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
Agenda

غا Who is an “insider”?  
غا Are insider threats a problem?  
غا Insider Threat Specification for threat mitigation.  
غا Logging for Insider Threat Specification (LUARM)  
غا A model for insider threats  
غا A DSL approach for specifying Insider Threats (ITPSL)  
غا 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
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

The ongoing WikiLeaks expose 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.

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

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 MARCHEL, VITAISTOCKPHOTO.COM

New Tensions Involving Police, Occupy Wall Street Protesters

Data Are Stolen From Hospitals

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 extended 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 Gun Hill 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:

“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.”
How many respondents had staff-related incidents?

**Figure 29**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Misuse of web access</td>
<td>80%</td>
<td>42%</td>
<td>16%</td>
<td>21%</td>
<td>22%</td>
<td>11%</td>
<td>8%</td>
</tr>
<tr>
<td>Misuse of email access</td>
<td>34%</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
</tr>
<tr>
<td>Unauthorised access to systems or data (e.g., using someone else's ID)</td>
<td>75%</td>
<td>58%</td>
<td>58%</td>
<td>58%</td>
<td>58%</td>
<td>58%</td>
<td>58%</td>
</tr>
<tr>
<td>Breach of data protection laws or regulations</td>
<td>45%</td>
<td>45%</td>
<td>45%</td>
<td>45%</td>
<td>45%</td>
<td>45%</td>
<td>45%</td>
</tr>
<tr>
<td>Misuse of confidential information</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
</tr>
<tr>
<td>Loss or leakage of confidential information</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
<td>46%</td>
</tr>
<tr>
<td>Any of the above</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Quantifying insider misuse manifestation (2)

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

<table>
<thead>
<tr>
<th></th>
<th>None</th>
<th>Up to 20%</th>
<th>21 to 40%</th>
<th>41 to 60%</th>
<th>61 to 80%</th>
<th>81 to 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malicious insider actions</td>
<td>59.1%</td>
<td>28.0%</td>
<td>5.3%</td>
<td>0.8%</td>
<td>3.8%</td>
<td>3.0%</td>
</tr>
<tr>
<td>Non-malicious insider actions</td>
<td>39.5%</td>
<td>26.6%</td>
<td>6.5%</td>
<td>8.9%</td>
<td>4.0%</td>
<td>14.5%</td>
</tr>
</tbody>
</table>

- **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

Insider (user) → Security policy → Audit records → Monitoring policy → Security analyst → Misuse Detection Engine → Misuse Scenario signatures
The basis for Insider Threat prediction

Event 1:
1. Net access
2. Google query

Event 2:
1. Net access
2. P2P download

Event 3:
1. Install P2P
2. P2P execution

Event 4:
1. Install P2P
2. P2P execution

Time

Event significance

Threat prediction

Threat detection

Security policy line
Conceptual Insider Threat mitigation flow

Environment

Threat conditions

Abstraction and logging

Record of conditions

Language Semantics and model

Threat detection/prediction?

YES

NO
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
“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)

Volatile data versus a collection of time-ordered volatile data.
Insider Threat Specification Logging

