RSA®CONFERENCE
CHINA 2012
RSA信息安全大会2012

THE GREAT CIPHER

MIGHTIER THAN THE SWORD
伟大的密码胜于利剑

Sharing Threat Intelligence Analytics for Collaborative Attack Analysis

Samir Saklikar
RSA, The Security Division of EMC

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012

Agenda

- Advanced Targeted Threats & Challenges
- Need for Collaboration and Threat Intelligence Sharing
  - Existing Standards
- Limitations in sharing incident analysis ***process*** details
- Proposals – Extend Threat Intelligence Sharing with
  - Machine-based Analytics Representation
    - Leverage existing standards
  - Human Analyst Actions Representation
    - Propose new standards
- Conclusions

The Attack

- Advanced Targeted Threats
  - Determined Cyber Adversaries
  - Custom Malware, 0-days, Social Engineering
  - Low-and-Slow Multi-Stage Lateral Movement
  - Diverse Concurrent Attack Vectors
  - P2P Encrypted C&C activity
  - Hidden in plain-sight (http, social media)

The Target

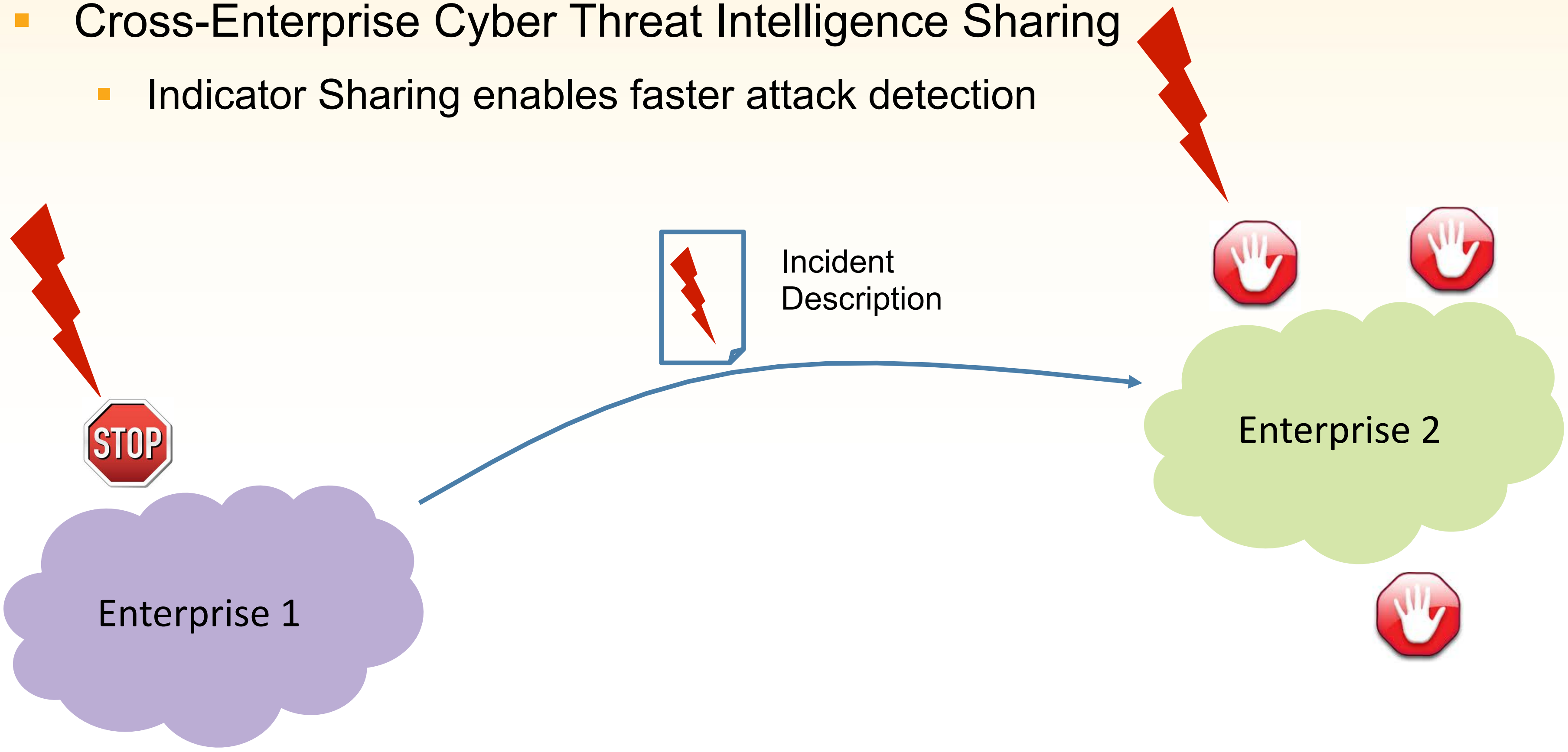
- Evolving and Complex IT Landscape
  - Movement to the Cloud
    - Large interdependent stacks, Newer points of attack insertion
  - More Layers in the IT stack
    - Virtualization (Server/Network)
    - Mobile Clients – “Bring Your Own Device”
    - More Layers → More Logs
  - Newer Security Data sources
    - Netflow, Full Packet Capture, Sandbox Indicators

