RSA®CONFERENCE
ASIA PACIFIC 2013

Security in knowledge

SHARING THREAT INTELLIGENCE ANALYTICS FOR COLLABORATIVE ATTACK ANALYSIS

Samir Saklikar

RSA, The Security Division of EMC

Session ID: CLE-T05

Session Classification: Intermediate

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

State of Cyber Security

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)

v/s

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

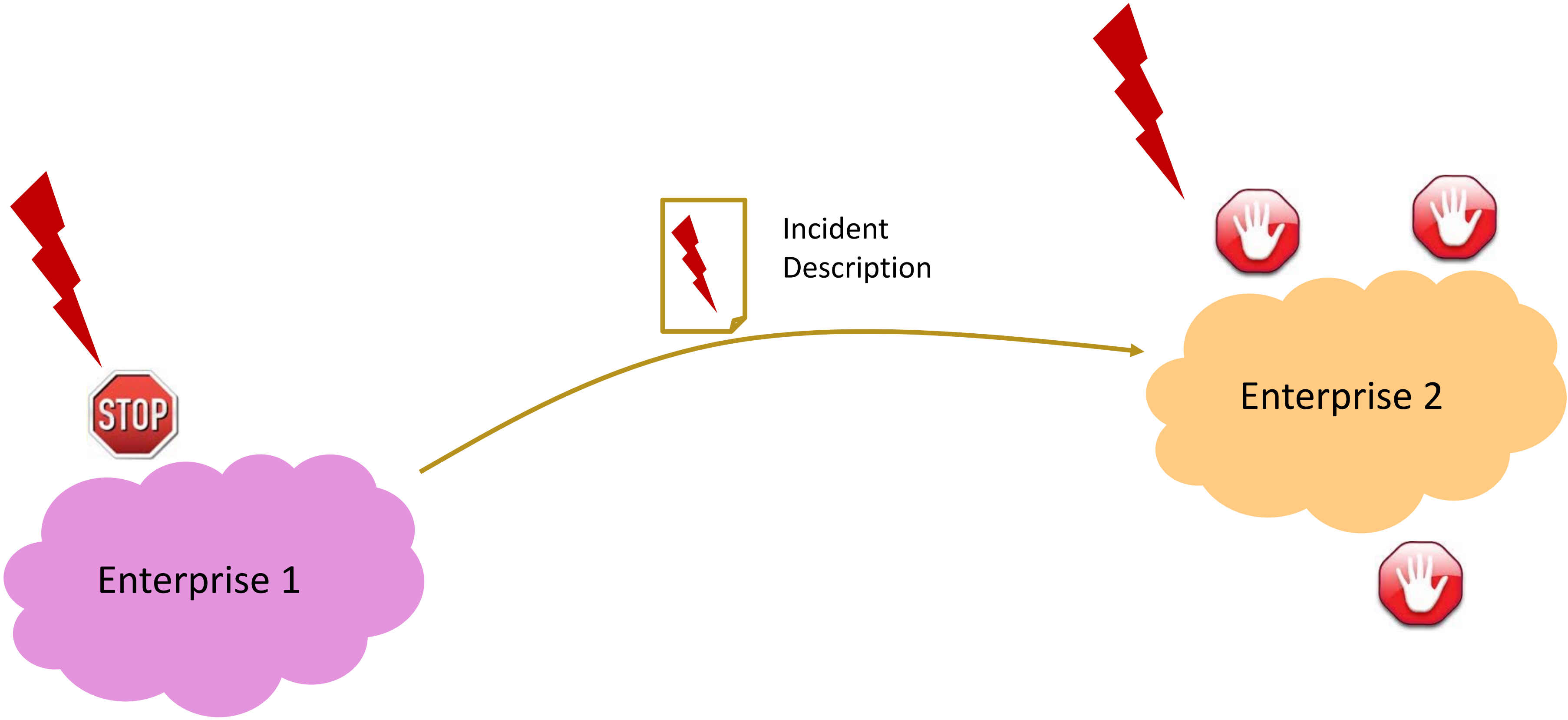
State of Cyber Defense

- The Tools
  - Intrusion Detection
    - Host and Endpoint-based tools
  - Security Incident Event Mgmt.
  - Security Analytics

