Explaining CEE
The Need for Event Standards

11 September 2008
The ‘Log’ Problem

MITRE was involved with research related to Network Operations and Security Center (NOSC) products for security information processing. This effort led to involvement with the analysis of Security Information Management (SIM) Systems.

While SIMs show a strong promise for NOSCs, we identified some limitations that would affect its success.
The Challenge

Today

- 1000s of devices
- 100s of events
- Multiple ways of expressing events

Tomorrow ++

A Common Event Expression will improve

- Log management capabilities
- Log correlation (SIM) capabilities
- Device intercommunication enabling autonomic computing
- Enterprise-level situational awareness
- InfoSec Adversarial Modeling through the integration of Red, Blue, and White Team reports with sensor logs and SIM reports

Solve the problem of “inconsistent log formats”, since “there is no consensus in the security community as to the standard terms to be used to describe the composition of log entries and files.” (NIST Publication 800-92)
Motivation
From Events to Logs
The Log Space

Log Records

Event Records

Reports

Debug
OODA Loop for CND

- **Observe** – Logs, Vulnerability Reports, News
- **Orient** – History, Policy, IT Information
- **Decide** – Good, Bad, Unknown, Watch, Ignore…
- **Act** – Block or Allow? Refine Rules or Policy?
… Why It Doesn’t Work

- CND designed for business efficiency
- Assume static, stable network architectures
- Rely on Normalization and Rules-based Signature Matching
Problem

LOGS ARE PRODUCED FOR THE WRONG AUDIENCE

Humans understand **semantics**
Systems understand **syntactics**
Why Standardize? (1)

- **Missing Event Details**
- **Cryptic Records**
  
  Sep 01 08:11:53 Last message repeated 5 times

- **Problem: Inconsistent Success/Fail**
  
  Apr 10 12:31:34 host sshd[16682]: error: PAM: Authentication failure for user from remote-pc.mitre.org

  Apr 10 12:31:39 host sshd[16701]: Accepted keyboard-interactive/pam for user from 192.168.0.1 port 2880 ssh2
Why Standardize? (2)

- **Inconsistent Event Descriptions**

  Sep 22 10:02:00 myhost login(pam_unix)[808]: session opened for user root by LOGIN(uid=0)

  Sep 26 12:17:32 myhost-- root[808]: ROOT LOGIN ON tty1

  Sep 26 13:00:40 myhost snort: [1:5503:6] POLICY ROOT login attempt [Classification: Misc activity] [Priority: 3]: {TCP} 6.7.8.9:32804 -> 1.2.3.4:23
CEE Enabled Process

CEE = Syntax + Vocabulary + Transport + Log Recommendations

- Events recorded guided by **Log Recommendations**
  - Events and details needed to be logged by devices (OS, IDS, FWs, etc.)
- Log messages exchanged via a **Common Log Transport**
- Log messages received in a **Common Log Syntax** for parsing out relevant data
- **Common Event Expression Taxonomy** to specify the event in a common representation
Scenario: An attacker has breached our network
– 1. Determine if any successful logins
   • What do we search for – ‘log in’, ‘log in’, ‘logged on’ etc.

Taxonomy – a reduced language set for consistent log messages
Right now they are semantically similar – ok for humans – but not for computers
Future Potential (1)

Move from IT to Mission-focused awareness
Future Potential (2)
How can You Help?

- Provide Your Log Requirements
- Join the Working Group
- Help Grow the Community
  - Request Vendors to Participate
  - Inform Others about the Effort

http://cee.mitre.org
Backups
There is no common agreement, understanding, or representation for "event".

Solution Strategy:
- Increase the density of meaningful information
- Decrease representation size of voluminous, raw log data

Characteristics:
- Layered representation volumes, but well understood
- Map categorical meaning to event descriptions

Event Messages buried in Log Files = Volume

Event Descriptions buried in Event Messages

Event Meanings buried in Event Description

x Parsing & Translation

Here there be dragons
CEE Challenge

Messages from Logs:

Tera-"lines" of log messages ...
Sep 26 00:03:10 zan postfix/smtpd[7949]: connect from unknown[208.66.74.58]
Sep 26 00:03:11 zan postgrey[18992]: cleaning up old logs...
Sep 26 00:03:11 zan postgrey[18992]: delayed 687 seconds: client=208.66.74.58, from=adds@mercedmedia.com, to=zander@intrusion.org
Sep 26 00:03:13 myhost - root [808]: ROOT LOGIN ON tty1
Sep 26 00:03:13 zan postfix/smtpd[7949]: 185D1381BC6: client=unknown[208.66.74.58]
Sep 26 00:03:14 myhost snort: [1:5503:6] POLICY ROOT login attempt [Classification: Misc activity] [Priority: 3] (TCP) 6.7.8.9:32804 -> 1.2.3.4:23

Descriptions from Messages:

Terabytes of words that describe events ...
Blah blah 6.7.8.9:2804 blah blah myhost blah blah blah login attempt blah blah blah 1.2.3.4:23 blah blah blah

Meanings from Descriptions:

These 3 log lines mean the same thing ...

PAM: Sep 26 12:00:00 myhost login(pam_unix) [808]: session opened for user root by LOGIN(uid=0)
Linux: Sep 26 00:03:13 myhost - root [808]: ROOT LOGIN ON tty1
Snort: Sep 26 00:03:14 myhost snort: [1:5503:6] POLICY ROOT login attempt [Classification: Misc activity] [Priority: 3] (TCP) 6.7.8.9:32804 -> 1.2.3.4:23

Common Event Representation:

???? Standards: IDMEF, SDEE ... Initiatives: CIEL ... log transport, format, SIM taxonomies ... ???
“[We] must keep it simple and stupid or it'll be ASN.1 before we know what hit us…” – Marcus Ranum
Example Log Messages

1. Sep 26 12:00:00 myhost-- root[808]: ROOT LOGIN ON tty1
2. Apr 10 12:30:34 hostname sshd[16682]: error: PAM: Authentication failure for user1 from host.domain.com
3. Sep 19 08:26:10 zuric CEF:0|security|threatmanager|1.0|100|worm successfully stopped|10|src=10.0.0.1 dst=2.1.2.2 src=1232

Syntax - details specific to event being logged
Format (1 and 2)
month day time host program[pid]: message

In CEE, each of these would be a possible syntax element, whose value and definition would be well defined by a Data Dictionary

Data Dictionary needed to enumerate details associated with the event
It needs to provide flexible syntax options by defining the elements and formats
Ex: dst in the CEF event – defined by a dotted quad IPv4 address

Transport - successful log transmission
Syslog (each log message is transmitted in a single UDP packet usually over port 514/UDP)
CEE Taxonomy

Event Message Details

1. ROOT LOGIN ON tty1
2. error: PAM: Authentication failure for user1 from host.domain.com
3. worm successfully stopped

- Scenario: An attacker has breached our network
  - 1. Determine if any successful logins

Example of CEE Event Taxonomy

1. Login (tty1) for user (root) successful
2. Login for user (user1) failed from host.domain.com
3. worm stopped