	tinyint
cday	tinyint
chour	tinyint
cmin	tinyint
csec	tinytint
dyear	int
dmonth	tinyint
dday	tinyint
dhour	tinyint
dmin	tinyint
dsec	tinyint

hardwareops

LUARM query examples

- Find all accesses of the file 'prototype.ppt' by users 'toms' OR 'georgem' between 9:00 and 14:00 hours on 23/10/2009.
  - SELECT * FROM fileinfo WHERE filename='prototype.ppt' AND ((username='toms') OR (username='georgem')) AND cyear='2009' AND cmonth='10' AND cday='23' AND chour >= '9' AND chour <= '13' AND cmin >= '0' AND cmin >= '59';
- Find all USB devices that were physically connected to the system when users 'toms' OR 'georgem' were logged on 23/10/2009.
  - SELECT * from hwinfo WHERE devbus='usb' AND ((userslogged RLIKE 'toms') OR (userslogged RLIKE 'georgem')) AND cyear='2009' AND cmonth='10' AND cday='23' AND chour >= '9' AND chour <= '13' AND cmin >= '0' AND cmin >= '59';

The Insider Threat Model

Computers & Security

Volume 21, Issue 1, 1st Quarter 2001, Pages 62–73

Events

Insider Threat Prediction Tool: Evaluating the probability of IT misuse

G.B Magklaras, S.M Furnell

Network Research Group, Department of Communication and Electronic Engineering, University of Plymouth, UK

Available online 2 February 2002.

http://dx.doi.org/10.1016/S0167-4048(02)00109-8, How to Cite or Link Using DOI

Describes the taxonomy of insider misuse and the threat evaluation process.

The Insider Threat Model (2)

Computers & Security

Volume 24, Issue 5, August 2005, Pages 371–380

A preliminary model of end user sophistication for insider threat prediction in IT systems

G.B. Magklaras [Author Vitae], S.M. Furnell [Author Vitae]

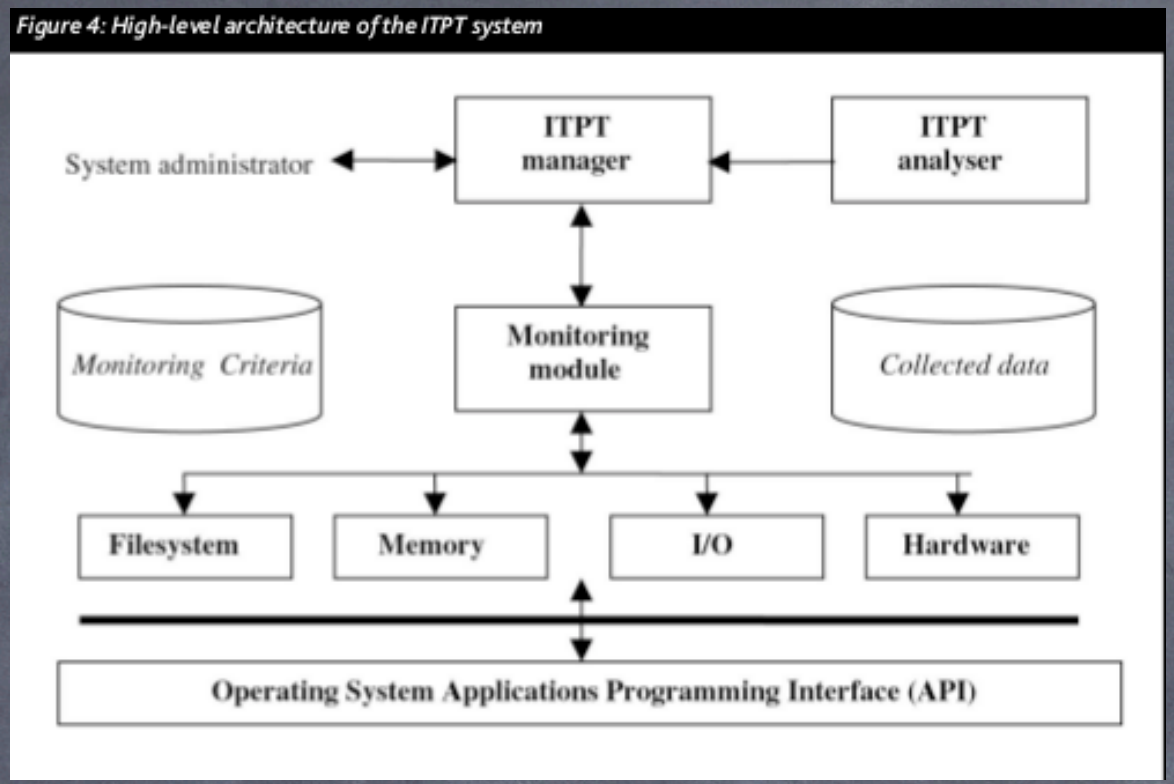
Network Research Group, School of Computing, Communications and Electronics, University of Plymouth, Plymouth, United Kingdom

