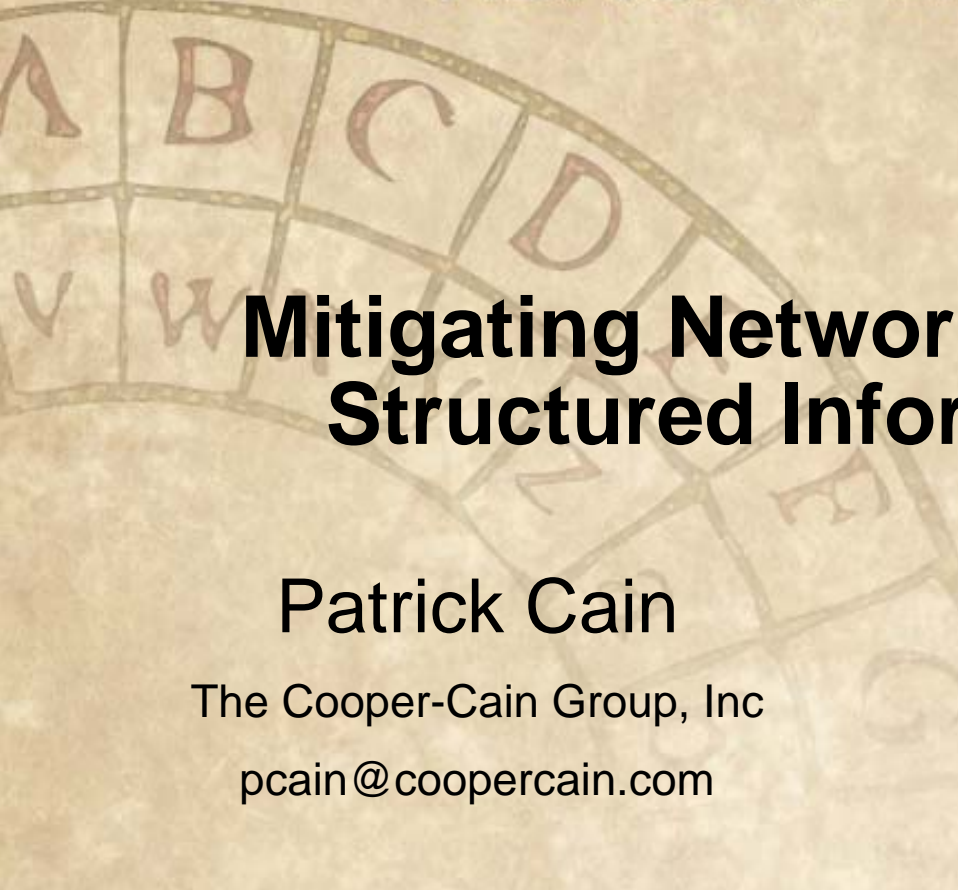




**RSA<sup>®</sup>CONFERENCE 2007**



**Mitigating Network Events Through  
Structured Information Sharing**

**Patrick Cain**

The Cooper-Cain Group, Inc  
pcain@coopercain.com

**Roman Danyliw**

CERT Program  
rdd@cert.org

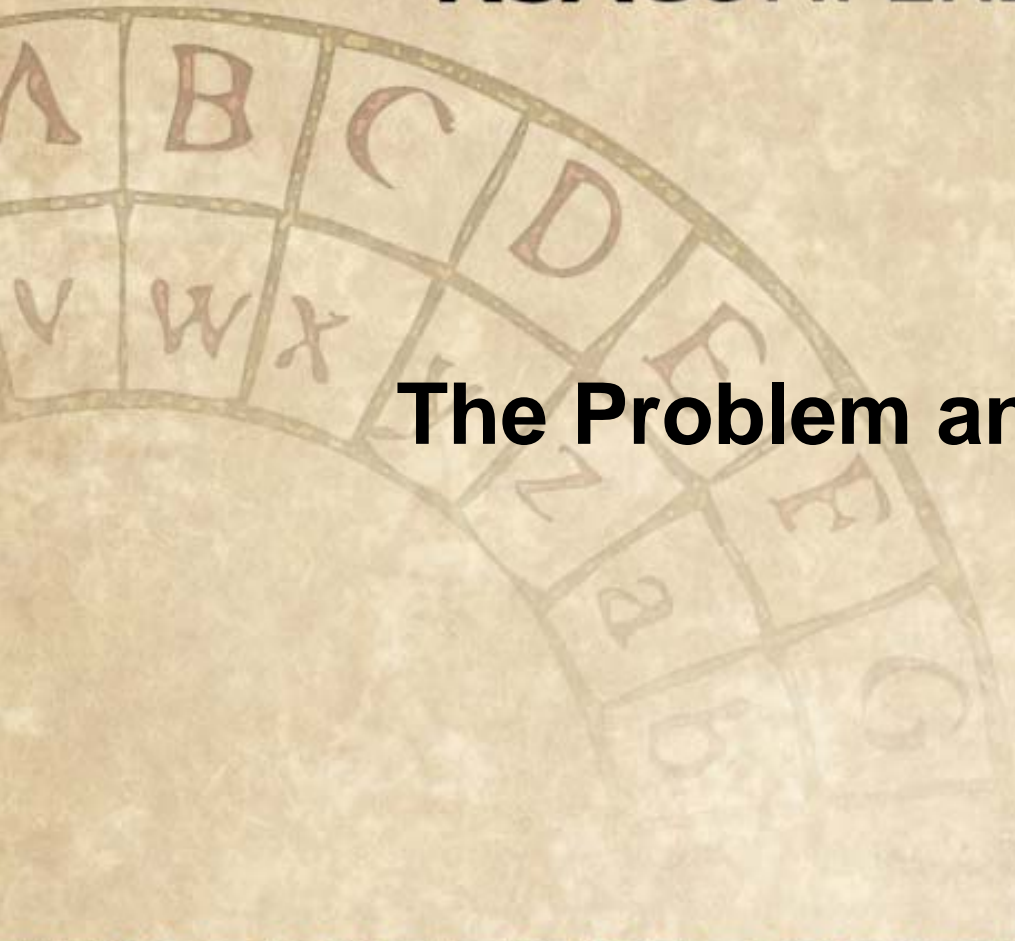


# Outline

- The Problem and Challenge
- Standardization efforts in the IETF
- Anti-Phishing Working Group: An Example Solution
- Lessons Learned



# **RSA<sup>®</sup>CONFERENCE 2007**



## **The Problem and Challenge**



# Defining the Problem

- Philosophy
  - Ignore the politics of whether we should share data, or if people will actually do it...
  - ... and focus on the communities who want to share incident data
- Sharing data is the means, not the end goal
- Particular use-cases will scope:
  - What to do with the data?
  - What is the right data?
  - How to share the data?
  - With whom should it be shared?

# Observations: Motivations

- The purpose of data sharing is security event mitigation
  - Timeliness is key to resolving ongoing activity
  - Retrospection is important to understanding trends
- Timeliness necessitates automation
  - Structured data -- defined semantics, protocols, failures, and errors
  - Ease of reporting -- integration with existing work-flow process
- Trending requires efficient archiving
  - Comparable structured data as above, but kept historically
  - Scalability may necessitate:
    - Aging – deletion after some period of time
    - Aggregation – derived and reduced data
  - Diversity in the observed data

# Observations: Sharing Partners

- External parties may:
  - Not speak my language
  - Not have my level of expertise
  - Not have the same detection, collection, or remediate infrastructure
- External parties have different requirements for the data
  - Remediation – source and target sites
    - Sufficient detail for making changes
  - Trending -- involved or interested 3<sup>rd</sup> parties (e.g., ISAC, Network Intelligence Services)
    - Aggregation, making fidelity less significant
  - Prosecution -- Law Enforcement Agencies (LEA)
    - Acquisition, custody, and retention issues
  - Research – universities, labs, R&D efforts

# Observations: Process

- The lowering the bar for participation will yield a greater number of participants
  - Readily available tools that support sharing
  - Lowering the threshold for the quality of accepted information
- Some privacy and confidentiality must be lost for some gain
  - The producer of the information must drive this trade-off
- A shared information model is more desirable than normalization
- Standardized information models need to be flexible
  - Understanding about an incident grows as more information is collected or analyzed
  - Every incident is different, in some way
  - What constitutes an “incident” varies by organization

