Center for Configuration Analytics and Automation

An NSF Industry/University Cooperative Research Center

STIX Analytics-- From Threat Information Sharing to Automated Response

Secure and Resilient Cyber Ecosystem Industry Workshop Presentation for DHS

Dr. Ehab Al-Shaer (PI), Dr. Bill Chu (PI)

University of North Carolina Charlotte

Ealshaer, billchu@{uncc.edu,ccaa-crc.org}

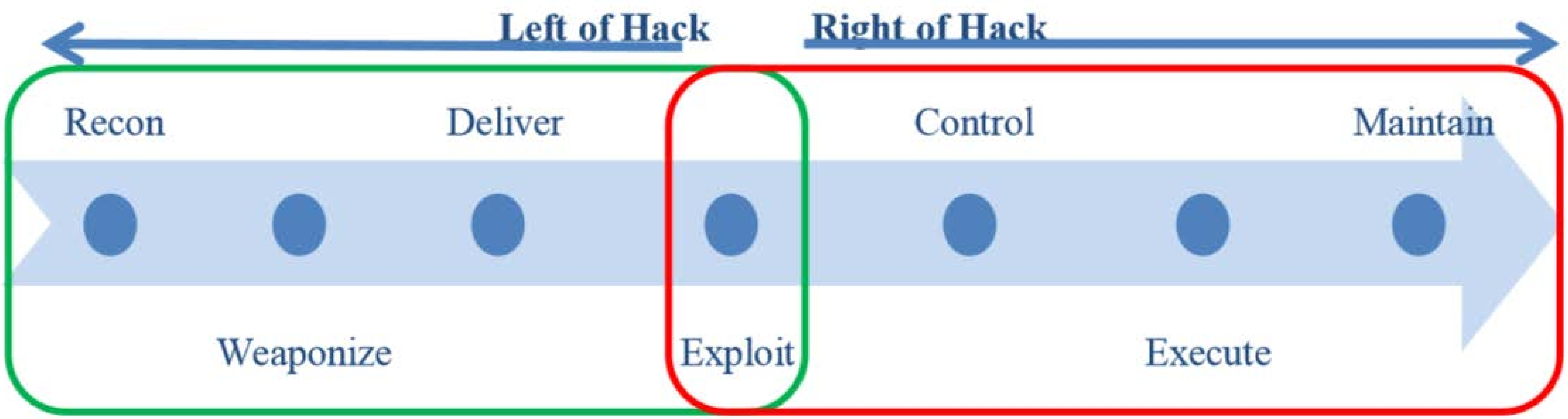
Agenda Outline

- Motivation of Cyber Threat Information (CTI) Sharing
- Background – STIX (if needed)
- Challenges to Effective CTI Sharing
- Our Research Proejcts/Directions
  - STIXChecker
    - Logical Formalization of STIX (OWL/SMT) and configurations
    - Impact analysis
  - STIXAnalytics
    - Determining Network Relevance
    - Visual Analytics
    - Reputation Analysis
  - ThreatMitigation
    - Impact Analysis: Killing the Cyber Kill-Chain
    - From STIX to Actions

What is STIX

- A language to *specify, capture, characterize and communicate* **Cyber Threat Information**.
- A standardized and structured way to represent threat information
- Both human readable and machine parsable.
- Built upon active participation and feedback from a broad spectrum of organizations and experts linked with government, academia and industry.
- Initial implementation has been done in XML Schema and JSON.
  - Plan to iterate and refine with real-world use