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

http://luarm.sourceforge.net

Universitetet i Oslo - AF Security Seminars - 2012

Saturday, April 21, 2012
LUARM: An Audit Engine for Insider Misuse Detection

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

Copyright: 2011
Volume: 3
Issue: 3
Pages: 13


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

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.
LUARM architecture

Universitetet i Oslo - AF Security Seminars - 2012

Saturday, April 21, 2012
<table>
<thead>
<tr>
<th>fileaccessid</th>
<th>bigint</th>
</tr>
</thead>
<tbody>
<tr>
<td>md5sum</td>
<td>text</td>
</tr>
<tr>
<td>filename</td>
<td>varchar</td>
</tr>
<tr>
<td>location</td>
<td>varchar</td>
</tr>
<tr>
<td>username</td>
<td>tinytext</td>
</tr>
<tr>
<td>application</td>
<td>text</td>
</tr>
<tr>
<td>pid</td>
<td>tinytext</td>
</tr>
<tr>
<td>size</td>
<td>bigint</td>
</tr>
<tr>
<td>cyear</td>
<td>int</td>
</tr>
<tr>
<td>cmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>cday</td>
<td>tinyint</td>
</tr>
<tr>
<td>chour</td>
<td>tinyint</td>
</tr>
<tr>
<td>cmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>csec</td>
<td>tinyint</td>
</tr>
<tr>
<td>dyear</td>
<td>int</td>
</tr>
<tr>
<td>dmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>dday</td>
<td>tinyint</td>
</tr>
<tr>
<td>dhour</td>
<td>tinyint</td>
</tr>
<tr>
<td>dmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>dsec</td>
<td>tinyint</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pentity</th>
<th>bigint</th>
</tr>
</thead>
<tbody>
<tr>
<td>md5sum</td>
<td>text</td>
</tr>
<tr>
<td>username</td>
<td>tinytext</td>
</tr>
<tr>
<td>pid</td>
<td>smallint</td>
</tr>
<tr>
<td>ppid</td>
<td>smallint</td>
</tr>
<tr>
<td>pccpu</td>
<td>decimal</td>
</tr>
<tr>
<td>pmem</td>
<td>decimal</td>
</tr>
<tr>
<td>command</td>
<td>text</td>
</tr>
<tr>
<td>arguments</td>
<td>mediumtext</td>
</tr>
<tr>
<td>cyear</td>
<td>int</td>
</tr>
<tr>
<td>cmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>cday</td>
<td>tinyint</td>
</tr>
<tr>
<td>chour</td>
<td>tinyint</td>
</tr>
<tr>
<td>cmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>csec</td>
<td>tinyint</td>
</tr>
<tr>
<td>dyear</td>
<td>int</td>
</tr>
<tr>
<td>dmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>dday</td>
<td>tinyint</td>
</tr>
<tr>
<td>dhour</td>
<td>tinyint</td>
</tr>
<tr>
<td>dmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>dsec</td>
<td>tinyint</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>endpointinfo</th>
<th>bigint</th>
</tr>
</thead>
<tbody>
<tr>
<td>md5sum</td>
<td>text</td>
</tr>
<tr>
<td>sourceip</td>
<td>tinytext</td>
</tr>
<tr>
<td>sourcedqdn</td>
<td>tinytext</td>
</tr>
<tr>
<td>destip</td>
<td>tinytext</td>
</tr>
<tr>
<td>destfqdn</td>
<td>tinytext</td>
</tr>
<tr>
<td>sourceport</td>
<td>smallint</td>
</tr>
<tr>
<td>destport</td>
<td>smallint</td>
</tr>
<tr>
<td>ipversion</td>
<td>smallint</td>
</tr>
<tr>
<td>cyear</td>
<td>int</td>
</tr>
<tr>
<td>cmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>cday</td>
<td>tinyint</td>
</tr>
<tr>
<td>chour</td>
<td>tinyint</td>
</tr>
<tr>
<td>cmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>csec</td>
<td>tinyint</td>
</tr>
<tr>
<td>dyear</td>
<td>int</td>
</tr>
<tr>
<td>dmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>dday</td>
<td>tinyint</td>
</tr>
<tr>
<td>dhour</td>
<td>tinyint</td>
</tr>
<tr>
<td>dmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>dsec</td>
<td>tinyint</td>
</tr>
<tr>
<td>username</td>
<td>tinytext</td>
</tr>
<tr>
<td>pid</td>
<td>int</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>hwdevd</th>
<th>bigint</th>
</tr>
</thead>
<tbody>
<tr>
<td>md5sum</td>
<td>text</td>
</tr>
<tr>
<td>devbus</td>
<td>tinytext</td>
</tr>
<tr>
<td>devstring</td>
<td>text</td>
</tr>
<tr>
<td>devvendor</td>
<td>text</td>
</tr>
<tr>
<td>application</td>
<td>text</td>
</tr>
<tr>
<td>userslogged</td>
<td>text</td>
</tr>
<tr>
<td>cyear</td>
<td>int</td>
</tr>
<tr>
<td>cmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>cday</td>
<td>tinyint</td>
</tr>
<tr>
<td>chour</td>
<td>tinyint</td>
</tr>
<tr>
<td>cmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>csec</td>
<td>tinyint</td>
</tr>
<tr>
<td>dyear</td>
<td>int</td>
</tr>
<tr>
<td>dmonth</td>
<td>tinyint</td>
</tr>
<tr>
<td>dday</td>
<td>tinyint</td>
</tr>
<tr>
<td>dhour</td>
<td>tinyint</td>
</tr>
<tr>
<td>dmin</td>
<td>tinyint</td>
</tr>
<tr>
<td>dsec</td>
<td>tinyint</td>
</tr>
</tbody>
</table>