# A Review of the Approaches

## Current

- Event is detected
- Event is reported
  - Reported to somebody
  - Reported to “correct” somebody
    - Maybe in the right language this time...
  - More info requested... (repeat)
  - Reported again... (repeat)
- Response started
- Attacker long gone

## Suggested

- Get appropriate and correct data in one report
  - Sufficient data for use by the audience (e.g., investigation)
  - Standardize on a common framework with some flexibility on semantics and taxonomy
- Use an already understood format to enhance acceptance (if possible)
- Make it easy-to-use





# **RSA<sup>®</sup>CONFERENCE 2007**



## **IETF Efforts**



# Extended Incident Handling working group (INCH)

- Define a transport format to encode information commonly exchanged between Computer Security Incident Response Teams (CSIRTs)
  - Data relevant across administrative domains
- Incident Object Description Exchange Format (IODEF)
  - XML Schema
  - Mix of free-form text and enumerated values
  - Recursive design reduces redundancy and obviates need for XML refs
  - Supports references rather than encapsulating the actual data
  - Ability to summarize and report the same information at different levels of detail
  - Incomplete for all purposes, but extensible

# INCH WG: Assumptions

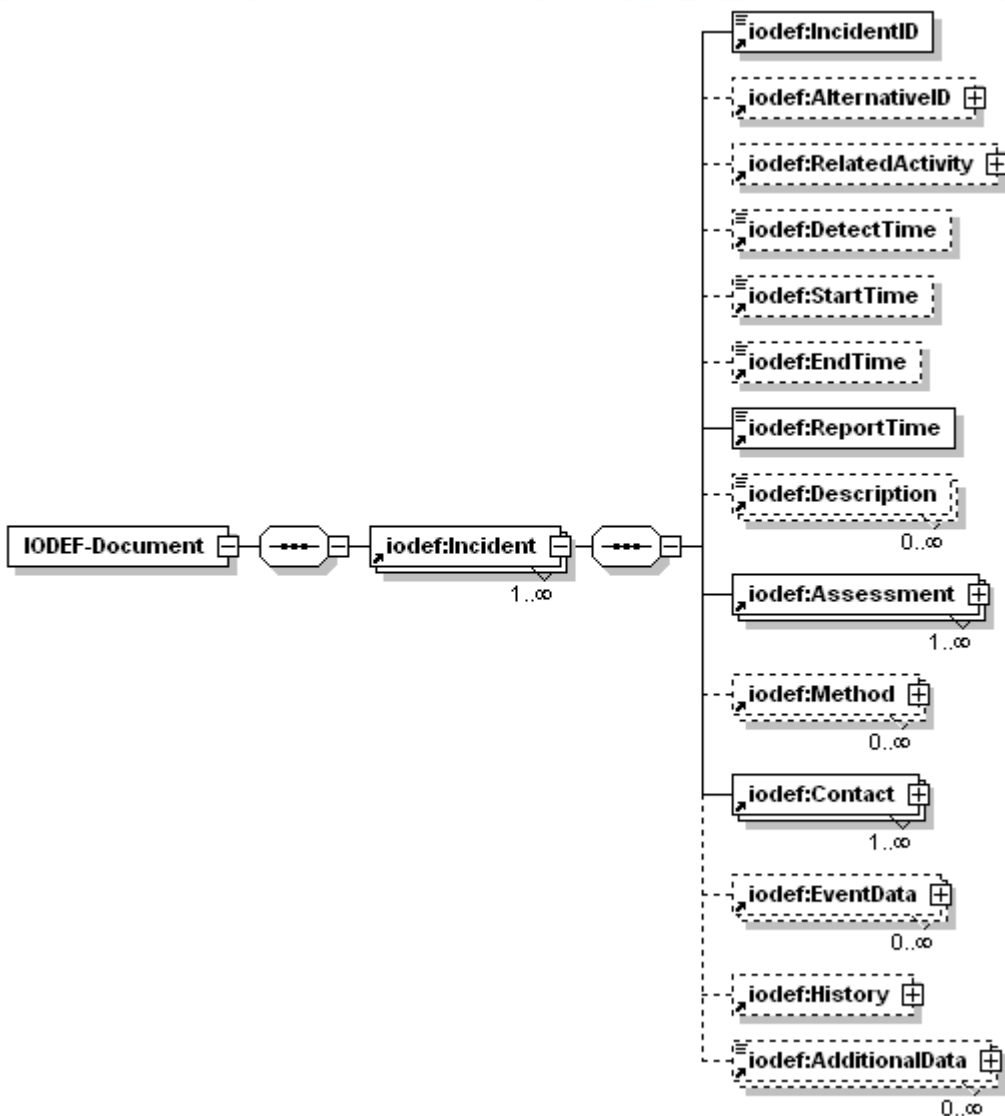
- Incidents are not IDS alarms
  - “Incidents are composed of events”
- Agnostic to specific incident taxonomies
  - “Your definition/threshold of an incident may be different than mine”
- Incidents are numbered and there is state kept about them
  - “Organizations assign incident IDs and have ticketing/handling/correlation systems that process them”
- Merely a wire format
  - “Sharing is different than storage and archiving”
- Incomplete information
  - “You may require more complete information than I need, can get, or have right now”