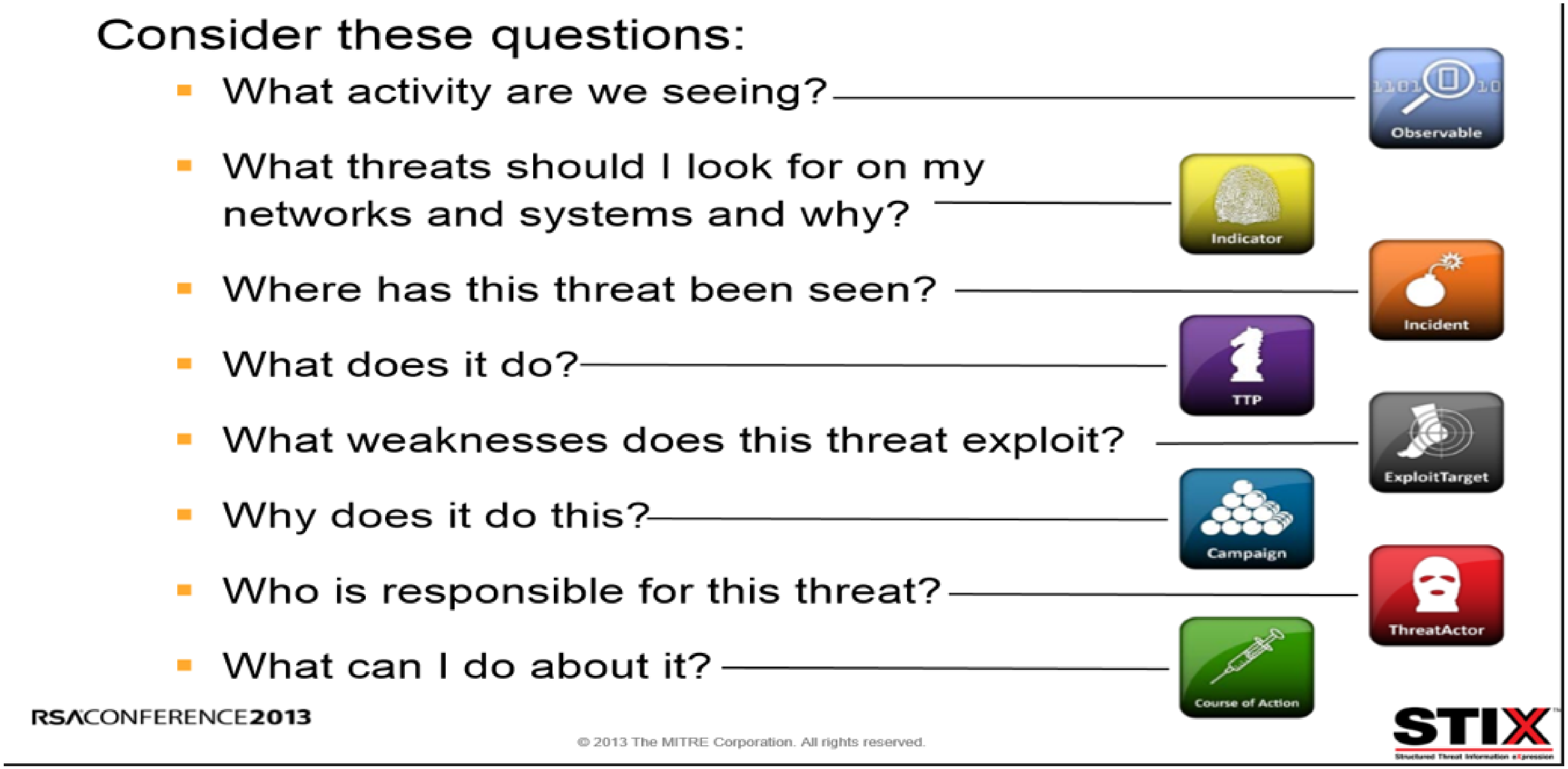
STIX Concept

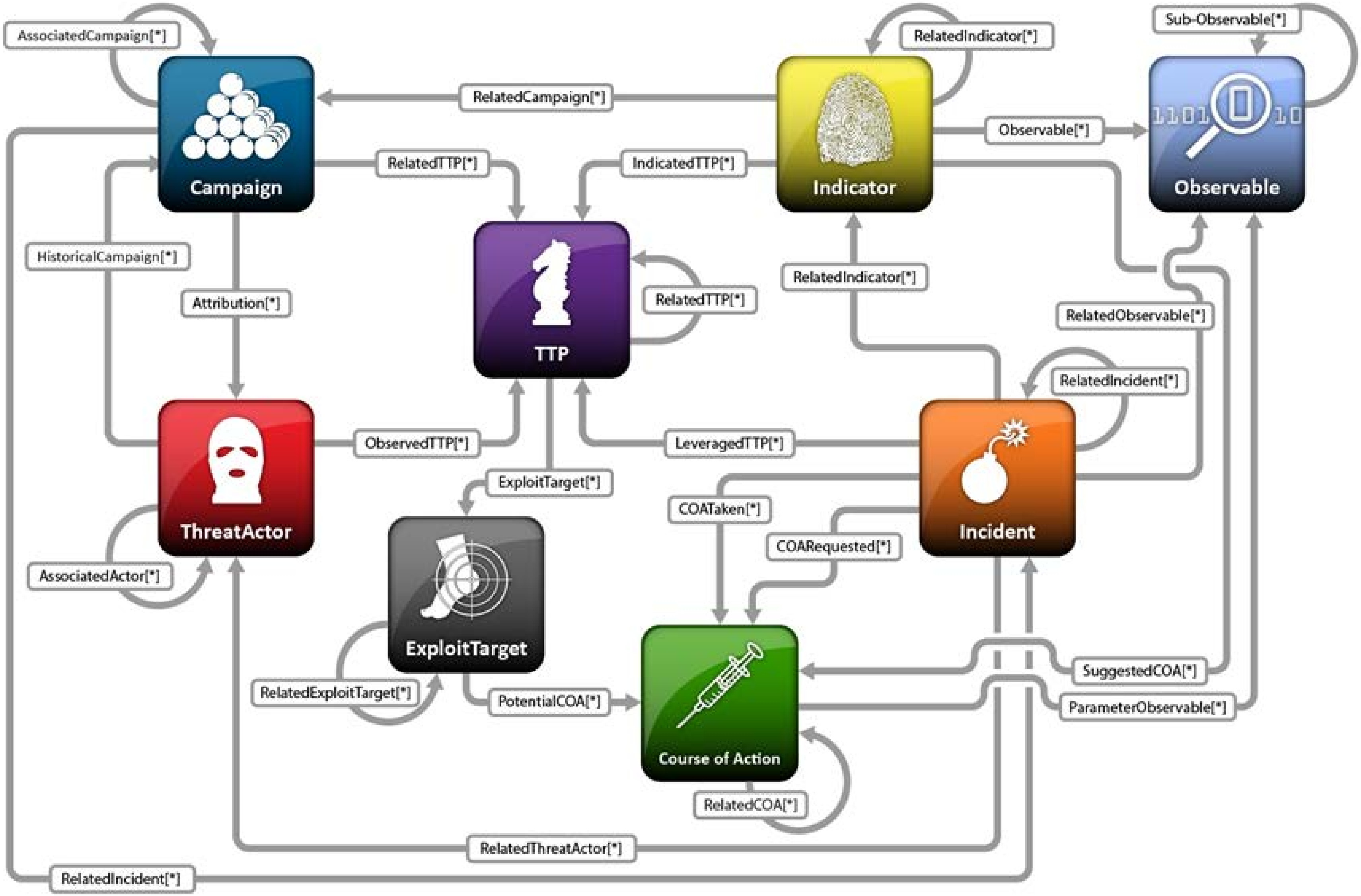


Example Cyber Kill Chain

- Proactive approach to resist adversary, preferably **before the exploit stage**.
- Only possible through adoption of Cyber Threat Intelligence

STIX Embedded with CTI

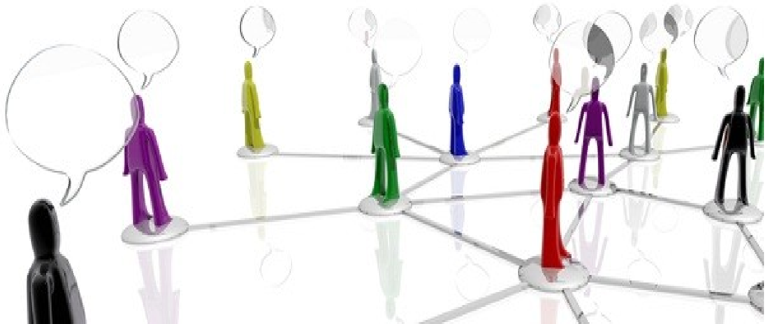



AssociatedCampaign[*]
Campaign
RelatedCampaign[*]
RelatedIndicator[*]
Indicator
Sub-Observable[*]
Observable
Observable[*]
1101
10
RelatedTTP[*]
IndicatedTTP[*]
TTP
RelatedTTP[*]
HistoricalCampaign[*]
Attribution[*]
RelatedIndicator[*]
RelatedObservable[*]
RelatedIncident[*]
ThreatActor
ObservedTTP[*]
LeveragedTTP[*]
Incident
ExploitTarget[*]
COATaken[*]
COARequested[*]
AssociatedActor[*]
ExploitTarget
RelatedExploitTarget[*]
PotentialCOA[*]
Course of Action
SuggestedCOA[*]
ParameterObservable[*]
RelatedCOA[*]
RelatedThreatActor[*]
RelatedIncident[*]

CHALLENGES TO EFFECTIVE CTI SHARING

Challenges

- Intelligence must be actionable otherwise it is useless. [CTI rules]
- Making Threat Intelligence Actionable. [RSA Conference 2015]
- Stix XML leads to interpretability and portability issues .
- Difficult to import as it is in existing analysis tools.
- Implementation independent solution is highly favourable.
- Automated inference and reasoning deficiency in XML.