Received 26 April 2004. Revised 7 October 2004. Accepted 11 October 2004. Available online 16 December 2004.

http://dx.doi.org/10.1016/j.cose.2004.10.003, How to Cite or Link Using DOI

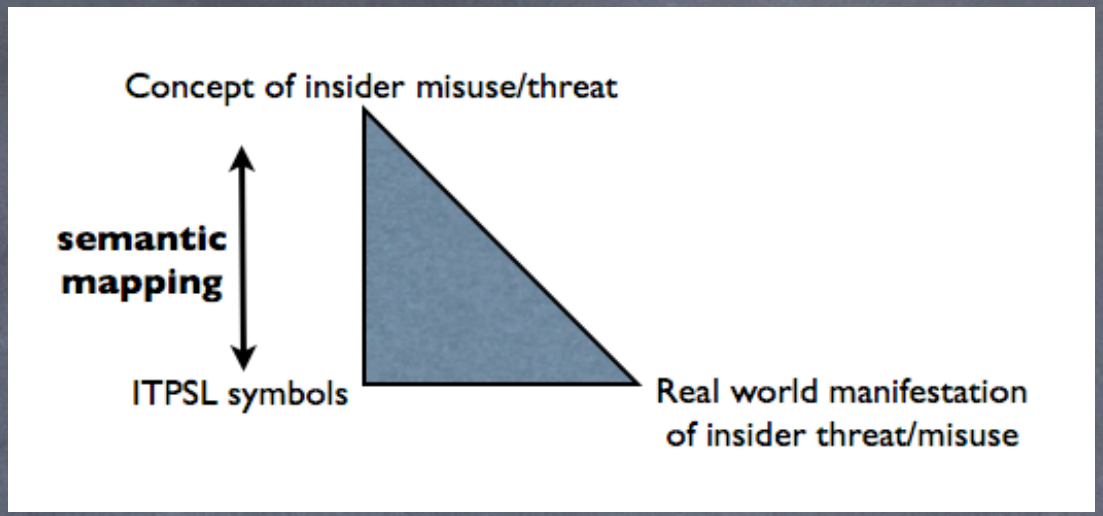
Describes how one can measure user sophistication as a threat metric.

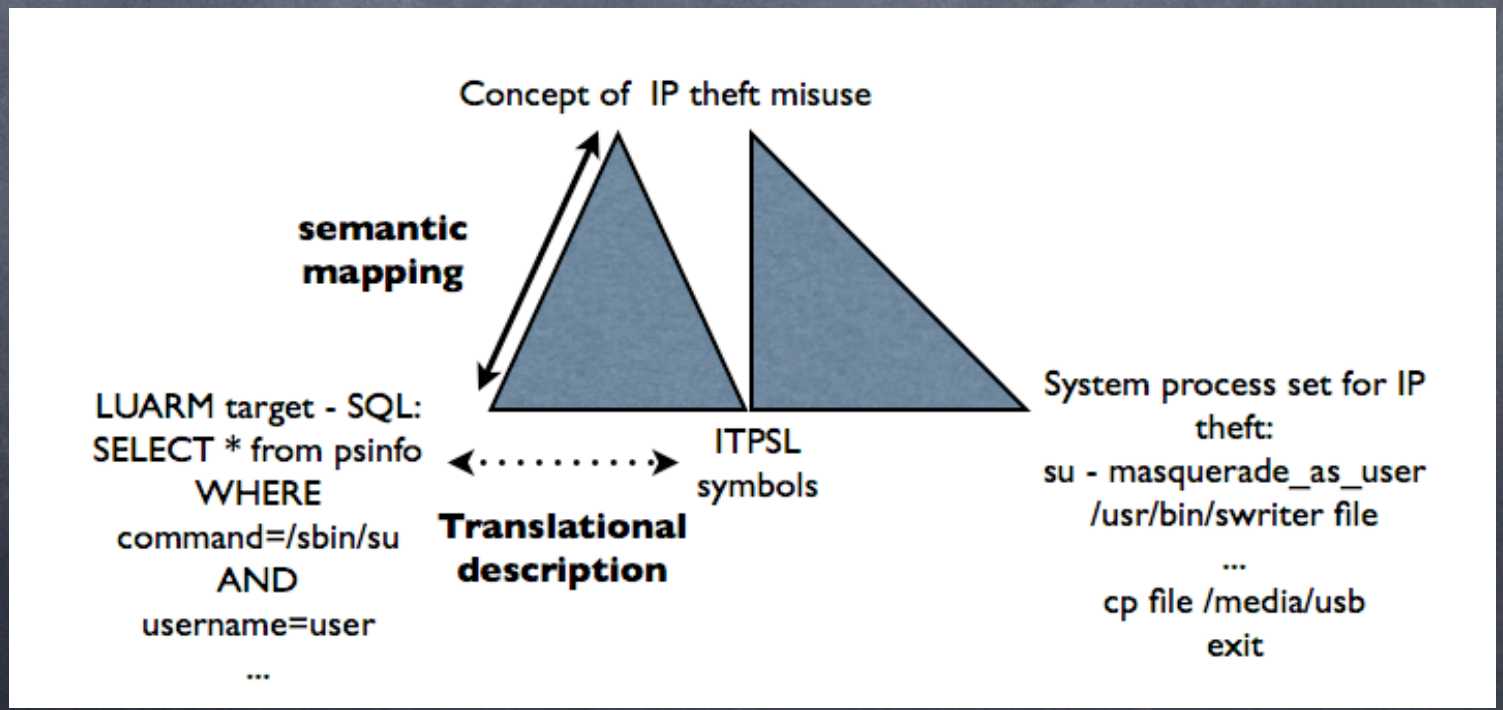
The Insider Threat Model (3)