# INCH WG: Status

- Status of the work
  - INCH WG has concluded
  - `draft-ietf-inch-iodef-10` under review by the Security Area Director for standards track RFC publication
  - All other documents are now individual drafts
  - Limited implementations
- Further reading
  - Summary Website
    - <http://www.cert.org/ietf/inch/>
  - Email Archive
    - <http://listserv.surfnet.nl/archives/inch.html>

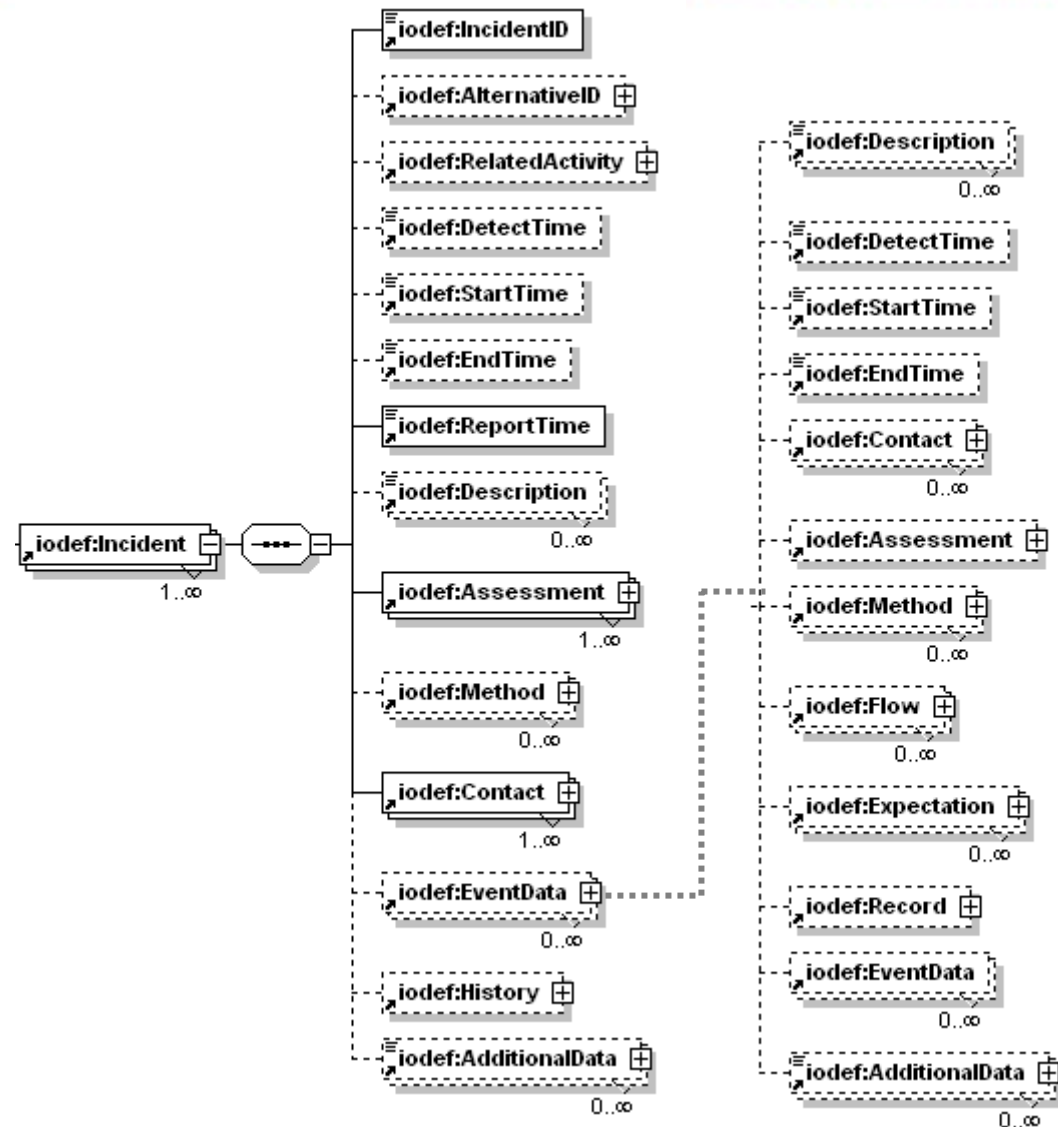


# IODEF Data Model: Meta Data



- CSIRT operations
  - Incident identifiers
  - Contact information
- Internationalization
  - Various encodings
  - Translations
- Data handling labels
  - Sensitivity
  - Confidence
- Extensibility of attributes and adding new elements

# IODEF Data Model: Core



- Timing information
- Enumeration of hosts or networks
  - e.g., IP addresses, ports, protocols, applications, etc.
- History and requested action
- Exploit and vulnerability references
- Impact expressed technically, financially, or by time
- Forensics information

# Implementing IODEF

- Prearranged “profiles” between parties are required to define:
  - Minimally required information (i.e., required “optional” fields)
  - Semantics of weights (e.g., “low” vs. “high”)
  - Extensions
- Data model is not completely machine-parsable
  - Text blobs
  - Unknown extensions
- Requires integration with existing incident handling system
  - IODEF does not readily capture internal workflow
  - Export and import filters are necessary to translate between IODEF and ticketing (correlation) system
    - Import = IODEF → [translator] → ticketing system
    - Export = Ticketing system → [translator] → IODEF

# Implementing IODEF (2)

- IODEF integration is not merely data translation
  - Honoring meta-data (e.g., sensitivity labels)
  - Establishing trust infrastructure (e.g., key infrastructure)