Challenges

- For the Network Admin (use-case 1)
  – STIX feeds requires extensive analysis to extract elements relevant to the network.
  – Mapping threats to their counter measures is a manual process and lacks cost-benefit and impact analysis.
- For the Cyber-Security Analyst (use-case 2)
  – Visualization of the ‘big picture’ of the cyber-threats landscape

Challenges

- Identified Problems when Stix mapped or used for a particular network.
- Thousands of threats shared every day using Stix.
  - How to identify threats relevant to organization infrastructure ?
  - Which one is important ?
  - Which has higher impact or critical ?
  - What is the likelihood of particular exploit ?
  - What could be the damage in terms of privacy , integrity, availability.
  - How much Cost will be affected ?
  - What nodes will be affected , if particular threat occurs.

STIXCHECKER – FROM STIX TO ACTION

STIXChecker Objectives

- Extend and develop **ontologies as a working model** for STIX, network and vulnerabilities
- Identify **relevance** of prevalent STIX threats according to network architecture .
- Quantitative estimation of the **impact** induced by STIX threats to the enterprise mission, assets and security requirements.
- Automatic transition from CTI to mitigation actions.
- C**ost-benefit mitigation analysis** to achieve an optimal level of security when provided with a limited budget.

STIXAnalytics Objectives

- Risk Analytics:

What is the impact of STIX-threats on the enterprise policy based on its network configuration and vulnerability scanning reports?

- Intelligence-Driven Proactive Cyber Defense:

What are the configuration changes and vulnerability fixes that will reduce the risk to an acceptable level without affecting the mission of the system?

- Visual Analytics

What are the most prevalent threats? How are they related? Which ones are instances of the same attack? Which bots co-host multiple malwares?

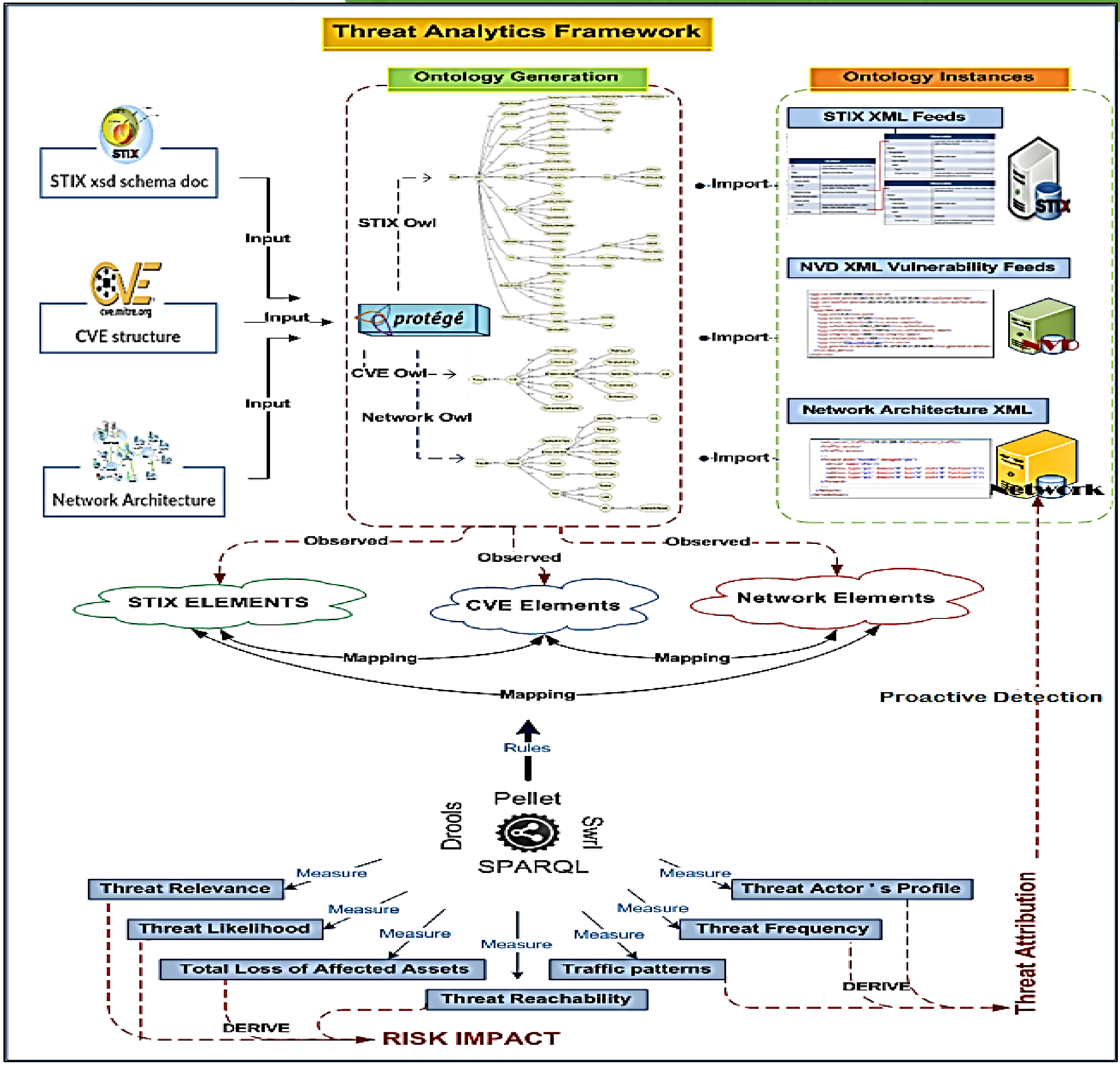
Industrial and Business Relevance

– **Proactive security**: Ensuring that the current implementation of the network reserves its mission in the face of cyber threats.

– **Automatic:** Automatic transition from “threat intelligence” to “mitigation actions”.

– **Cost-effective**: fixes of critical vulnerabilities and risky configurations are based on cost, usability and security requirements.

STIXChecker Process Flow



LOGICAL FORMALIZATION OF STIX (OWL/SMT)

