Carnegie Mellon
Software Engineering Institute

CERT
Situational
Awareness

AirCERT: Building a Framework for Cross-Administrative Domain Data Sharing

Roman Danyliw <rdd@cert.org>

FloCon 2004: Complementary Architecture Panel

CERT® Network Situational Awareness Group
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213-3890

The CERT Network Situational Awareness Group is part of the Software Engineering Institute. The Software Engineering Institute is sponsored by the U.S. Department of Defense.

Background

- Form situational awareness for the SEI, its sponsors, and the Internet community
  - Big picture view of threats
- Constraints
  - Situational awareness can only be formed with data from many organizations – all data is governed by the constraints of its owners
  - Must provide a reasonable value-proposition for data sharing
  - Strict hierarchies in data sharing will not scale
  - Solutions must be built with open and transparent architectures

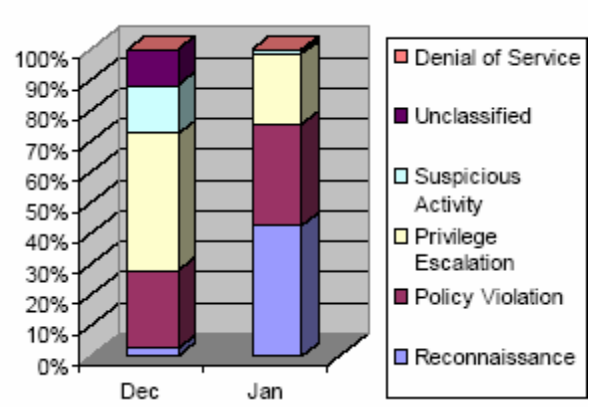
Analytical Concerns

Focus on merging and analyzing data from multiple view points

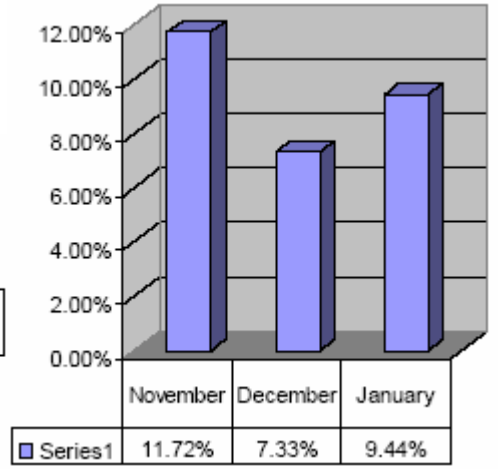
- Distinguish between targeted, localized, and Internet-wide activity
  - Widely targeted services
  - Clusters of attacks
    - Passive detection of new tools
  - Attack techniques *de-jour*
  - Attack sources
- Historical trending
  - Enable forward estimation of expected intruder activity of a site

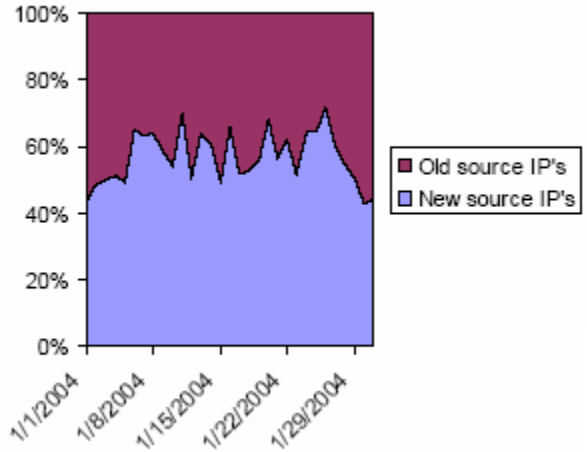
Current Results

- Generating "Top 10" lists and volumetric measures based on
  - *Packet/Flow features*: IP addresses, ports, protocols, signature, etc.
  - *Context*: timing, vulnerability, country, net-blocks, etc.



Share Source IP addresses Targeting multiple organization





Implementation

- **`http://aircert.sourceforge.net`**

- Gather data from existing security solutions already deployed
  - Partner with security operations in the federal civilian community and in academia
- Write “glue” to integrate, convert, analyze, and share the data across organizations
- Provide analytical results back to participants and sponsors

Synthesized Data

- Categorization
  - SIM/SEMs (e.g., ArcSight)
  - IDS (e.g., Snort)
- Discovery
  - Flow data (e.g., argus)
- Refinement
  - Network topology information
  - IT/data data sharing policies
- Context
  - Vulnerability (e.g., CERT/CC KB)
  - Artifacts (e.g., CERT/CC AC)

Collection Infrastructure

- Provides infrastructure to *automatically* extract relevant information from existing instrumentation
  - If human intervention is required, sharing is too expensive
- Wrote “normalizers” to handle the reformatting and semantic transformation of the data
  - Too many vendor to write one-off tools for each
  - Write transformation engine that understands the underlying data-store: text files, RDBMS, etc.