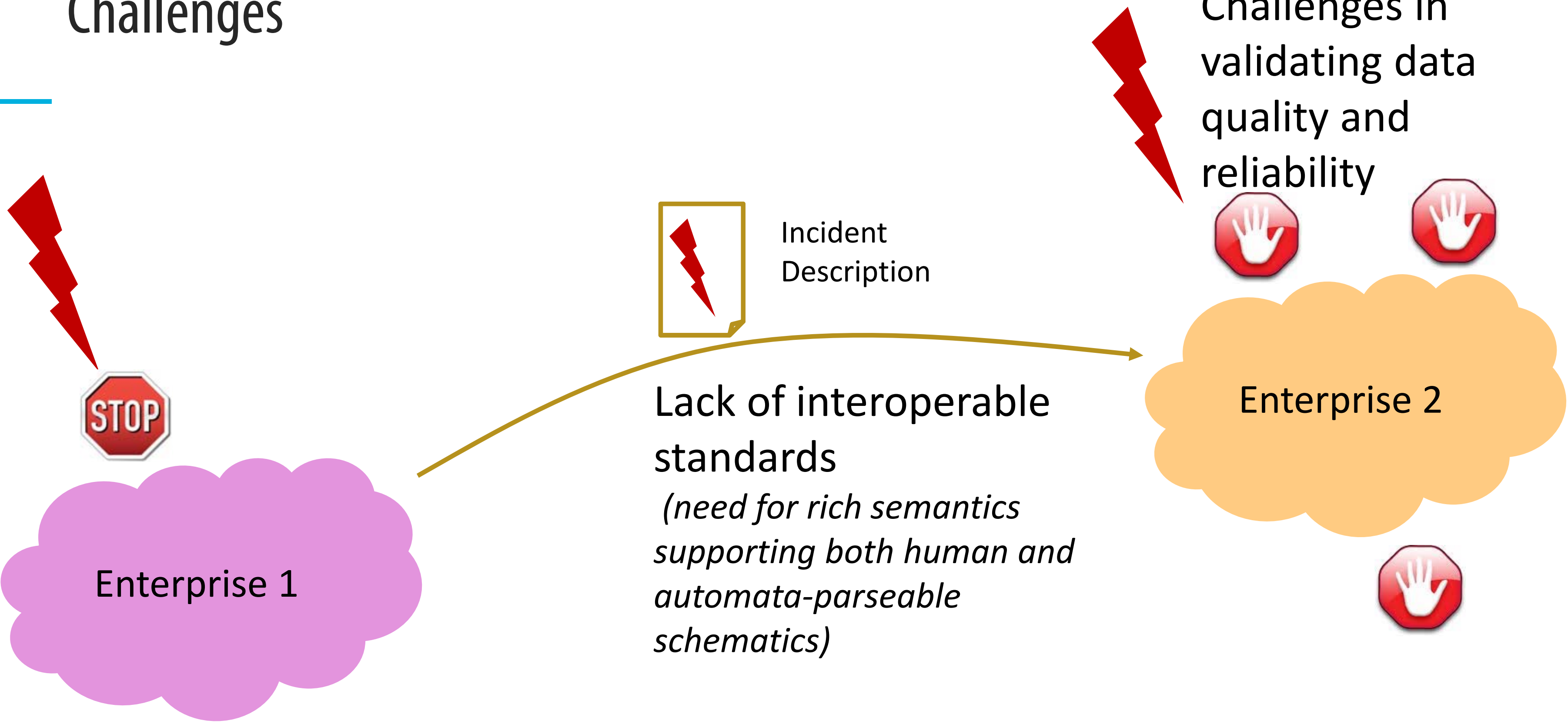
- The Expertise
  - CIRT/SOC teams overburdened
  - Lack of sufficient in-house expertise
    - Malware Analysis, Network Intrusion Detection, Remediation

Collaboration is the key

- Cross-Enterprise Cyber Threat Intelligence Sharing



Challenges



Challenges in validating data quality and reliability

Lack of interoperable standards

(need for rich semantics supporting both human and automata-parseable schematics)