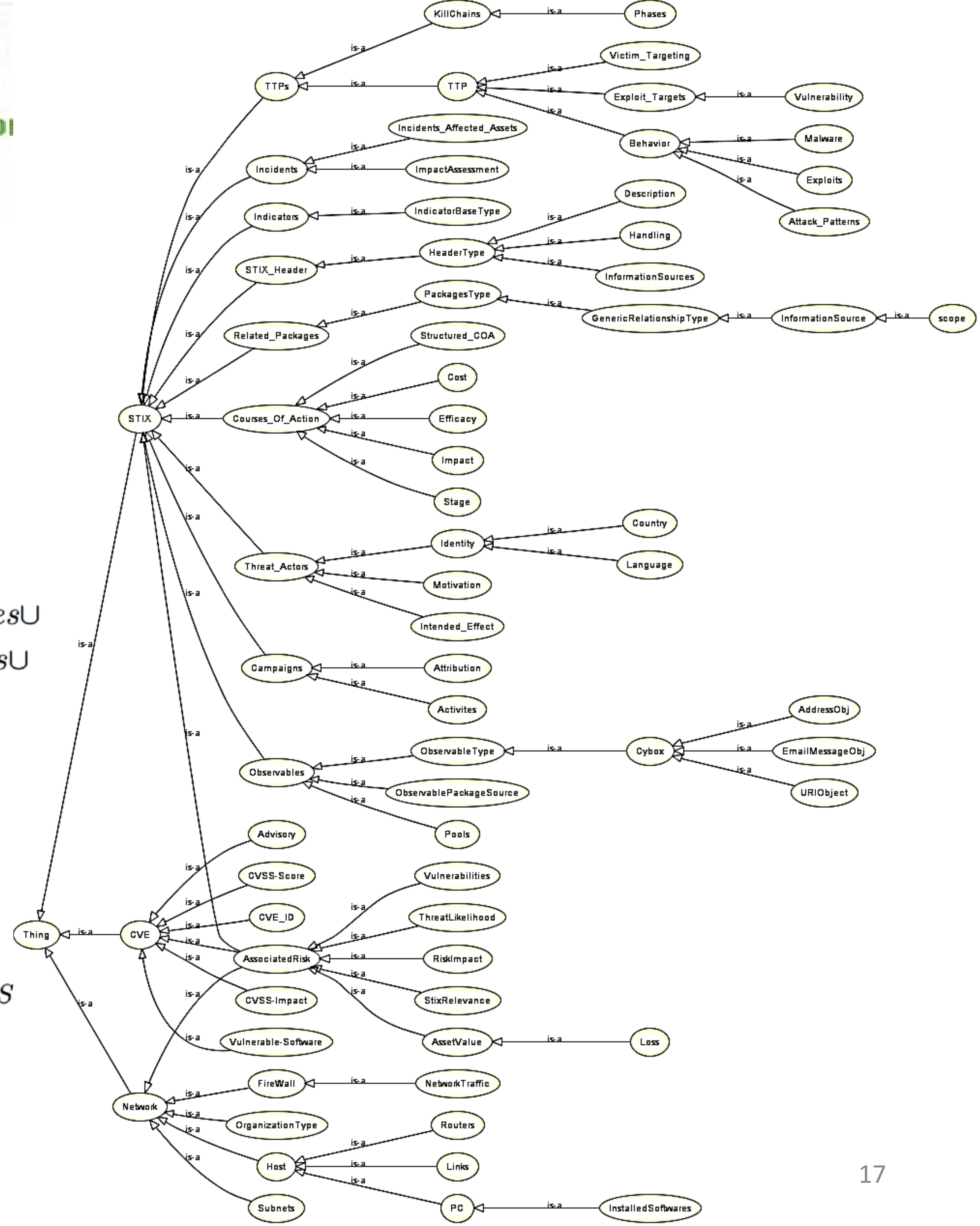
Ontology

- Leveraged
  - Existing work from Vistology
  - NVD Database
- Domains and Restrictions

$$
\begin{aligned}
Indicator \equiv STIX \cup\; & \\
& \in hasObservable\ has\ Observables \cup \\
& \exists\ hasIndicators\ some\ Indicators \cup \\
& \exists\ hasRelatedTTP\ some\ TTP
\end{aligned}
$$

$$
\begin{aligned}
Host \equiv Network \cup\; & \\
& \exists\ hasFirewall\ some\ Firewall \cup \\
& \geq\ hasConnectedHost\ some\ Host \cup \\
& \geq\ hasRouter\ min\ 1 \cup \\
& \geq\ hasHostName\ min\ 1 \cup \\
& \geq\ hasHostIP\ min\ 1 \cup \\
& \exists\ hasVulnerableSoftware\ some\ InstalledS
\end{aligned}
$$

$$
\begin{aligned}
CVE \equiv owl:Thing \cup\; & \\
& \leq\ hasCVE\_ID\ max\ 1 \cup \\
& \leq\ hasCVSS\_baseScore\ max\ 1 \\
& \geq\ hasVulnerableSoftware\ min\ 1 \cup \\
& \geq\ hasAdvisory\ min\ 1
\end{aligned}
$$



DETERMINING NETWORK RELEVANCE

Relevance Factors and Associated Weights

- Relevance Scoring

$$S_i = \frac{E_s \cap E_n}{\bar{E}}$$

where :

E_s is set of relevance elements found in STIX

E_n is set of relevance elements received in network

$\bar{E}$ is set of all available relevance elements

Threat Likelihood

$$L = \max_{0 \le S_i \le 1} \frac{\sum_{i=0}^{N} S_i \times W_i}{\sum_{i=0}^{N} \bar{S}_i \times W_i}$$

where :