The Defense

- The Tools
  - Intrusion Detection
    - Host and Endpoint-based tools
  - Security Incident Event Management
  - Vulnerability Scanners
  - Memory/Disk Analysis
- The Expertise
  - CIRT/SOC teams overburdened
  - Lack of sufficient in-house expertise
    - Malware Analysis, Network Intrusion Detection, Remediation

Collaboration is the key

- Cross-Enterprise Cyber Threat Intelligence Sharing
  - Indicator Sharing enables faster attack detection



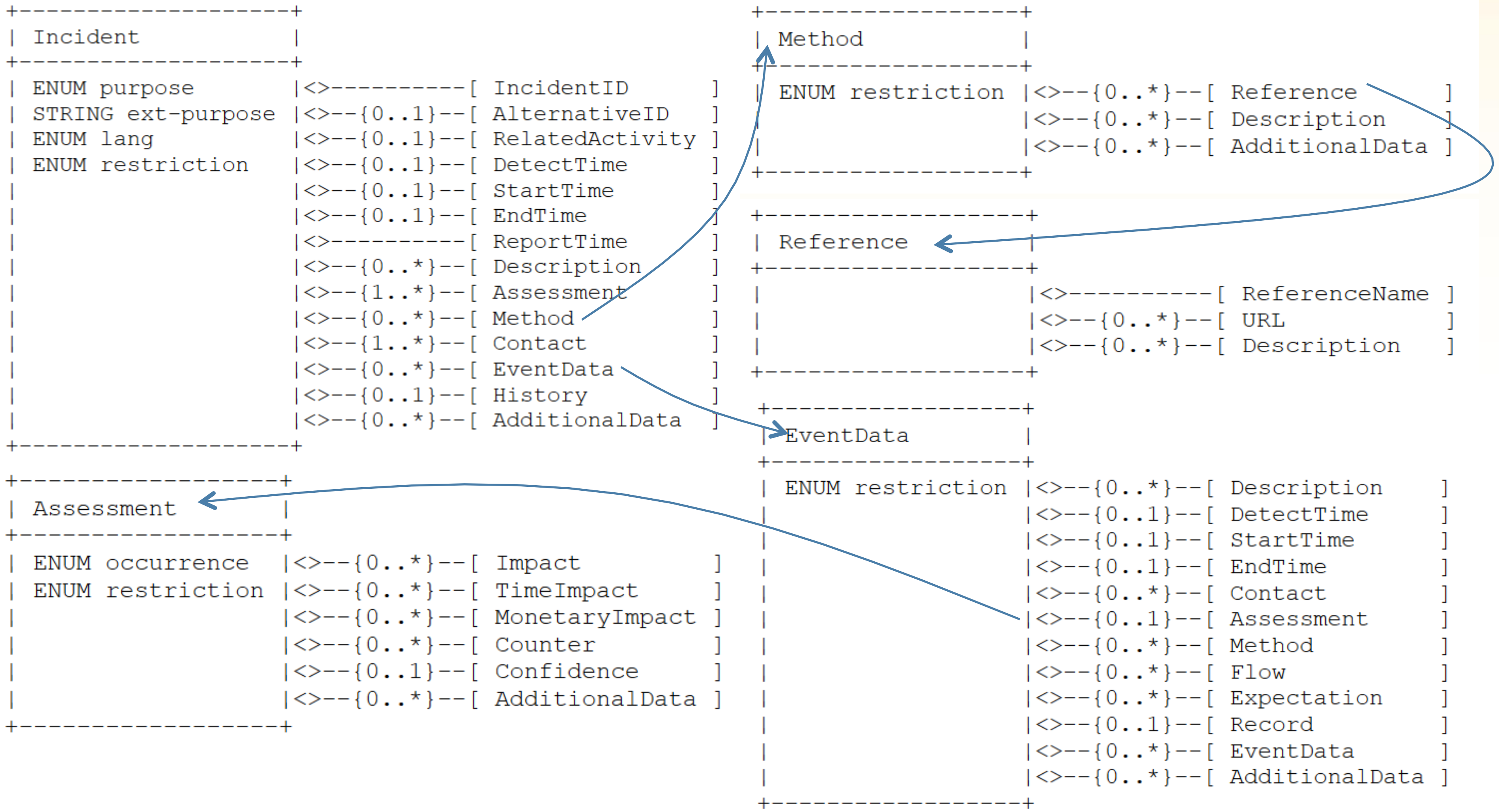
Threat Intelligence Sharing Challenges

- Lack of interoperable standards
- Challenges in validating data quality and reliability
- Risk of information leakage
- Untested methods for governing use of sensitive information
- Shortage of skilled security expertise
- Legal and Data confidentiality requirements

Standards to the Rescue

- Incident Object Description Exchange Format (IODEF)
  - Encoding Threat Information
- Real-time Internetwork Defense (RID)
  - "Envelopes" for IODEF and other electronic incidents
- Malware Attribute Enumeration and Classification
  - Describe malware and its behavior
- Common Attack Pattern Enumeration and Classification (CAPEC)
  - Common methods for subverting software etc
- Cyber Observable eXpression (CybOX)
  - Any observable event; Abstract schemas of activity
- Advanced Forensics Framework (AFF4)
- Security Content Automation Protocol (SCAP)

IODEF - Data Model



```
+--------------------+
| Incident           |
+--------------------+
| ENUM purpose       |<>----------[ IncidentID      ]
| STRING ext-purpose |<>--{0..1}--[ AlternativeID   ]
| ENUM lang          |<>--{0..1}--[ RelatedActivity ]
| ENUM restriction   |<>--{0..1}--[ DetectTime      ]
|                    |<>--{0..1}--[ StartTime       ]
|                    |<>--{0..1}--[ EndTime         ]
|                    |<>----------[ ReportTime      ]
|                    |<>--{0..*}--[ Description     ]
|                    |<>--{1..*}--[ Assessment      ]
|                    |<>--{0..*}--[ Method          ]
|                    |<>--{1..*}--[ Contact         ]