*Universitetet i Oslo - AF Security Seminars - 2012*
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.
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


Describes how one can measure user sophistication as a threat metric.
The Insider Threat Model (3)

\[
EPT = \sum F_{\text{threat components}} \Rightarrow EPT = F_{\text{accessrights}} \\
+ F_{\text{behavior}} \Rightarrow EPT = C_{\text{role}} + C_{\text{criticalfiles}} + C_{\text{hardware}} \\
+ C_{\text{utilities}} + C_{\text{sysadm}} + F_{\text{behavior}} \Rightarrow EPT = C_{\text{role}} \\
+ C_{\text{criticalfiles}} + C_{\text{hardware}} + C_{\text{utilities}} + C_{\text{sysadm}} \\
+ F_{\text{sophistication}} + F_{\text{fileops}} + F_{\text{execops}} + F_{\text{network}}
\]
From LUARM data to a language

Concept of insider misuse/threat

semantic mapping

ITPSL symbols

Real world manifestation of insider threat/misuse

Concept of IP theft misuse

semantic mapping

LUARM target - SQL:
SELECT * from psinfo
WHERE
command=/sbin/su
AND
username=user
...

Translational description

ITPSL symbols

System process set for IP theft:
su - masquerade_as_user
/usr/bin/swriter file
...
cp file /media/usb
exit

Universitetet i Oslo - AF Security Seminars - 2012

Saturday, April 21, 2012
# 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)


**Keywords:** Data security, Information systems

**Article type:** Concept paper

**DOI:** [10.1108/09685220610690826](10.1108/09685220610690826) (Permanent URL)

**Publisher:** Emerald Group Publishing Limited

---


---

*Universitetet i Oslo - AF Security Seminars - 2012*
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)

ITPSL XML Signature → Schema Validation → Valid signature?

NO → ABORT

YES → ITPSL Schema Definition (XSD) → ITPSL Xpath/Twig Compiler → LUARM SQL Statements
Example 1:

Pornographic access detection scenario

Saturday, April 21, 2012
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>
</itpslsig>
```

\[ \sum_{n=1}^{\text{wevent}_n} = \text{EPMO} \]

EPMO -> Evaluated Potential Misuse Occurrence (0...1)
n-> number of specified events

\[ <\text{weightmatrix}>n\text{events, wevent}_1,\text{wevent}_2,\ldots,\text{wevent}_n </\text{weightmatrix}> \]
The ITPSL compiler

- submitsig.pl
- searchesig.pl
- getsig.pl
- validate.xsd
- signatures_repository.sql
- hostA
- hostB
- hostN

LUARM
LUARM
LUARM

Universitetet i Oslo - AF Security Seminars - 2012

Saturday, April 21, 2012
Performing an ITPSL ontology search

1. searchsig.pl
2. signatures_repositories.sql
3. sigid1 sigid2 ... sigidn
4. getsig.pl
5. ITPSL .sig
6. itpslc.pl

Saturday, April 21, 2012
Running an ITPSL signature to the compiler

1. ITPSL.sig
2. signatures_repository.sql
3. itpslc.pl
4. validate.xsd
5. EPMO/0/1

hostA

hostB

hostN

LUARM

LUARM

LUARM
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.
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