N is number of relevance factors F

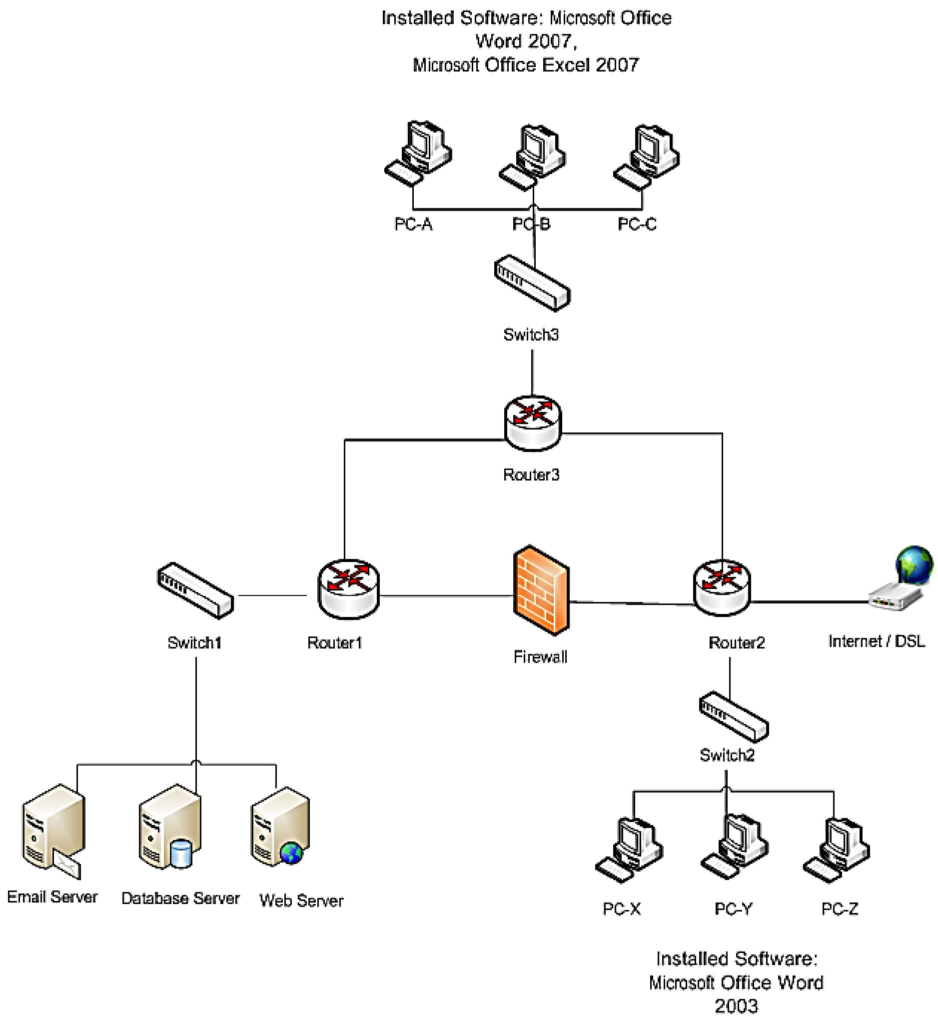
S_i is recieved relevance score for F_i

$\bar{S}_i$ is maximum relevance score for F_i

W_i is assigned weight for F_i

F	W_i
hasCVE_Relevance	5
hasAssetsRelevance	4
hasCIA_Relevance	4
hasTargetedLocationRelevance	3
hasMotivationRelevance	2
hasOrganizationRelevance	1
hasImpactRelevance	1
hasTargetedLanguageRelevance	1
hasSecurityCompromiseRelevance	1

Example Case study– Red October APT & Ashley Madison



	Red October STIX Elemenets (E_s)	Network Elements (E_n)	Relevance Score(S_i)
hasCVE_Relevance	CVE-2012-0158, CVE-2010-3333, CVE-2009-3129	CVE-2009-3129, CVE-2010-3333, CVE-2012-0158	1
hasAssetsRelevance	Servers, Routers, Switches, Persons, PCs and Mobile Phones	Servers, Routers, Switches and PCs	0.66
hasTargetedLocation Relevance	USA	USA	0.25
hasTargetedLanguage Relevance	English	English	1
hasCIA_Relevance	Confidentiality, Integrity, Availability	Confidentiality, Integrity, Availability	1

Network-STIX Relevance

F	$S_i \times W_i$	$S \times W_i$
hasCVE_Relevance	5	5
hasAssetsRelevance	2.64	4
hasCIA_Relevance	4	4
hasTargetedLocationRelevance	0.75	3
hasMotivationRelevance	0	2
hasOrganizationRelevance	0	1
hasImpactRelevance	0	1
hasTargetedLanguageRelevance	1	1
hasSecurityCompromiseRelevance	0	1
SUM	13.39	22
L		13.39/22 = 0.60

Threat Likelihood

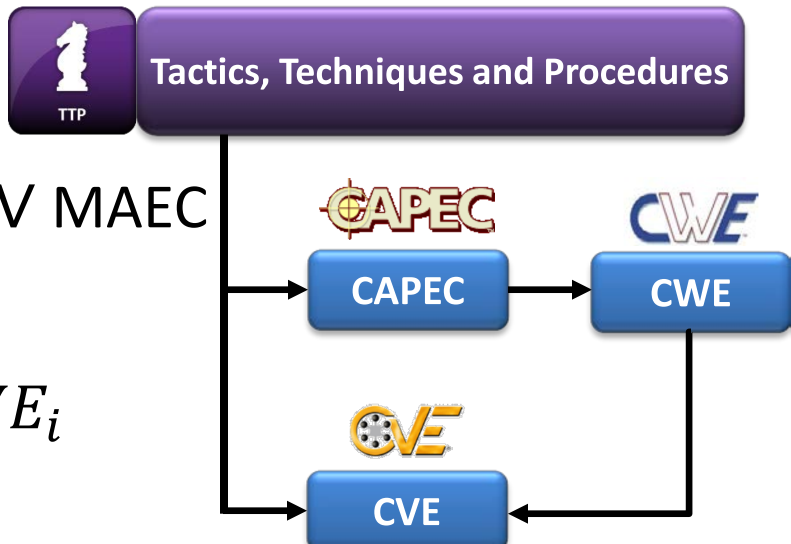
IMPACT ANALYSIS
&
KILLING THE CYBER KILL-CHAIN

STIX threat Modeling

- STIX feeds can be generically formalized as steps within a kill chain.
- A phase can be decomposed into one or multiple TTPs

Kill chain-phase= $\forall_i^{p-1} TTP_i \longrightarrow TTP_{i+1}$

- TTPs can further be broken up into:
  - Attack Patterns given through CAPEC
  - Exploit Targets given through CVE
  - Malwares behavior using MAEC



TTP = CAPEC $\vee$ CVEs $\vee$ MAEC

- CAPEC can be broken into CWEs

CAPEC = $\vee_i^n CWE_i$

Proposed Impact Metric

- Measures the damage inflicted by STIX threats and the contribution of each kill chain phase to the total damage.

$$\text{Impact}(d, \mathcal{S}) = \begin{cases} V_d * A_d * \text{progress} & \mathcal{S} = 0 \text{ or } \mathbb{r} = 0 \\ V_d * A_d * \text{progress} + (w * \sum_{i \epsilon \mathbb{r}} Impact(i, \mathcal{S} - 1)) & \mathcal{S} > 0 \text{ or } \mathbb{r} > 0 \end{cases}$$

Where

V_d: *vulnerability score of the host (likelihood * severity).*

A_d : the asset value of the host d.

$\mathbb{r}$: set of the reachable hosts which are vulnerable to the next **kill chain phase.** (i.e., hosts that have vulnerabilities which enable the next phase).