Risk of information leakage

(usability of shared intelligence v/s risk of potential security posture compromise)

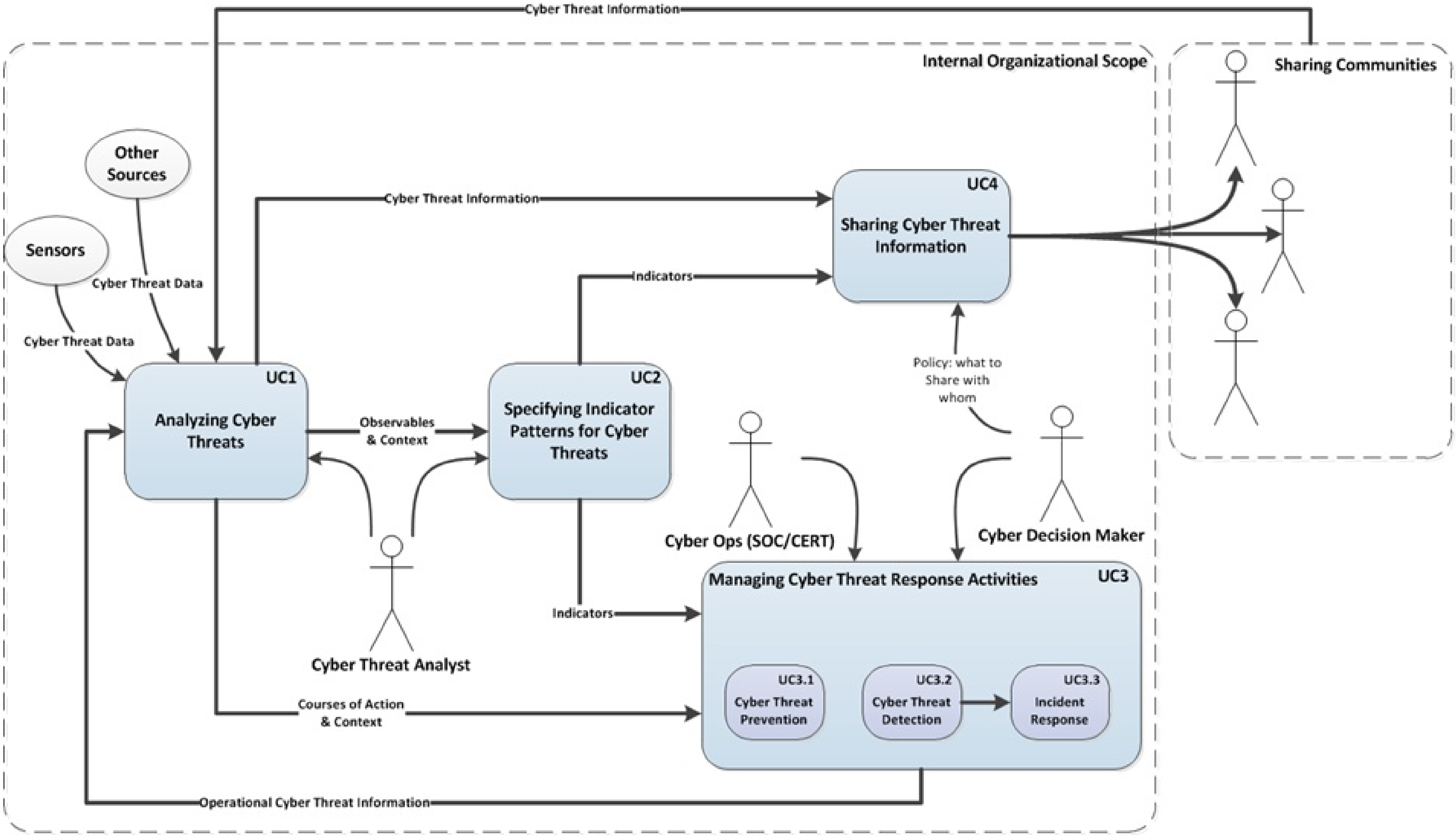
- Untested methods for governing 3rd party use of sensitive information
- Shortage of skilled security expertise
- Legal and Data confidentiality requirements