|                    |<>--{0..*}--[ EventData       ]
|                    |<>--{0..1}--[ History         ]
|                    |<>--{0..*}--[ AdditionalData  ]
+--------------------+
```

```
+------------------+
| Assessment       |
+------------------+
| ENUM occurrence  |<>--{0..*}--[ Impact         ]
| ENUM restriction |<>--{0..*}--[ TimeImpact     ]
|                  |<>--{0..*}--[ MonetaryImpact ]
|                  |<>--{0..*}--[ Counter        ]
|                  |<>--{0..1}--[ Confidence     ]
|                  |<>--{0..*}--[ AdditionalData ]
+------------------+
```

```
+------------------+
| Method           |
+------------------+
| ENUM restriction |<>--{0..*}--[ Reference      ]
|                  |<>--{0..*}--[ Description    ]
|                  |<>--{0..*}--[ AdditionalData ]
+------------------+
```

```
+------------------+
| Reference        |
+------------------+
|                  |<>----------[ ReferenceName ]
|                  |<>--{0..*}--[ URL           ]
|                  |<>--{0..*}--[ Description   ]
+------------------+
```

```
+------------------+
| EventData        |
+------------------+
| ENUM restriction |<>--{0..*}--[ Description    ]
|                  |<>--{0..1}--[ DetectTime     ]
|                  |<>--{0..1}--[ StartTime      ]
|                  |<>--{0..1}--[ EndTime        ]
|                  |<>--{0..*}--[ Contact        ]
|                  |<>--{0..1}--[ Assessment     ]
|                  |<>--{0..*}--[ Method         ]
|                  |<>--{0..*}--[ Flow           ]
|                  |<>--{0..*}--[ Expectation    ]
|                  |<>--{0..1}--[ Record         ]
|                  |<>--{0..*}--[ EventData      ]
|                  |<>--{0..*}--[ AdditionalData ]
+------------------+
```

Existing Threat Intelligence Sharing

- Answering questions such as
  - What was the attack
  - When did it happen
  - Where was it found
  - What does the attack look like
  - Who found it
  - How is it affecting the environment
  - How quickly was it solved
  - What was the impact
  - What was the surrounding context
  - ...