$$\mathrm{EPT} = \sum F_{\text{threat components}} \Rightarrow \mathrm{EPT} = F_{\text{accessrights}} + F_{\text{behavior}} \Rightarrow \mathrm{EPT} = C_{\text{role}} + C_{\text{criticalfiles}} + C_{\text{hardware}} + C_{\text{utilities}} + C_{\text{sysadm}} + F_{\text{behavior}} \Rightarrow \mathrm{EPT} = C_{\text{role}} + C_{\text{criticalfiles}} + C_{\text{hardware}} + C_{\text{utilities}} + C_{\text{sysadm}} + F_{\text{sophistication}} + F_{\text{fileops}} + F_{\text{execops}} + F_{\text{network}}$$

From LUARM data to a language





ITPSL publications

Towards an insider threat prediction specification language

Document Information:

Title:	Towards an insider threat prediction specification language
Author(s):	G.B. Magklaras, (Network Research Group, School of Computing, Communications and Electronics, University of Plymouth, Plymouth, UK), S.M. Furnell, (Network Research Group, School of Computing, Communications and Electronics, University of Plymouth, Plymouth, UK), P.J. Brooke, (School of Computing, University of Teesside, Middlesbrough, UK)
Citation:	G.B. Magklaras, S.M. Furnell, P.J. Brooke, (2006) "Towards an insider threat prediction specification language", Information Management & Computer Security, Vol. 14 Iss: 4, pp.361 - 381
Keywords:	Data security, Information systems
Article type:	Conceptual paper
DOI:	10.1108/09685220610690826 (Permanent URL)
Publisher:	Emerald Group Publishing Limited