Standards/Specifications to the Rescue

- IETF
  - Incident Object Description Exchange Format (IODEF)
  - Real-time Internetwork Defense (RID)

- MITRE
  - Trusted Automated eXchange of Indicator Information (TAXII)
  - Structured Threat Information eXpression (STIX)
  - Malware Attribute Enumeration and Classification (MAEC)
  - Common Attack Pattern Enumeration and Classification (CAPEC)
  - Cyber Observable eXpression (CybOX)

Threat Intelligence Sharing Use-Cases

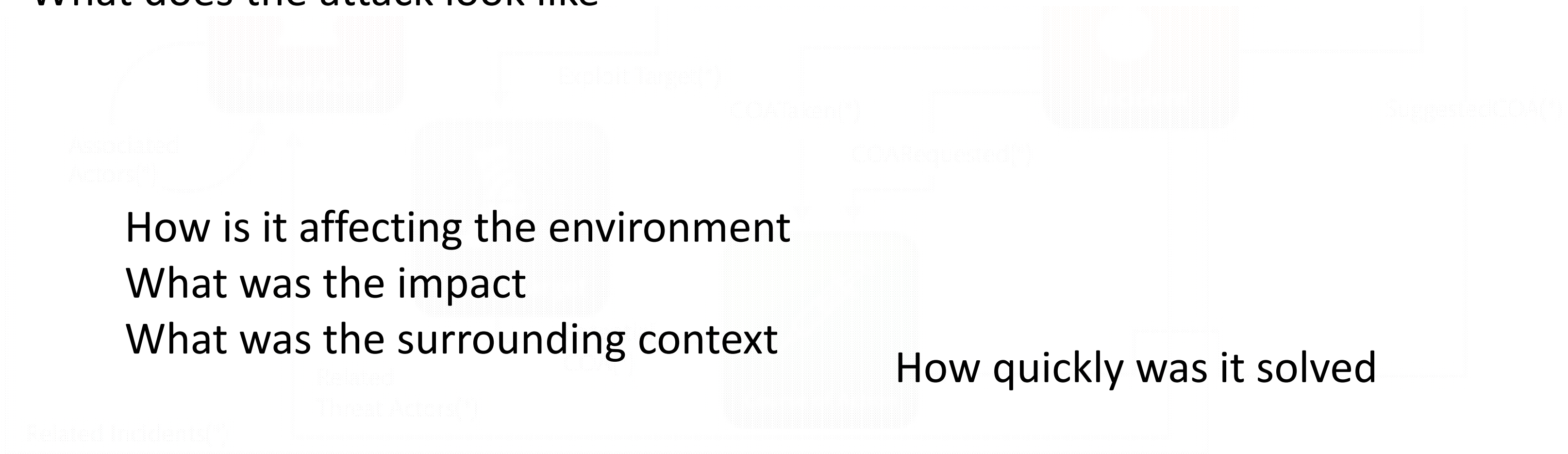


** from STIX Use-cases document*

Threat Intelligence Sharing – What questions does it answer?

What was the attack
When did it happen
Where was it found

What does the attack look like



How is it affecting the environment
What was the impact
What was the surrounding context

How quickly was it solved

** from STIX Architecture document*

Opportunities to extend Threat Intelligence indicators

Richer Indicator/TTP Semantics

How was the Indicator identified

Which analytics worked better and why?

What changed which helped in attack identification?

What was the confidence level in the indicator?