Opportunities - Extend Indicator Sharing to

- Convey Indicator Identification Process
  - How was the Indicator identified Machine-based Analytics, Analyst Expertise
  - Which analytics worked better and why? Analyst Opinion, Comparative Results
  - What changed which helped in attack identification? Analyst Strategy
  - What was the confidence level in the indicator? Analyst Opinion
- Validation of Indicator Authenticity
  - No means of conveying indicator authenticity with backing data Supporting Data Sets
- Guidelines for Indicator Portability
  - Porting Indicator detection requires mostly human consumption of threat feed Mapping between Analytics and Environment
- Composition of multiple Indicators – multistage attacks
  - Requires human presence to understand and write higher-level indicators Analytics, Expertise Markup enabling composition

Proposal - Indicator Sharing Extensions

- Extend Indictor Sharing Description with
  - Machine Analytics Representation to
    - Describe which and how machine-based analytics techniques were used to identify the attack
      - For e.g. rule-based, or data-mining or machine-learning techniques
    - Include a sampling of the input data to help in easier portability of machine analytics techniques
  - Analyst Actions Representation to
    - Describe what actions were manually performed by the human analyst to identify the attack
    - How did the analyst interpret the results from machine-based analytics
    - What was the analyst's opinion about the attack

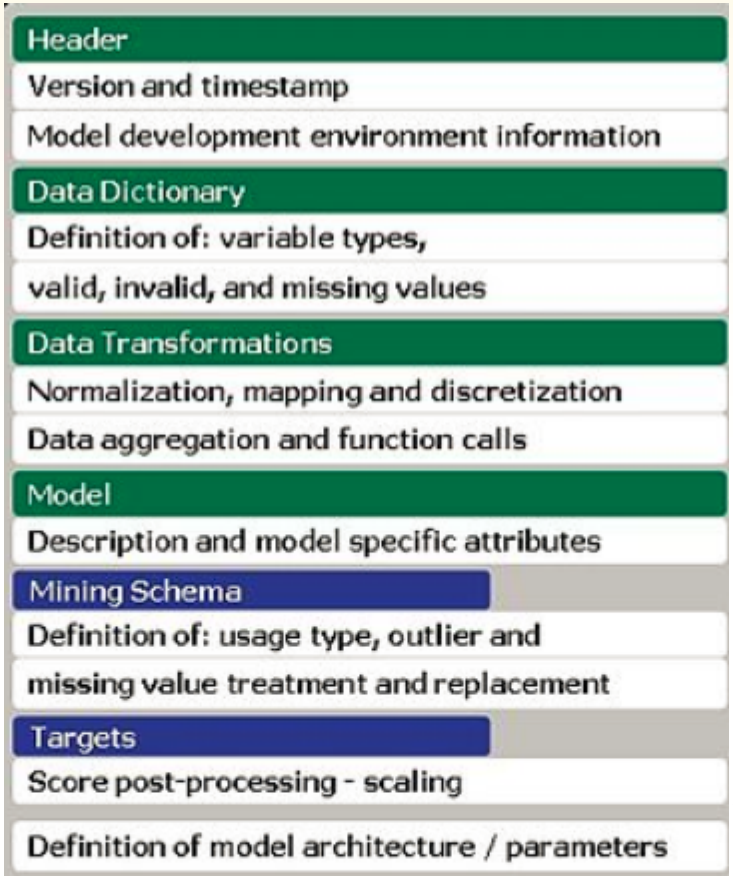
Analytics and Actions Representation

- Machine Analytics Representation
  - Leverage and extend existing standards
  - Predictive Modeling Markup Language (PMML)
    - For representing data-mining and machine learning techniques
- Analyst Actions Representation
  - Develop/Propose new standard

Predictive Modeling Markup Language

- Standardized Representation of mining models and data
- Encompasses the various stages in a typical data-mining/analytics task
  - Data Dictionary definition
  - Data Transformations
  - Handling missing or outlier data values
  - Model Definition
  - Outputs
  - Post-Processing steps
  - Model Explanation
  - Model Verification
- Supported by leading Data analytics tools vendors (commercial and open-source likewise)

