Safely Sharing Data Between CSIRTs: The SCRUB* Security Anonymization Tool Infrastructure

William Yurcik*
<byurcik@gmail.com>
Clay Woolam, Greg Hellings, Latifur Khan, Bhavani Thuraisingham

University of Texas at Dallas

20th Annual FIRST Conference
Vancouver Canada
June 2008
The SCRUB* Architecture

Organization Enabled for Distributed Sharing

(1) SCRUB-tcpdump

(2) SCRUB-PACCT

(3) SCRUB-NetFlows

(4) SCRUB-Alerts

CANINE (format converter)

MSSP

CERT

NetFlows (Cisco, Argus, IPFix)

IDS

Firewall

Virus

Other Organizations

ISAC

packet traces

commands

processes
**SCRUB* Motivation**

**Why Should We Share Security Data?**

- **Event correlation across administrative domains is needed based on shared data**
  - We cannot continue to stop attacks at organizational borders, we need to cooperate with law enforcement and each other.
  - Chasing attackers away to other organizations does not improve security
- **Need to share security data between organizations in order to**
  - Detect attacks
  - Blacklist attackers and attacker techniques
  - Distinguishing normal versus suspicious network traffic patterns
State-of-the-Art in Security Data Sharing

All too common

1

Data Sharing
Organization
Policy Domain

2

3

4

5

6

All too common
For Safe Data Sharing: Two Types of Data To Protect

- **Private Data**
  - User-identifiable information
    - user content (Email messages, URLs)
    - user behavior (access patterns, application usage)
  - Machine/Interface addresses
    - IP and MAC addresses

- **Sensitive Data**
  - System configurations (services, topology, routing)
  - Traffic patterns (connections, mix, volume)
  - Security defenses (firewalls, IDS, routers)
  - Attack impacts
SCRUB* TOOL 1:

- Anonymizes packet traces
  - packet traces can contain the most private/sensitive data
  - packet traces are the authoritative raw security source
- Leverage a popular existing tool – tcpdump
- Anonymizes any/all packet fields (12)
- Each field has multiple anonymization options
  - none/low/medium/high levels of protection for protecting the same data field
**SCRUB* TOOL 2: SCRUB-PACCT**

- Anonymizes process accounting logs
  - process accounting records contain user IDs and user command behavior
  - process accounting records contain precise timing information for event correlation between systems
- Anonymizes any/all process accounting fields (16)
- Each field has multiple anonymization options
  - none/low/medium/high levels of protection for protecting the same data field
SCRUB* TOOL 3: SCRUB-NetFlows

- Anonymizes NetFlow logs
  - NetFlows logs efficiently aggregate packet traffic by connections
  - Most commonly shared security data
- Anonymizes any/all NetFlow fields (5)
- Each field has multiple anonymization options
  - none/low/medium/high levels of protection for protecting the same data field
SCRUB* Fields of Interest Between Data Sources

1. Transport Protocol Number
   data sources: packet, NetFlows, alerts

2. IP Address
   data sources: packet, NetFlows, alerts

3. Ports
   data sources: packet, NetFlows, alerts

4. Payload
   data sources: packet, alerts

5. Timestamp
   data sources: packet, process accounting, NetFlows, alerts
Multi-Level Anonymization Options

- Black Marker (filtering/deletion)
- Pure Randomization (replacement)
- Keyed Randomization (replacement)
- Annihilation/Truncation (time, accuracy reduction)
- Prefix-Preserving Pseudonymization (IP address)
- Grouping (accuracy reduction)
  - Bilateral Classification
- Enumeration (time, adding noise)
- Time Shift (time, adding noise)
A Problem with Anonymization for Sharing: Privacy vs. Analysis Tradeoffs

while anonymization protects against information leakage it also destroys data needed for security analysis

- Zero-Sum? (more privacy <> less analysis & vice versa)
- to date, no quantitative measurements of how useful anonymized data is for security analysis
Empirically Measuring Anonymization
Privacy/Analysis Tradeoffs

- Series of experiments to test effects of different anonymizations options
- Use snort IDS alarms as a metric for security analysis
Summary

• There is a critical need for security data sharing between organizations.

• Anonymization can provide safe data sharing:
  – Multi-Field: prevent information leakage
  – Multi-Level: no one-size-fits-all anonymization solution

• A practical data sharing infrastructure is needed which supports multiple data sources:
  – SCRUB* tool suite for packet traces, process accounting, NetFlows, alerts

• Privacy/analysis anonymization tradeoffs can be characterized:
  – Zero-Sum tradeoff? (not always, more complex than this)
  – Multi-Level anonymization options can/should be tailored to requirements of sharing parties to optimize tradeoffs
  – More tradeoff measurements are in progress
References

Background on Using Anonymization to Safely Share Security Data


SCRUB* Tool (1) SCRUB-tcpdump <http://scrub-tcpdump.sourceforge.net/>


SCRUB* Tool (2) SCRUB-PACCT <http://security.ncsa.uiuc.edu/distribution/Scrub-PADownLoad.html>


SCRUB* Tool (3) SCRUB-NetFlows <http://scrub-netflows.sourceforge.net/> >