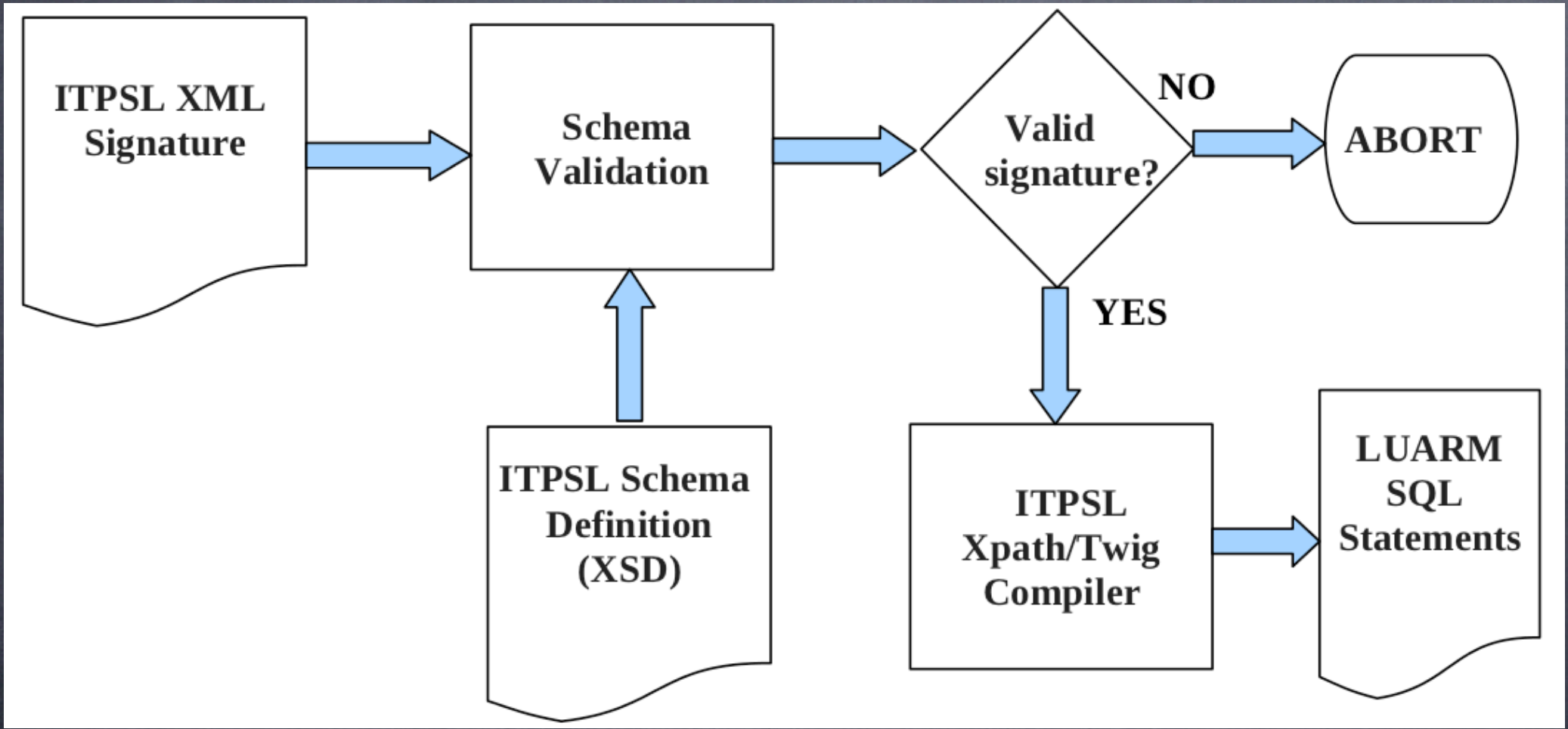
INC2012
11-12 July, Port Elizabeth, South Africa

Magklaras G., Furnell S. (2012), "The Insider Threat Prediction and Specification Language", Ninth International Network Conference, 11-12 July, Port Elizabeth, South Africa.

High level language requirements

- Descriptive power for insider misuse detection and prediction
- Machine and human readable form
- LUARM audit record <-> Language semantics
- Focused on the domain - Domain Specific Language - DSL
- Should facilitate the creation of threat scenario repositories/ontologies.

Insider Threat Prediction and Specification Language (ITPSL)



Signature header with insider scenario ontology

Main body that describes the elements of the scenario/threat

```
<itpslsig>
<itpslheader>
        <signid> <md5sum of date and second, type of OS, current number of processes>
        </signid>
        <signdate>
                <year> dddd </year>
                <month> dd </month>
                <day> dd </day>
        </signdate>
        <ontology>
                <reason> "intentional" | "accidental" </reason>
                <revision> d.d </revision>
                <user_role> "admins" | "advanced_users" | "ordinary_users" </user_role>
                <detectby>  "file" | "exec" | "network" | "hardware" | "multi" </detectby>
                <multihost> yes | no </multihost>
                <hostlist> host1,hostgroup1,...hostn,hostgroupn </hostlist>
                <weightmatrix>n_events, w_event1,w_event2,...,w_eventn </weightmatrix>
                <os> "linux" | "windows" | "macosx" | "unix" </os>
                <osver> "2.4" | "2.6" |  "2000" | "Vista" | "7" </osver>
                <threatkeywords> keyword1 keyword2 ... keyword5
                 </threatkeywords>
                [ <synopsis> "text that describes the signature's purpose and function"
                </synopsis>]
        </ontology>
</itpslheader>
<itpslbody>
        <mainblock>
                <mainop> "AND"|"OR"|"XOR"|"as_a_result_of" | "justone"</mainop>
                <subblock>
                        <subop> "AND"|"OR"|"XOR"|"as_a_result_of"| "single" </subop>
                        <filestatements> ....</filestatements>
                        <execstatements>....</execstatements>
                        <netstatements>...</netstatements>
                </subblock>
                <subblock>
                        <subop> "AND"|"OR"|"XOR"|"as_a_result_of"| "single" </subop>
                        <filestatements> ....</filestatements>
                        <execstatements>....</execstatements>
                        <netstatements>...</netstatements>
                </subblock>
        </mainblock>
</itpslbody>
</itpslsig>
```