PMML - Mapping to Threat Intelligence



Header

Version and timestamp

Model development environment information

Data Dictionary

Definition of: variable types,

valid, invalid, and missing values

Data Transformations

Normalization, mapping and discretization

Data aggregation and function calls

Model

Description and model specific attributes

Mining Schema

Definition of: usage type, outlier and

missing value treatment and replacement

Targets

Score post-processing - scaling

Definition of model architecture / parameters

Description of which security event data was analyzed

May leverage CybOX and similar standards

Any pre-processing done by the source enterprise over the event data

The analytics (data mining) model used by the analyst to process the event data

Any specific treatment for missing values etc. performed in the analytics

Any post-processing of the security analytics results

Must match to the Incident data shared in the IODEF object

May leverage CybOX and similar standards

Proposed Extensions to PMML

- Allow incomplete data and mining models for privacy reasons
  - For e.g. Allow Mining models to show only Data Relationships without actual weights.
  - Enables sharing the relevant security event data which was used in the incident detection, but NOT how it is related to the sharing enterprise
- Allow wild-carded/pattern-matched data-model and mining-model representations
  - Enables recipient enterprise to leverage the mining model to their own enterprise network architecture
- Enable versioning of the shared data and mining-model
  - Enables the source organization to share multiple versions of the data and mining-model over time.
  - Enables the recipient organization to learn the evolution of the mining model and make suitable changes to self network model
- Allow Model Filter templates – typically intelligence sharing handled via a separate sub-org

Machine-based Analytics not enough

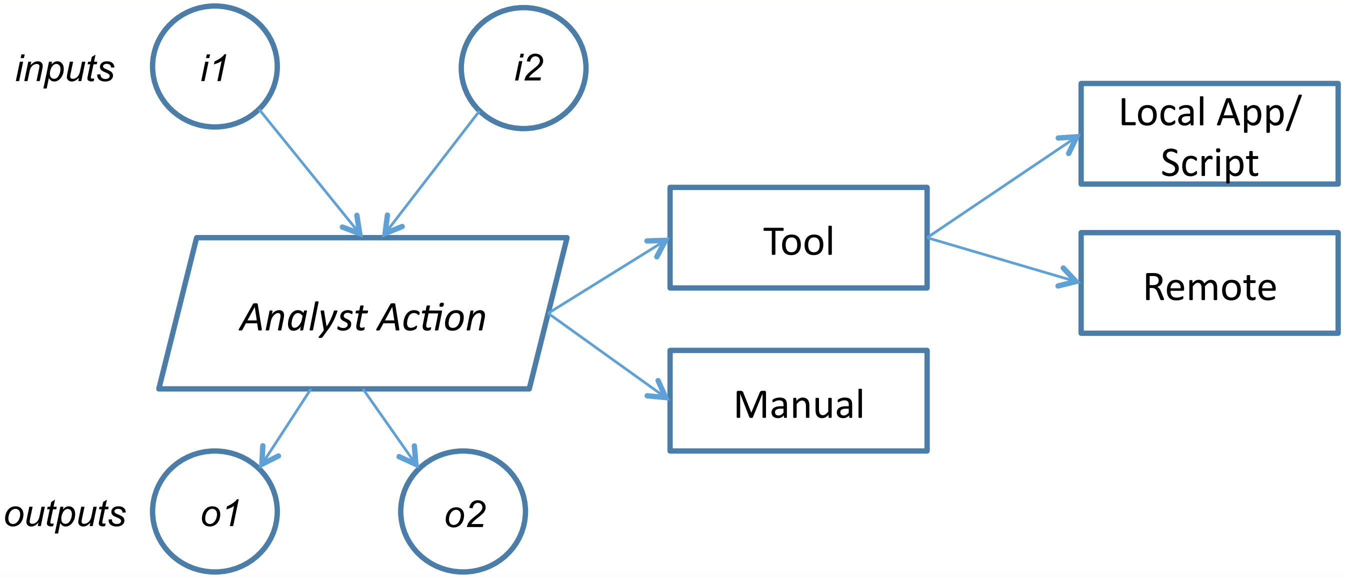
- Security Analysts use a variety of tools and processes for Incident Analysis
  - IODEF and proposed Machine Analytics extensions can convey tools information
- Yet, Incident Analysis process is intricately complex, requiring human intelligence and a trial-and-error methods at times
  - Human Expertise needed for “Connecting the Dots”
  - Discontinuous, brittle and human-coupled Analytics chain
- Sharing standardized Machine Analytics information helps but not enough
  - Need for sharing Analysts Actions over Threat intelligence feeds