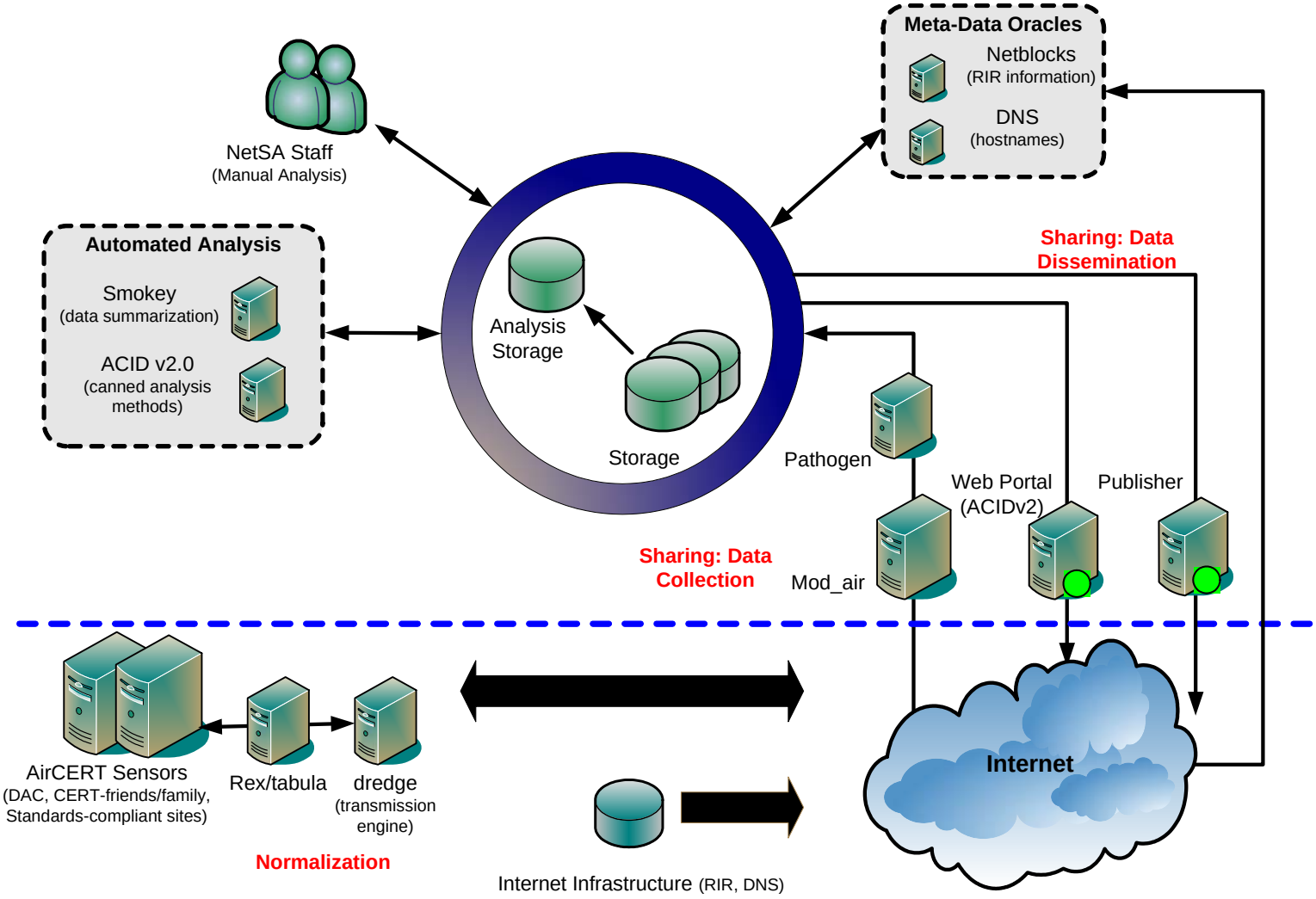
Sharing Infrastructure: Collection

- The key to facilitating data sharing across organizations is
  - Making it seamless – no human interaction
  - Ensuring policy compliance
- All “normalizers”, “publishers”, and the underlying storage architecture have a notion that all data has an owner
  - Dissemination respects site’s local policy
  - Sanitization of sensitive data
  - Tagging of all data with a source identifier