Example 1: Pornographic access **detection** scenario

```xml
<?xml version="1.0"?>
<itpslsig xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
        <itpslheader>
                <signid> 4938724b6b41a834ac695529dd298ed0 </signid>
                        <signdate>
                                <year>2011</year>
                                <month>1</month>
                                <day>20</day>
                        </signdate>
                <ontology>
                        <reason>intentional</reason>
                        <revision>1.0</revision>
                        <user_role>ordinary_users</user_role>
                        <detectby>file</detectby>
                        <multihost>no</multihost>
                        <hostlist>cn1</hostlist>
                        <weightmatrix> 0 </weightmatrix>
                        <os>linux</os>
                        <osver>2.6</osver>
                        <keywords>pornography xxx adult web browser</keywords>
                        <synopsis> This signature locates users that use the web browser to
                        connect to certain pornographic websites
                        </synopsis>
                </ontology>
        </itpslheader>
        <itpslbody>
                <mainblock>
                        <mainop>justone</mainop>
                                <subblock>
                                        <subop>single</subop>
                                        <fileexists>
                                                <filename>places.sqlite</filename>
                                                <type>any</type>
                                                <location>userhome/.mozilla/</location>
                                                <singlefile>yes</singlefile>
                                                <withcontents>
                                                        <stringsearch> "mybadsite1.com" OR
                                                        "mybadsite2.com" </stringsearch>
                                                </withcontents>
                                        </fileexists>
                                </subblock>
                </mainblock>
        </itpslbody>
</itpslsig>
```

Example 2: ITPSL header for threat prediction

```xml
<?xml version="1.0"?>
<itpslsig xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
        <itpslheader>
                <signid> 5938724b6b41a834ac695529dd104ed0 </signid>
                        <signdate>
                                <year>2010</year>
                                <month>12</month>
                                <day>20</day>
                        </signdate>
                <ontology>
                        <reason>intentional</reason>
                        <revision>1.0</revision>
                        <user_role>ordinary_users</user_role>
                        <detectby>multi</detectby>
                        <multihost>no</multihost>
                        <hostlist>proteas,dionisos,slart,cn1,panoptis</hostlist>
                        <weightmatrix>3,10,20,60</weightmatrix>
                        <os>linux</os>
                        <osver>2.6</osver>
                        <keywords>DoS software install DoS loiq </keywords>
                        <synopsis> This signature predicts the usage of the Low Orbit Ion Cannon tool for DDoS attacks.
                        </synopsis>
                </ontology>
        </itpslheader>
```