Provable validation of Indicator Authenticity to recipient organization

Guidelines for Indicator Portability

** from STIX Architecture document*

Proposal – Extend Indictor Sharing Description with

- Machine Analytics Representation to
  - Describe analytics techniques used
    - For e.g. rule-based, or data-mining or machine-learni techniques
  - Include a sampling of the input data to help in validation and portability

Leverage existing standards such as PMML

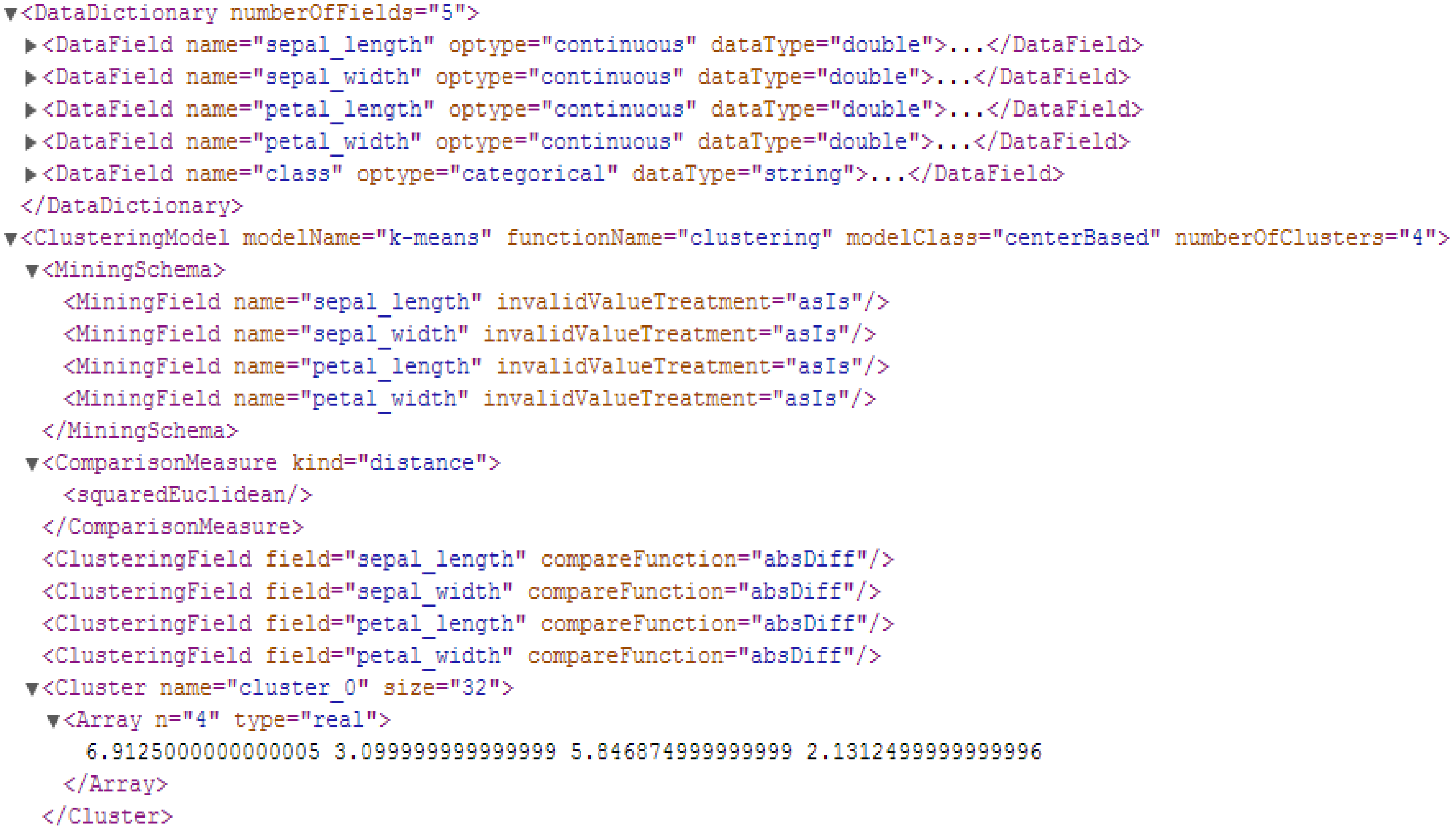
- Analyst Actions Representation to
  - Describe actions performed by the human analyst
  - Describe analyst's interpretation of machine analytics