- Transport considerations
  - Real-time Inter-network Defense (RID) protocol
    - Message semantics to IODEF
    - **draft-moriarty-post-inch-rid-00\***
  - SOAP wrapper for RID
    - Transport binding for RID over BEEP and HTTP/TLS
    - **draft-moriarty-post-inch-rid-soap-00\***

\* <http://www.ietf.org/internet-drafts/{file-name}>



# Related Standards Work

## IP Flow Information Export (IPFIX)

- Define a data model to describe IP flows and an associated protocol to exchange it
- Standardize “Netflow/flow/cflow/argus”

## • Packet Sampling (PSAMP)

- Extend the IPFIX data model to support packets

## • Cross Registry Information Service Protocol (CRISP)

- Structured and extensible “whois” query protocol

## • Intrusion Detection (IDWG)

- Standardized IDS alerts
- Intrusion Detection Message Exchange Format (IDMEF)



# RSA<sup>®</sup>CONFERENCE 2007



## An Example Solution

The APWG repository



# The Anti-Phishing Research Group (APWG)

- An independent organization of ~2500 international corporate, individual, law enforcement, and research members
- It's goal is to disperse anti-phishing and anti-phraud information and experiences
- Hosts a repository of ~600,000 phish and fraud attempts since '03
  - Mostly email, some other; additional 80-90,000/month received
  - Anyone can report phishing/fraud attempts
  - Every 5 minutes a list of URLs to block is generated and distributed to many web browser blockers, spam filterers, and anti-viral vendors

Anti-Phishing  
Working Group

APWG

# The APWG Repository



- Phishing/Fraud Reports as Data In
  - Email
  - 'Real-time'
- Database
- Data Out
  - Statistics
    - The famous monthly report
  - Searches
    - To compare amongst brands
    - To gather information for investigations
  - Products
    - URLs-to-Block list