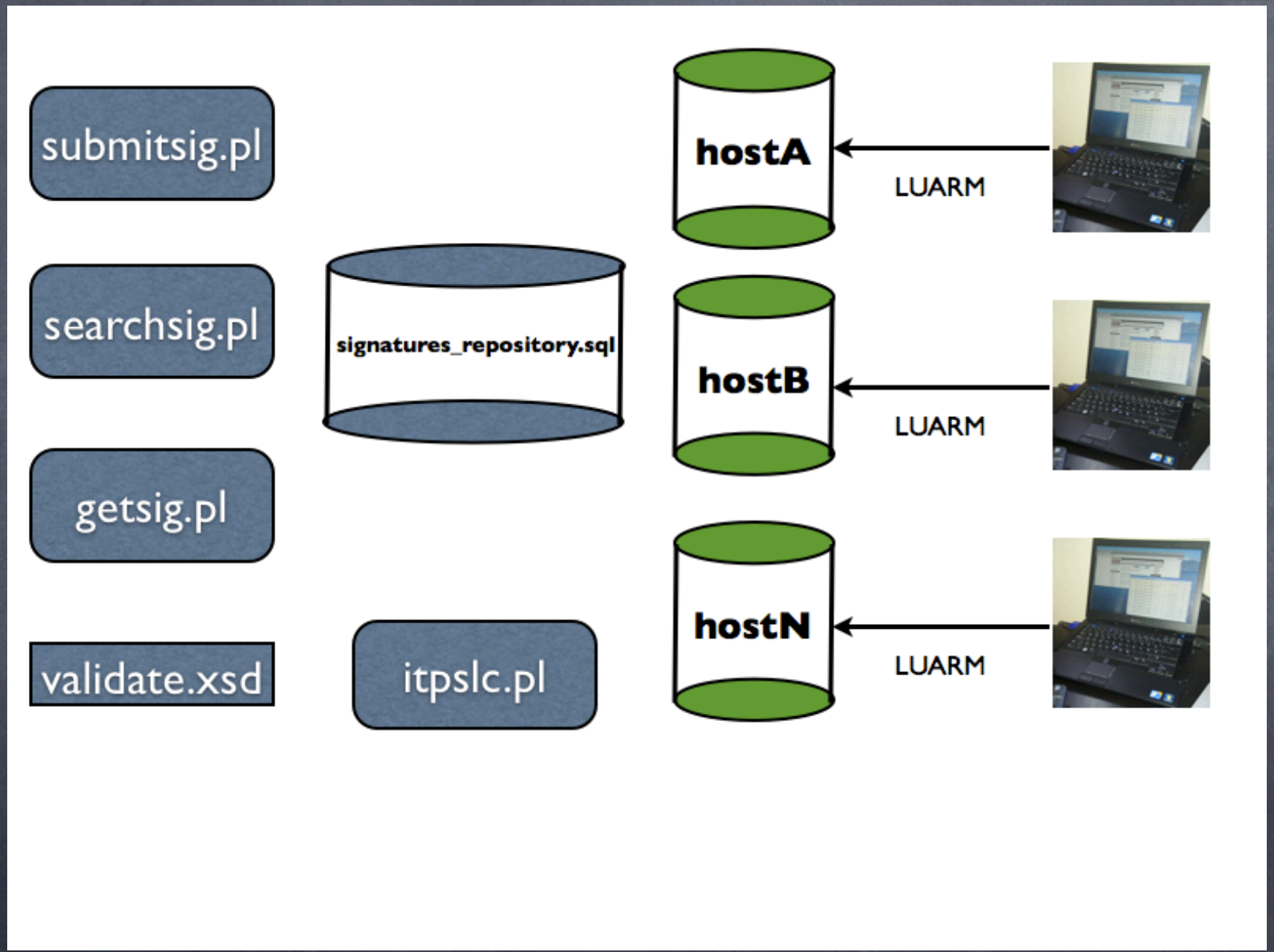
$$\sum wevent_n = EPMO$$

EPMO -> Evaluated Potential Misuse Occurrence (0...1)
n-> number of specified events
<weightmatrix>nevents, $wevent_1$,$wevent_2$,...,$wevent_n$ </weightmatrix>