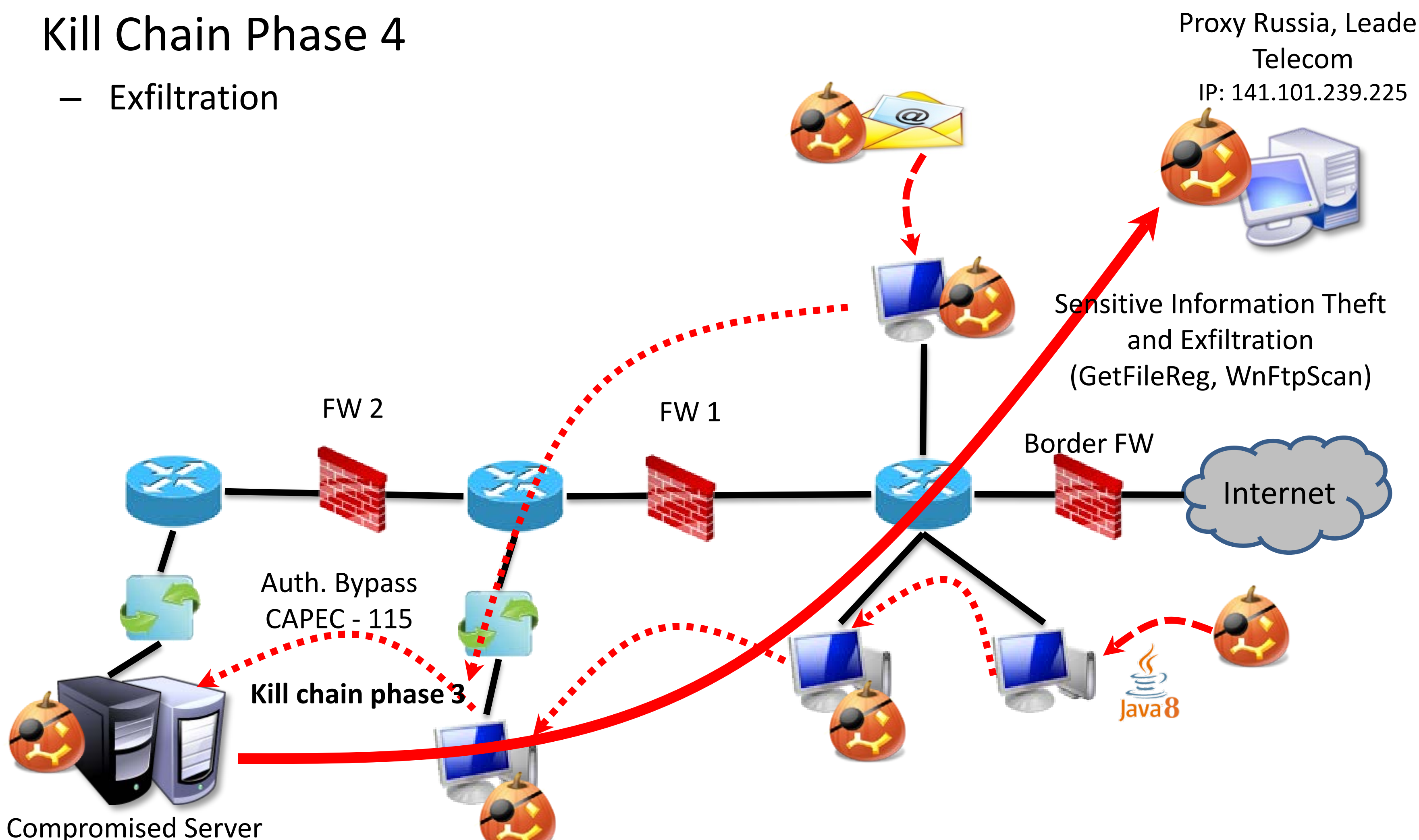
$\mathcal{S}$: the number of maximum **recursion steps** (recursion threshold) S = 0 means a leaf node.

W: a weight variable (keeps getting smaller with every recursive call due to indirect damage)

progress: the percentage of completed attack phases upon successfully compromising the selected host.