Propose new extensions

Predictive Modeling Markup Language

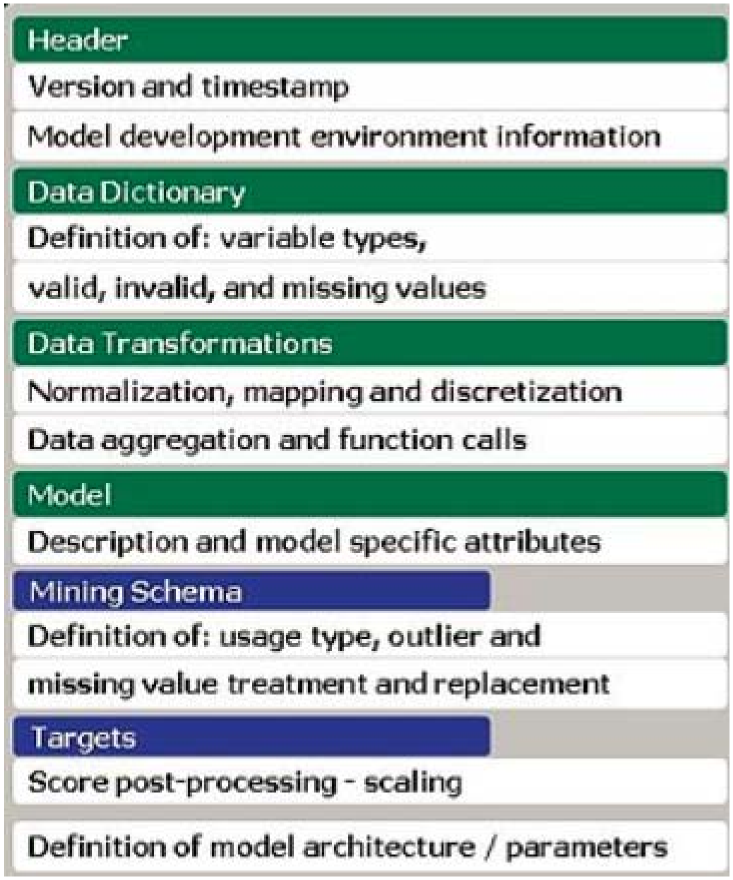
- Standardized Representation of mining models and data
- Encompasses the various stages in a typical data-mining/analytics task
  - Data Dictionary definition
  - Data Transformations
  - Handling missing or outlier data values
  - Model Definition
  - Outputs
  - Post-Processing steps
  - Model Explanation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

PMML Example (from www.dmg.org)



```
▼<DataDictionary numberOfFields="5">
  ▶<DataField name="sepal_length" optype="continuous" dataType="double">...</DataField>
  ▶<DataField name="sepal_width" optype="continuous" dataType="double">...</DataField>
  ▶<DataField name="petal_length" optype="continuous" dataType="double">...</DataField>
  ▶<DataField name="petal_width" optype="continuous" dataType="double">...</DataField>
  ▶<DataField name="class" optype="categorical" dataType="string">...</DataField>
 </DataDictionary>
▼<ClusteringModel modelName="k-means" functionName="clustering" modelClass="centerBased" numberOfClusters="4">
  ▼<MiningSchema>
     <MiningField name="sepal_length" invalidValueTreatment="asIs"/>
     <MiningField name="sepal_width" invalidValueTreatment="asIs"/>
     <MiningField name="petal_length" invalidValueTreatment="asIs"/>
     <MiningField name="petal_width" invalidValueTreatment="asIs"/>
   </MiningSchema>
  ▼<ComparisonMeasure kind="distance">
     <squaredEuclidean/>
   </ComparisonMeasure>
   <ClusteringField field="sepal_length" compareFunction="absDiff"/>
   <ClusteringField field="sepal_width" compareFunction="absDiff"/>
   <ClusteringField field="petal_length" compareFunction="absDiff"/>
   <ClusteringField field="petal_width" compareFunction="absDiff"/>
  ▼<Cluster name="cluster_0" size="32">
    ▼<Array n="4" type="real">
       6.9125000000000005 3.099999999999999 5.846874999999999 2.1312499999999996
     </Array>
   </Cluster>
```

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
- Allow wild-carded model representations
- Enable versioning of the shared data and mining-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
  - 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