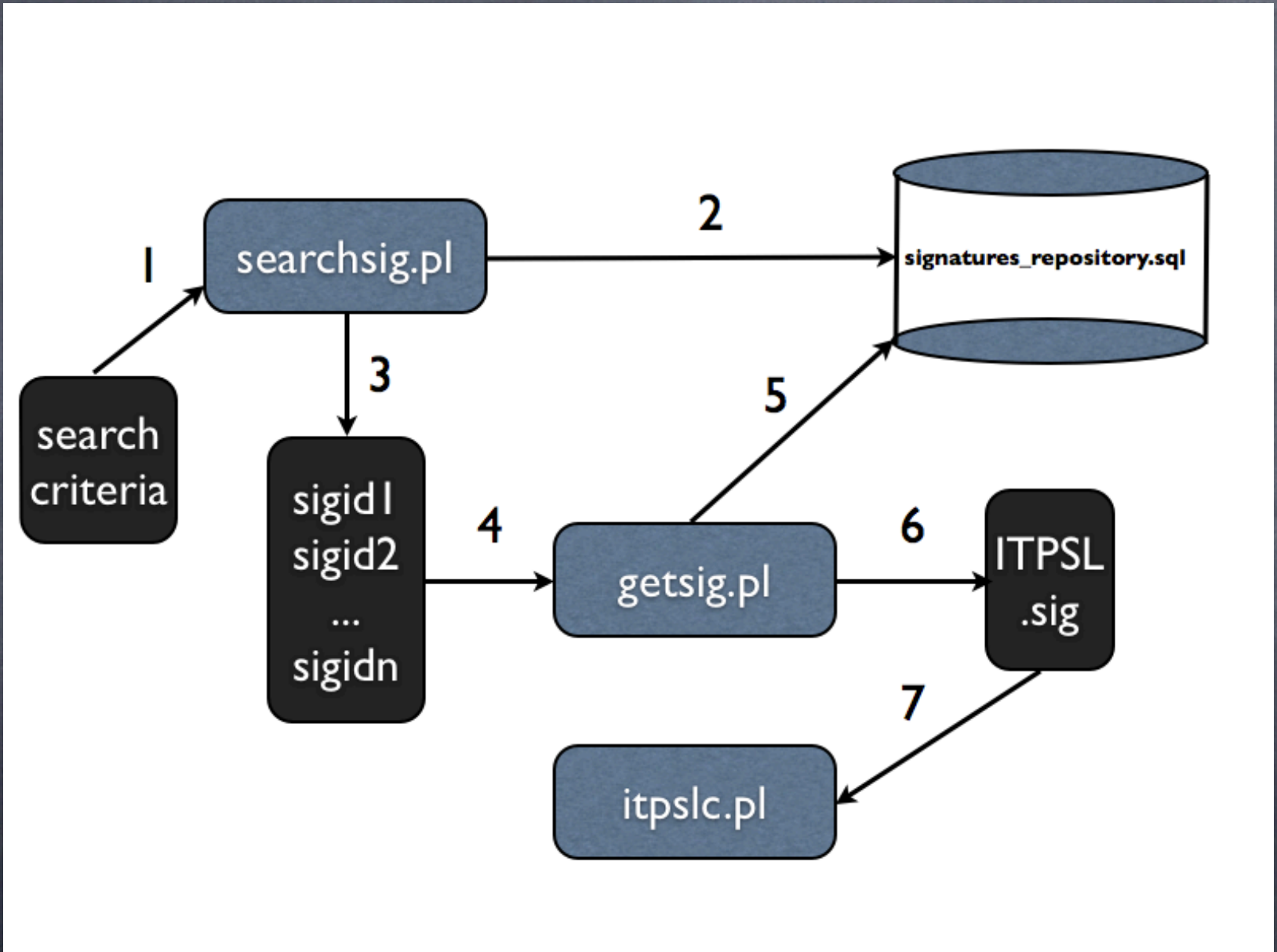
Example 2: DDoS attack initiation **prediction** scenario

```
        <itpslbody>
                <mainblock>
                        <mainop>as_a_result_of</mainop>
                                <subblock>
                                        <subop>AND</subop>
                                        <fileexists>
                                                <filename>loiq</filename>
                                                <type>executable</type>
                                                <location>OR (#userhome#/*,/site/*,/tmp/*,/temp/*)</location>
                                                <singlefile>yes</singlefile>
                                                <ownedbyuser>johnc</ownedbyuser>
                                        </fileexists>
                                        <fileexists>
                                                <filename>loiq.pro</filename>
                                                <type>textdata</type>
                                                <location>OR(#userhome#/*,/site/*,/tmp/*,/temp/*)</location>
                                                <ownedbyuser>johnc</ownedbyuser>
                                                <singlefile>yes</singlefile>
                                        </fileexists>
                                        <fileexists>
                                                <filename>loiq.qrc</filename>
                                                <type>textdata</type>
                                                <location>OR(#userhome#/*,site/*,/tmp/*,/temp/*)</location>
                                                <singlefile>yes</singlefile>
                                                <ownedbyuser>johnc</ownedbyuser>
                                        </fileexists>
                                </subblock>
                                <subblock>
                                        <subop>single</subop>
                                        <userexec>
                                                <username>johnc</username>
                                                <name>OR (file-roller,tar,bunzip2)</name>
                                                <path>OR(/usr/bin/,/usr/local/bin)</path>
                                                <singleprocess>yes</singleprocess>
                                                <argumentlist>loiq*.bz2</argumentlist>
                                                <pattern>any</pattern>
                                        </userexec>
                                </subblock>
                                <subblock>
                                        <subop>single</subop>
                                        <fileexists>
                                                <filename>*</filename>
                                                <type>any</type>
                                                <location>OR (#userhome#/.mozilla/*,#userhome#/.opera)</location>
                                                <singlefile>yes</singlefile>
                                                <withcontents>
                                                        <stringsearch>"http://sourceforge.net/projects/loiq"</stringsearch>
                                                </withcontents>
                                                <ownedbyuser>johnc</ownedbyuser>
                                        </fileexists>
                                </subblock>
                </mainblock>
        </itpslbody>
</itpslsig>
```