# Phishing and Other Frauds ☹️

- Phishing-specific challenges:
  - The phished institution is always the last to know
  - Most victims are hooked in the first  $n$  hours, where  $1 < n < 5$
- To { block | react | cry } requires quick reaction
  - How could reporters identify phishing sites easily and quickly so they get included in the URL block list?
    - Quickly → automated, no humans
    - Easily → machine generated and processed

# Concerns in a Solution

- How could we get quick acceptance?
  - Ease of use and reporting
  - Simple creating and data mining tools
  - Make it so \*ALL\* incident repositories accept the same format
- Make sure solution is expandable
  - Incidents evolve
- Quick implementation for reporters

# A Solution ?

- “Brew our own” ideas....
- The IETF defined an XML-based format to report incidents among CSIRTs! [IODEF]
- We created extensions to the IODEF format for phishing & crimeware
- Use the structured XML report to shorten the reported → URLlist time

# PhraudReport Structure

- A Phishing or Phraud Report contains:
  - Type of Attack
  - Brand Name involved
  - Info about the Data Collection Site
  - How the attack was Detected
  - Forensic/Archived Data about the Attack
  - Lots of Comment Areas
  - Information about Related sites or attacks
  - Info about Email (Headers, Content, etc)



# Does it work?

- The machine processing has been a big win
  - Incomplete reports can be dealt with automatically
  - Invalid reports can be rejected promptly
  - A URL shows up on the block list about 10 minutes after it is reported
- Some interoperability testing occurred
  - There is at least two implementations
  - Negotiation with other phish reporters is ongoing
  - U2 can send in XML reports ([report\\_iodef@antiphishing.org](mailto:report_iodef@antiphishing.org))
- Can the same processing model work for other sharing projects?

```
<?xml version="1.0" encoding="UTF-8"?>
<IODEF-Document xmlns="urn:ietf:params:xml:ns:iodef-1.0" lang="en-US">
  <Incident purpose="reporting" restriction="default">
    <IncidentID name="internetidentity.com">192620</IncidentID>
    <ReportTime>2006-11-03T16:32:07-08:00</ReportTime>
    <Assessment>
      <Confidence rating="high" />
    </Assessment>
    <Contact type="organization" role="creator">
      <ContactName lang="en-US">Internet Identity</ContactName>
      <Timezone>-08:00</Timezone>
    </Contact>
    <EventData>
      <AdditionalData dtype="xml">
        <PhraudReport xmlns="urn:ietf:params:xml:ns:iodef-phish-1.0" FraudType="phishemail">
          <FraudParameter>http://www.suntrust.com.ibswebsuntrust.cmserver.minuer.cc/sc/welcome/confirm.cfm.htm</FraudParameter>
          <FraudedBrandName>SunTrust</FraudedBrandName>
          <LureSource>
            <System xmlns="urn:ietf:params:xml:ns:iodef-1.0" category="source">
              <Node>
                <Address>unknown</Address>
                <NodeName>unknown</NodeName>
              </Node>
            </System>
          </LureSource>
          <OriginatingSensor OriginatingSensorType="human">
            <FirstSeen>2006-11-02T17:51:22-08:00</FirstSeen>
            <System xmlns="urn:ietf:params:xml:ns:iodef-1.0">
              <Node>
                <NodeName>www.internetidentity.com</NodeName>
              </Node>
              <Description>Internet Identity SHARC</Description>
            </System>
          </OriginatingSensor>
          <DCSite DCType="web">
            <DCSiteData DCSiteType="web">
              <SiteURL>http://www.suntrust.com.ibswebsuntrust.cmserver.minuer.cc/sc/welcome/confirm.cfm.htm</SiteURL>
            </DCSiteData>
          </DCSite>
        </PhraudReport>
      </AdditionalData>
    </EventData>
  </Incident>
</IODEF-Document>
```



# **RSA<sup>®</sup>CONFERENCE 2007**



## **Lessons Learned**

# What we learned...

- Writing a standard against a moving target is hard
- Target audience and platform remains ill-defined
- Presentation and update semantics are difficult
  - Reports get updated (a lot)
  - Many non-technical people look at reports
- Consensus on data model easier than the transport protocol
- Things are still missing
  - Common taxonomies and terminology
  - Completeness of forensics information





# **RSA<sup>®</sup>CONFERENCE 2007**



Thank You

