FACTOIDS

Carlos M. Martinez-Cagnazzo

FIRST Miami 2010
A Little About Me...

- I have been involved in incident response since 2005
  - So i have suffered all of the problems I'm going to talk about today
- This work is still a work in progress so any:
  - Criticism
  - Comments
  - In fact, anything you have to say about it
- is very welcome!
The Problem

- "How (this or that network activity) is seen from other networks?"
  - Many times, when responding to computer security incidents this question pops up
- Other people (researchers, software developers) often face similar problems
  - How to obtain datasets of network activity from other groups?
The Problem (II)

- How do we "solve" it today?
  - Public domain datasets
    - CAIDA
    - Others
    - Limited, many times outdated
  - Try to generate data ourselves
    - Limited view, can be "expensive"
  - Get data from other groups
    - A trust relationship must be in place
    - There are regulations that must be taken into account
Datasets

- Research and CSIRT groups build datasets that can be of great value to other groups
  - Data collected by sensors (honeypots, IDSs)
- However, multiple issues make the exchange of these datasets difficult
  - Trust
    - Do you trust your peer to share data?
  - Local and regional regulations
    - Personal data protection laws
  - Information Security policies within each organization
Trust - How do We Manage It Today?

- We do what we are doing here today at this FIRST event. We meet and get to know each other...
- We establish some framework
  - MoUs (memorandum of understanding)
  - NDAs (non-disclosure agreements)
- We then filter the data we decide to share
- Problems:
  - Not very flexible, once an MoU is in place it's very hard to modify
  - Language barriers, different laws make compatibility difficult
Is There a Better Way?

- What if?
  - Many groups (CSIRTs) exposed at least a *sanitized* view of their sensor data for other groups to use?
- Sanitization
  - *Hide what you don't want other people to see*
  - Typically
    - Internal IPs / hostnames
- Then...
  - Incident responders / researches would then be able to tap on these sources of data
    - Downloading only the data they need, maybe using a specialized query language
A Better Way...

1. Mutually prove identities
2. Instantiate policies

3. CSIRT #2 transmit uses an API exposed by #1
4. CSIRT #1 outputs data filtered according to the instantiated policy
What Do We Need to Make this Possible?

- Data models
  - We need to represent sensor data in an uniform way
- Automatic sanitization
  - The data each group sends to another must be automatically sanitized (no human intervention)
- Query languages / APIs
  - Efficient transport
- A directory of available information
  - Which group has which kind of information?
FACTOIDS – Vision

• FACTOIDS: "Models and Tools for Analysis and Secure Exchange of Sensor-Collected Data"

• Vision:
  • Efficient creation and management of trust relationships between groups who need to share security event data
    – Have control over which data is shared with whom
  • Improve transport efficiency of large sets of related data (datasets) regarding security events
    – Transport only what is needed
FACTOIDS – Potential Uses

- Real-time data exchange when responding to incidents
- Current and "real" datasets for research purposes
- Law enforcement agencies
- Software developers
FACTOIDS – Architecture

FACTOIDS

IDS

Firewall / Router

Honeypot

Netflow Source

Fuentes de contexto

A1

A2

A3

A4

B2

Correlación de eventos

Detección temprana de incidentes

Enriquecimiento contextual

Base de eventos normalizados

Base de datos de incidentes

Repositorio de políticas

Procesador de comandos

Control de acceso

Motor de sanitización

FACTOIDS

Sistema #1

Sistema #2

Interfaz Humana
Dataset Transport

- **Normalization**
  - The same event is almost always detected by several sensors
  - We must normalize our dataset to avoid duplicates

- **Size**
  - A honeynet, firewall or IDS can by itself generate large-size datasets
  - Instead of moving huge files back and forth, why not expose an API and/or a query language?
Sanitization Policies

- **Sanitization**: output data filtered according to a policy

- **Techniques**:
  - Non-property preserving
    - "Black Marker"
  - Property preserving
    - Depending on the type of data
Conclusions

- We believe there is a need for a set of models and techniques to enable more data sharing among groups
- Many stand to benefit from a more open data sharing environment
- Roadblocks need to be overcome
  - Comply with regulations
  - Ensure groups control what information gets shared
FACTOIDS – Next Steps

- Implement a sub-set of the system
  - Representing events as XML documents using IODEF (RFC 5070)
  - Implement a proof-of-concept sanitization engine and policy repository
  - Implement event loading from our honeypot network
Thank you for your patience!

¿Questions? (carlos.martinez@csirt-antel.com.uy)