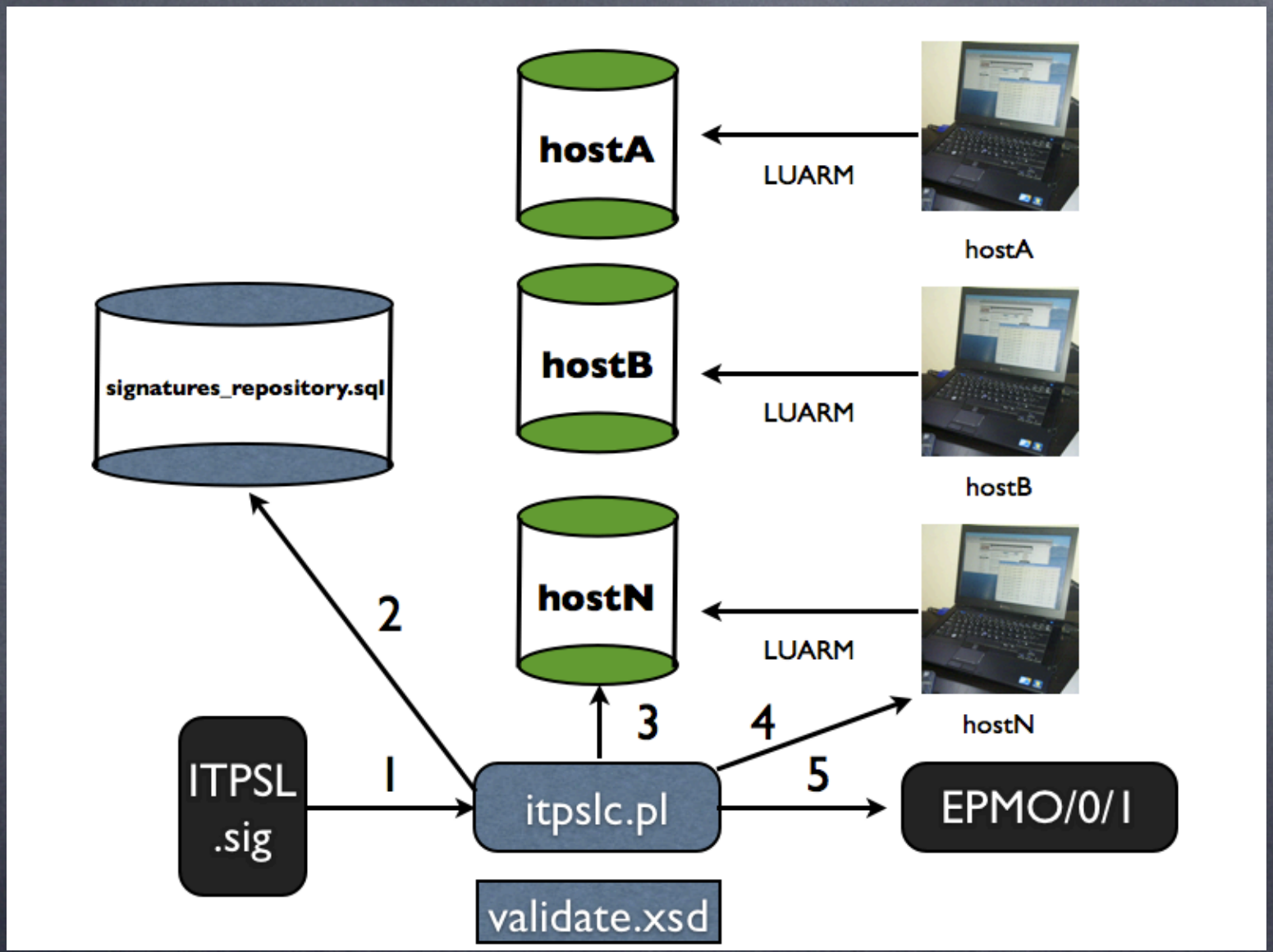
The ITPSL compiler



Performing an ITPSL ontology search



Running an ITPSL signature to the compiler



Overview of achievements:

- LUARM: Have been used in controlled experiments and in the real world. Installed base to date: 350 users.
- ITPSL: In constant development
- LUARM: Has successfully resolved more than 3000 cases of insider misuse: accidental and intentional.

Current and future research issues

- Forensics: I detected/predicted something in a reliable manner. Will it stand in a Court of Law?
- Privacy: How do I ensure I comply with the Law and protect the misuse of LUARM data?
- Scalability: Hundreds of hosts? Feasible. Thousands/millions?

Questions and references

georgios.magklaras@plymouth.ac.uk

http://folk.uio.no/georgios
http://luarm.sourceforge.net