Network Impact – Red October Example

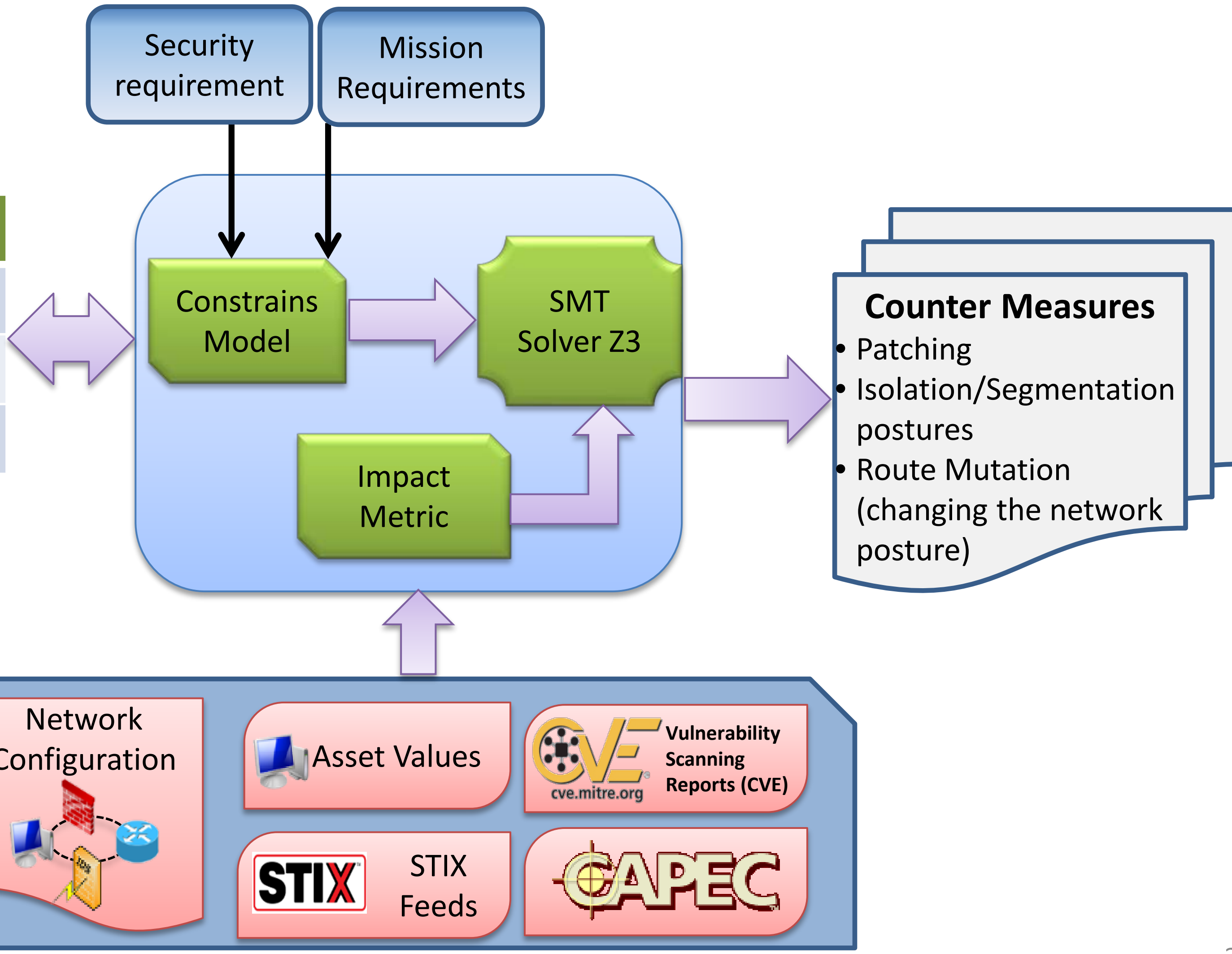
- Kill Chain Phase 4
  - Exfiltration



FROM STIX TO ACTIONS

Inside the Reasoning Engine -- From STIX To Actions

Phase	Potential impact	Prevention cost
1	$X	$A
2	$Y	$B
n	$Z	$C



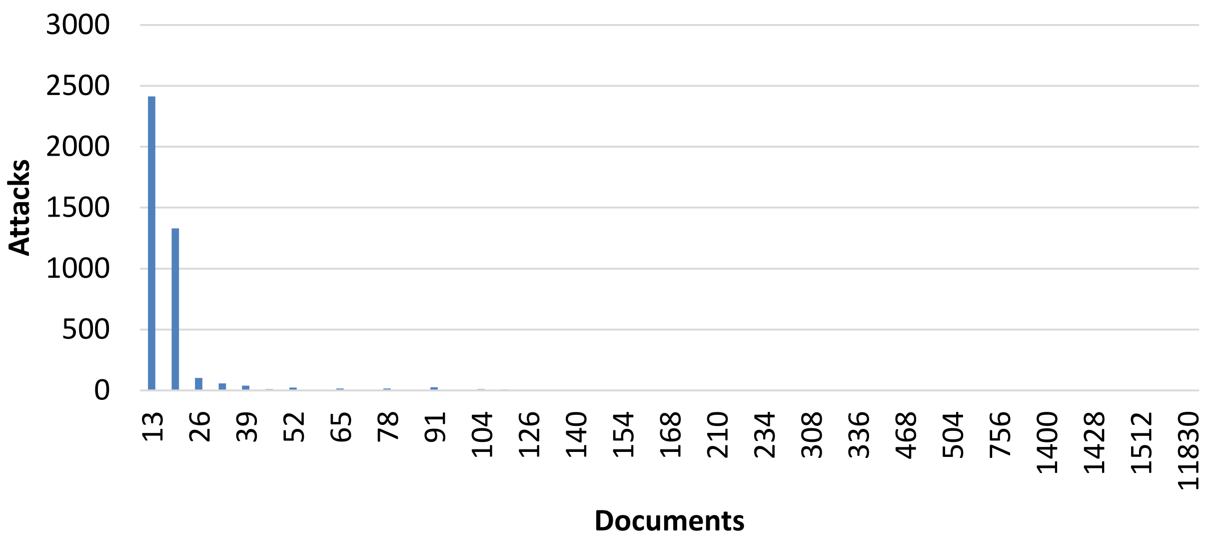
DATA-DRIVEN VISUAL ANALYTICS:
REPUTATION … ETC

Visualization: Initial Exploratory Experiments

- Initial Dataset
  - 2 categories polled from Hailataxii
    - MalwareDomainList
    - CyberCrime_Tracker
  - 10-minute time window
    - 2015-06-25T13:00 - 2015-06-25T13:10
  - 158,510 STIX Documents retrieved
  - ~ 482 MB Size

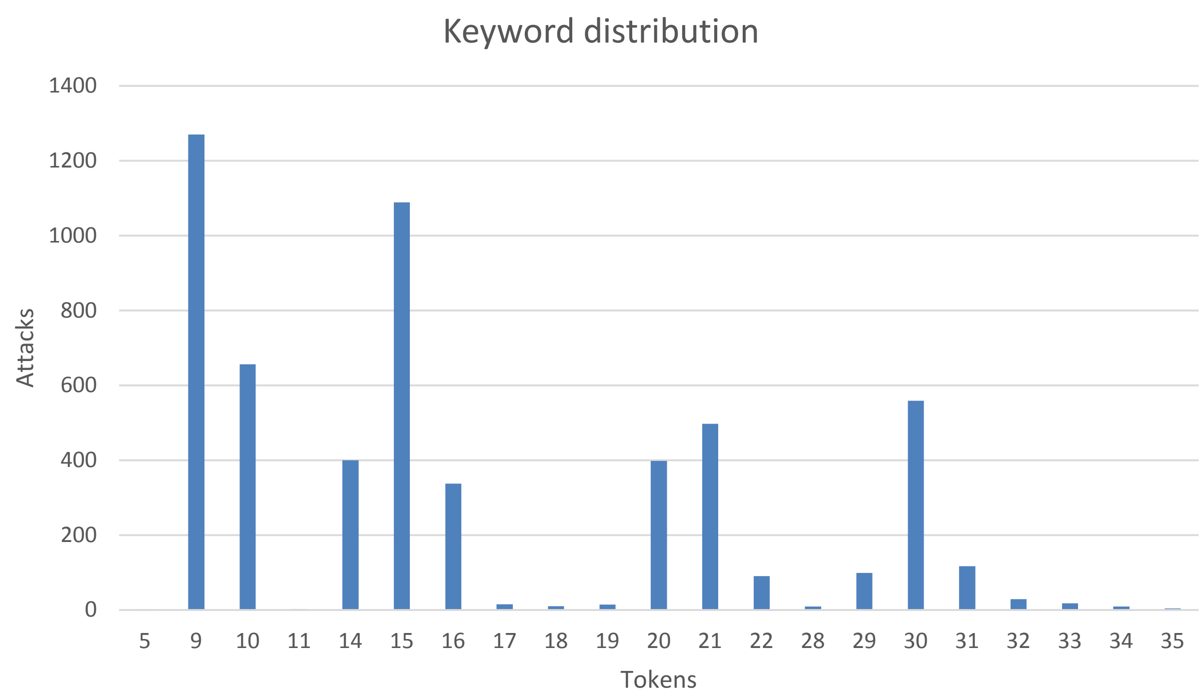
Problem: Dealing with a Huge Dataset

- Experiment 1: Grouping of 158,510 documents
  - Considered XML structures
    - 6 different groups
  - Considered content
    - 5,623 different groups (attacks)
      - 96.5 % reduction
    - Each group represents an attack
      - Avg docs in each group: 28



Problem: Dealing with a Huge Dataset (Continued)

- Experiment 2: Richness of attack information
  - Avg: 16.3 words
  - Excluding: stop words