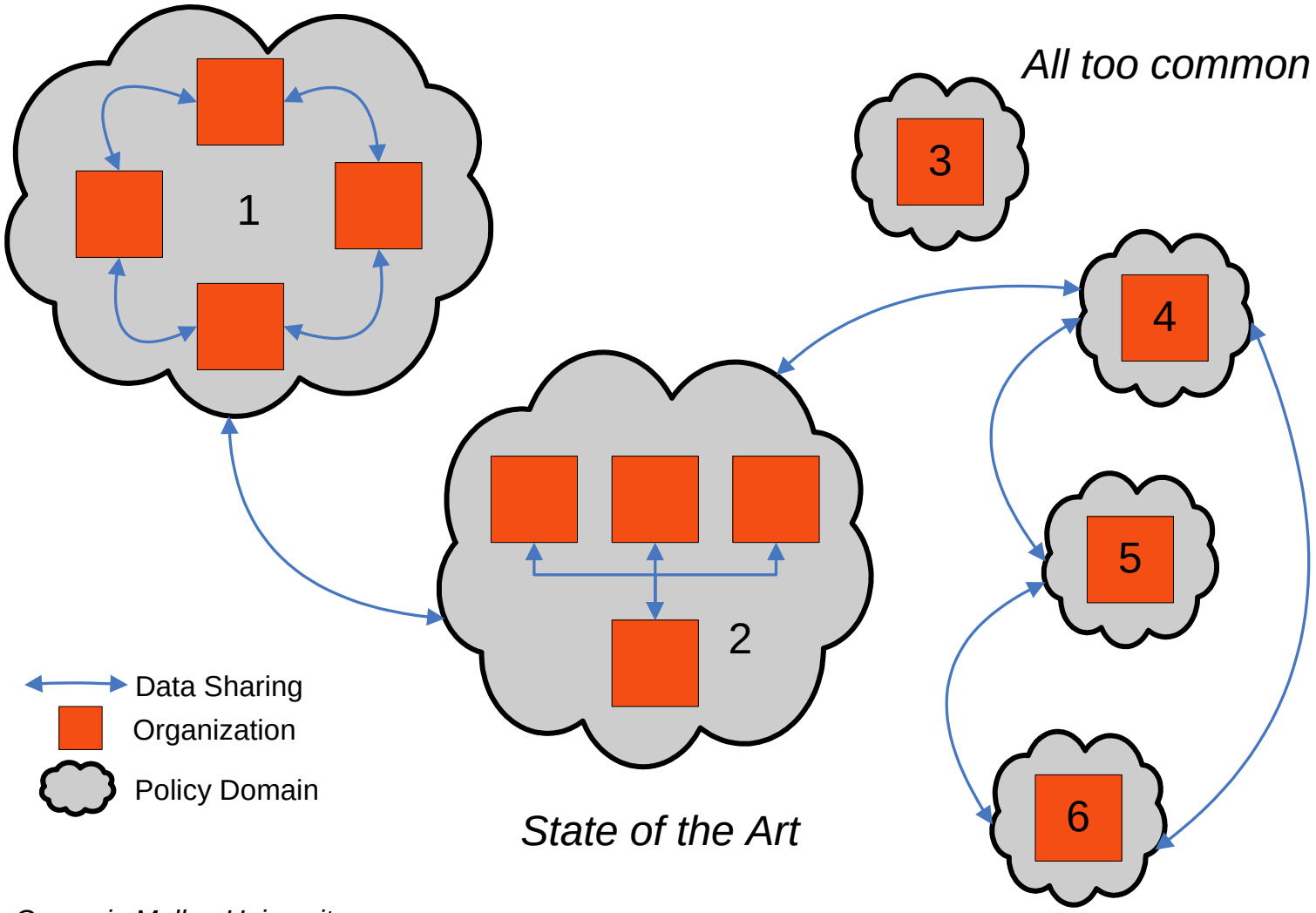
Sharing Infrastructure: Dissemination

- Sharing data with us, is no different than data with others
- Tailor channel for the audience
  - Web-portal for pre-digested snapshot
  - Export bulk-data in a machine-readable format (e.g., XML, RSS)

Architecture



Big Picture Architecture



Challenges and Solutions

- Many different formats used by the SEM and IDS products
  - Support standards efforts: IDMEF, IODEF, IPFIX, PSAMP
  - Storage-specific normalization tools
- Normalizing signatures across IDS products
  - Using CVE and custom classification taxonomies
- Analyzing the correct signature set
  - Use only explicit malicious activity
  - Filtering out policy violations and poorly written signatures
  - Use the correct tool for the task
    - Deploy non-IDS sensors next to the IDS
- Data loops
  - “Checksums” in the meta-data of the data stream

Challenges and Solutions (2)

- Need both push and pull model, while supporting varied levels of automation
  - Unified presentation engine (ACIDv2)
  - Publisher for bulk-data transfer

Ongoing Work

- Intelligent end-points that summarize instead of sending all data
- Automated ways to overlay the context provided by vulnerability and artifact information
- Continued support for standards work
- Improved attention focusing techniques for flow data-to-IDS and vice versa
- Improved approaches for integrating the analytical products into operations