Analyst Actions Representation

- Monitor, Log and Report on Analyst actions while handling a particular incident
  - Relevant monitoring, and logging tools deployed on analyst workstation
- Monitored Analyst actions can include
  - Analyst interactions with the workstation (keyboard inputs, clicks etc)
  - Network interactions data (server access, downloads, network tools)
  - Interactions with local or remote applications used in Incident Analysis
- Proposal
  - Create multiple Analyst Action Charts for each analyst working on a particular incident
  - Outputs a single final Action Chart which collates the various actions performed by the analysts while handling the incident

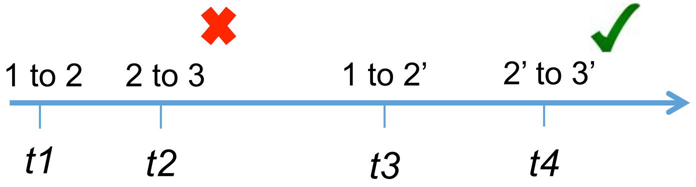
Analyst Action Chart Data Model

- Each Analyst action/step captured with
  - Tools/Process description used in the step
  - Process may be visual interpretation by human analyst
  - Inputs to the tools/process
  - Outputs of the tools/process
  - Pre/Post conditions of the step

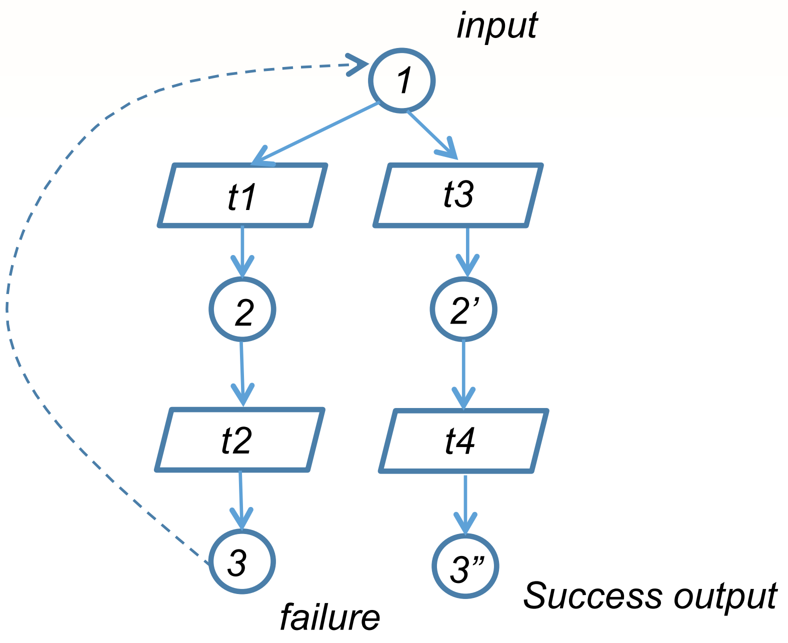


Analyst actions correlation

- Individual Steps are correlated; Output of previous step = Input of next step
- Analyst Activities monitored in time-sequence but may result in dead ends
- Failure paths result in dead ends in the graph structure
- Show success paths from inputs to final incident analysis output



Analyst Actions on input 1 to reach output 3'



Analyst Activity Chart Annotations

- Analyst Annotations
  - Human Inference of results (reasoning towards a particular conclusion)
  - Significant meta-data about outputs
    - IP Addresses, Strings, Files/Certs extracted, Signature of Author etc.
  - Distinguishing behavior signature for identifying the APT
  - Distinguishing binary signature for malware (used by APT)
  - Opinion of Attack Attribution

Support different Analysis types

- Automated (Machine-based Analytics)
  - Link to machine-based analytics (data mining, sandboxing results etc.)
- Semi-automated activities (Human + Tool-based Analytics)
  - Extract results from tool and perform some processing
  - For e.g. searching IDB files for memory dump files
  - Can by logged by monitoring human activities
- Manual activities (entirely Human-based activities)
  - Visual interpretation by human analyst of previous results
  - Human reasoning
  - Human-coded tasks such as setting break-points, identifying strings, decoding encryption/decryption routines, identifying hooking process
  - May need human assisted annotations

Conclusion

- Need for richer threat intelligence sharing
- Machine Analytics and Analyst Actions representations
- Complete picture of Incident Description, Identification and Analysis
- Proposals as IODEF extensions, leverage PMML standards

Thank You

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012