- Need for sharing Analysts Actions over Threat intelligence feeds

Analyst Actions Representation

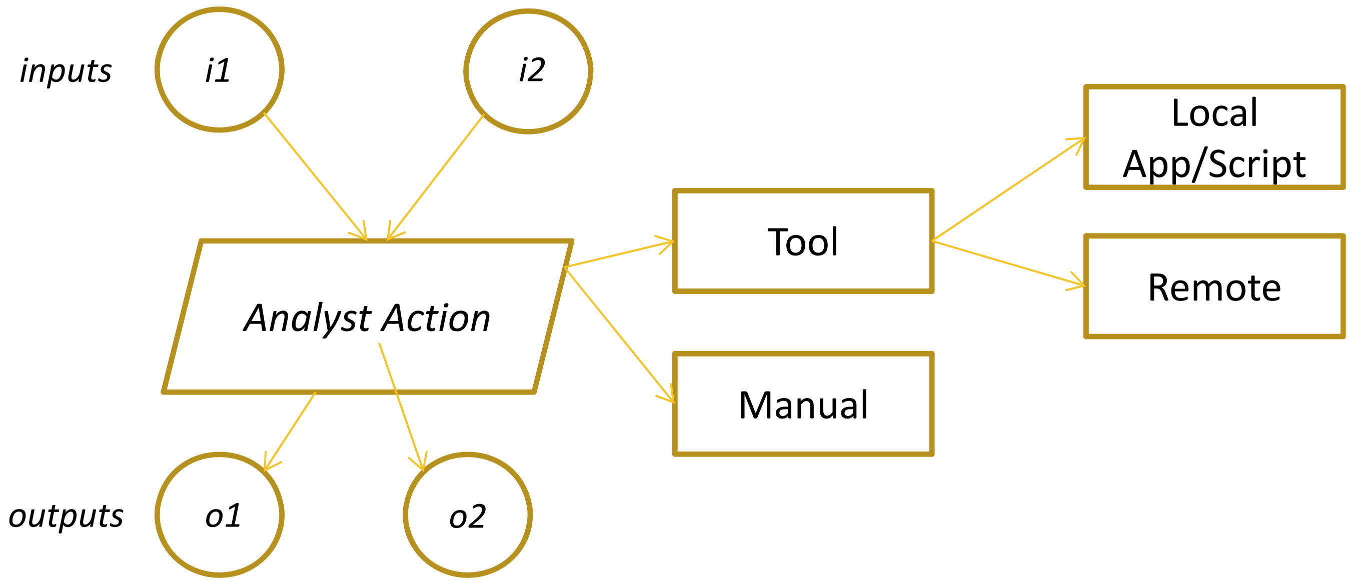
- Monitor, Log and Report on Analyst actions while handling a particular incident
  - Relevant monitoring, and logging tools deployed on analyst workstation

- 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

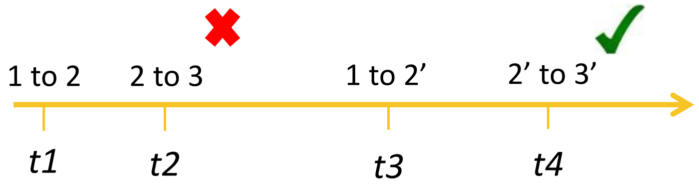
Analyst Action Chart Data Model

- 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

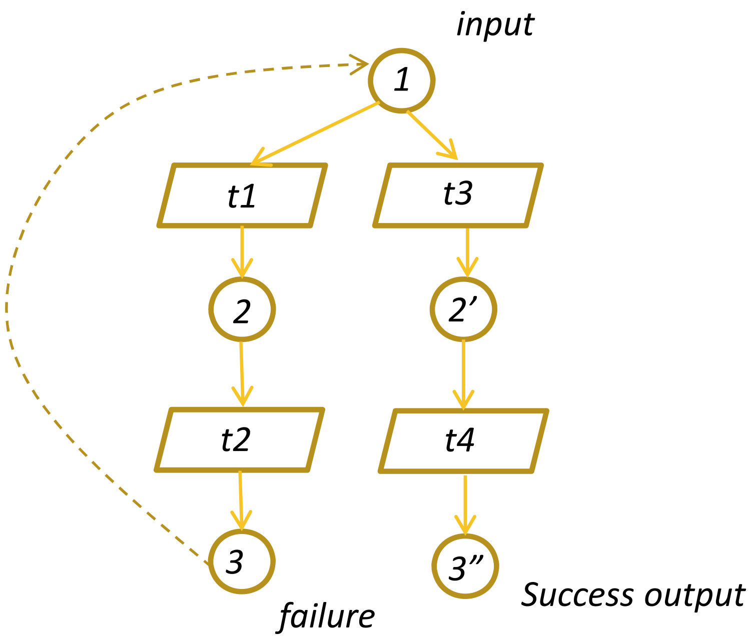


Analyst actions correlation

- 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

Example Usage – Spear Phishing

- STIX Representation (* STIX Use-Cases document)

```
<cybox:Observable ..>
  <cybox:StatefulMeasure ..>
    <cybox:DefinedObject .. xsi:type=“EmailMessage.obj..”>
      <EmailMessage.Obj:Header> {attachments,recipient,from,subject..}
    <cybox:DefinedObject .. xsi:type=“FileObj..”>
      <FileObj> {Name,extension, size, hash..}
    <cybox:DefinedObject .. xsi:type=“URIObj..”>
      <URIObj> {URL,DomainName..}
      <cybox:RelatedObj> {WHOIS,DNSQuery,DNSRecord,IPAddress,URLs}
    <cybox:DefinedObject .. xsi:type=“DNSQuery..”>
      <DNSQuery> {Qname,Qtype,Qclass, Question, Answer..}
    <cybox:DefinedObject .. xsi:type=“DNSRecord..”>
      <DNSRecord> {Address Object, Resolved to..}
    <cybox:DefinedObject .. xsi:type=“WHOIS..”>
      <WHOIS> {URI Obj..}
    ...
```

Analyst Actions

```
<cybox:Observable ..>
  <cybox:StatefulMeasure ..>
    <cybox:DefinedObject .. xsi:type=“CmdLineObj..”>
      <CmdLineObj> {shell,command,time,parameters,pipes,..}
    <cybox:DefinedObject .. xsi:type=“GUIActionObj..”>
      <CmdLineObj> {GUIApp,time, click position,key-press..}
    <cybox:DefinedObject .. xsi:type=“FileObj..”>
      <FileObj> {Name,extension, size, hash..}
  <cybox-ext:AnalysisMeasure>
    <cybox:DefinedObject .. Xsi:type=“AnalystActivityObjList”>
      <AnalystActivityObjList>
        <AnalystActivityObj>
          <src xsi:type=“CmdLineObj” id=“9097123123..”>
          <dst xsi:type=“FileObj” id=“8712313..”>
        <AnalystActivityObj>
          <src xsi:type=“FileObj” id=“8712313..”>
          <dst xsi:type=“GUIActionObj” id=“67823232”>
```

Machine-based Analytics

```
<cybox:Observable ..>
<cybox-ext:AnalysisMeasure>
    <cybox:DefinedObject .. Xsi:type="MachineAnalyticsObj">
        <MachineAnalyticsObj>
          <PMML>
            <DataDictionary>
              <DataField name="RecipientSubOrgObj">
              <DataField name="FromAddress">
                {size,attachments,time etc}
              <DataField name="LDATopic1" optype="categorical">
                <value="urgent", "europe", "opportunity", "millions"..
              <DataField name="LDATopic2" optype="categorical">
                <value="escalation","customer","bugreport"..
            <ClusteringModel modelName="k-means" functionName"..">
              <MiningSchema>
              <ComparisonMeasure>
              <Cluster name="cluster1">
              <Cluster name="cluster2">
```

Conclusion

- Need for richer indicator semantics description
- Need for Machine Analytics and Analyst Actions representations

- Leverage PMML and proposed analyst actions for Incident description, identification and analysis representation
- Opportunity for IODEF/STIX extensions

Questions?