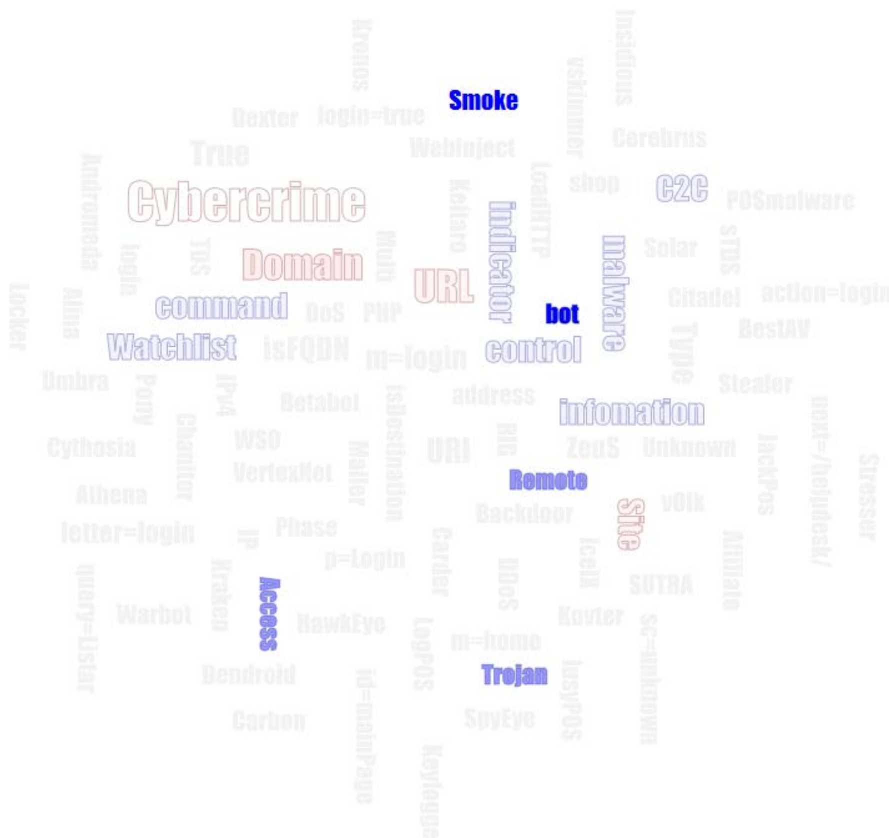


Problem: Visualization of Data (Continued)

- Word Cloud (created from contents)
  - To explore the content of the dataset and to see what is relevant to INW
  - http://cyberdna.uncc.edu/inw/stix/words.php

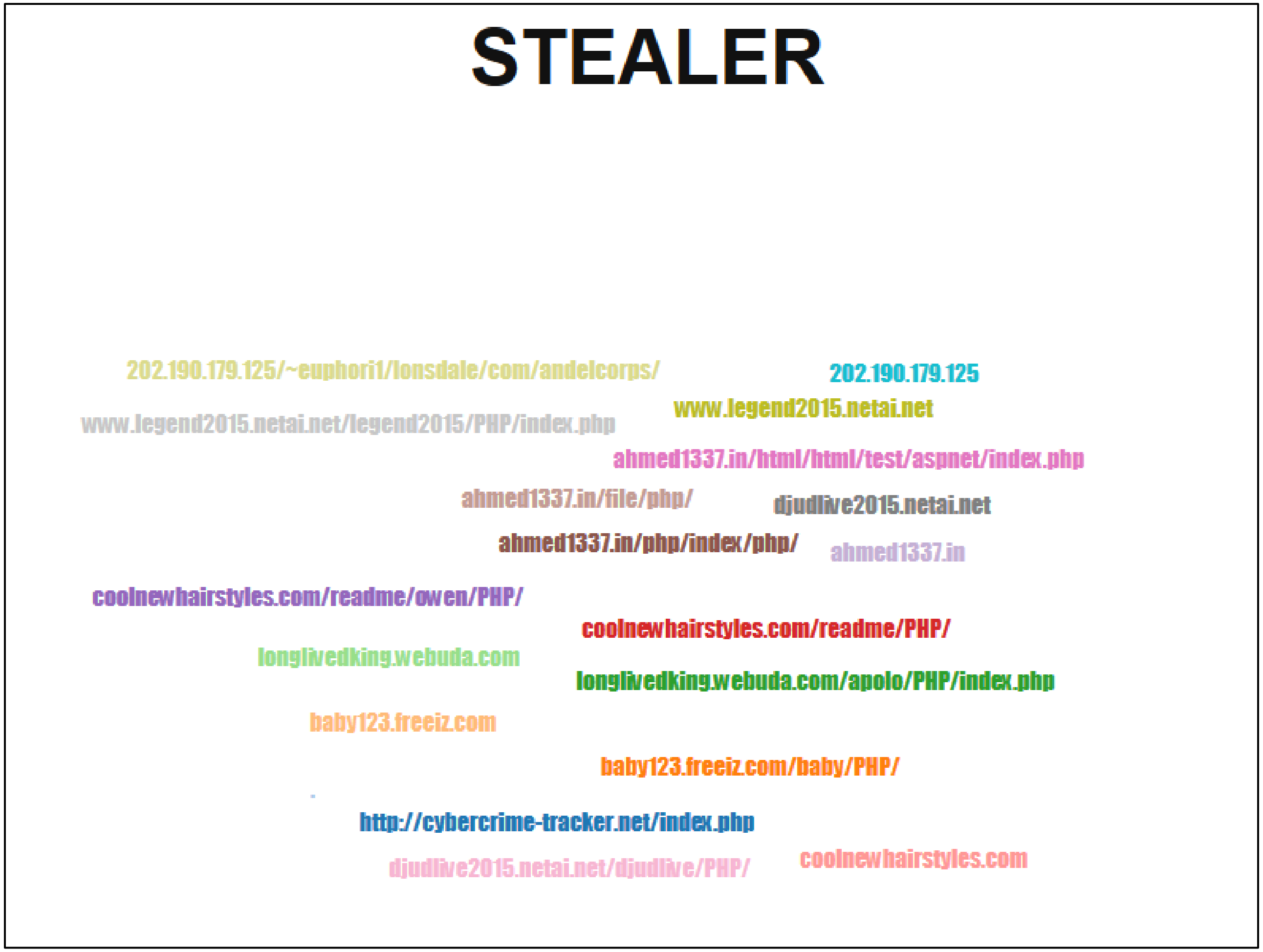
Problem: Visualization of Data (Continued)

- Keyword correlation
  - Intensity of blue => Stronger relationship
  - Gray without border => Not a related word



Problem: Visualization of Data (Continued)

- Urls related to each keyword (co-occurrence)



Problem: Visualization of Data (Continued)

- Related STIX Documents
  - User-friendly sampling of files for each attack (group)
  - Number of STIX documents in each attack
  - Attack Duration

PONY

Attack ID	Sample File	Start Date	End Date	Count
13	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_00_889585_00_00.xml	2015-06-25 17:00:00	2015-06-25 17:09:38	13
14	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_00_916096_00_00.xml	2015-06-25 17:00:00	2015-06-25 17:09:40	494
22	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_00_915034_00_00.xml	2015-06-25 17:00:00	2015-06-25 17:09:38	13
62	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_019254_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
65	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_026258_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
71	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_040738_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
77	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_056161_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
91	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_130903_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
93	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_137958_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
96	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_144995_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
104	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_166655_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
106	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_173758_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
108	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_180305_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:38	13
128	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_306155_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:43	39
144	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_363733_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:43	26
208	guest.CyberCrime_Tracker_STIX111_t2015_06_25T17_00_01_523118_00_00.xml	2015-06-25 17:00:01	2015-06-25 17:09:44	26
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Initial Exploratory Experiments

- Initial Dataset
  - 1 Channels in Hailataxii
    - CyberCrime_Tracker
  - 8-hour time window
    - 2015-06-25T12:00 - 2015-06-25T19:00
  - 3,476,792 STIX Documents
  - About ~10GB

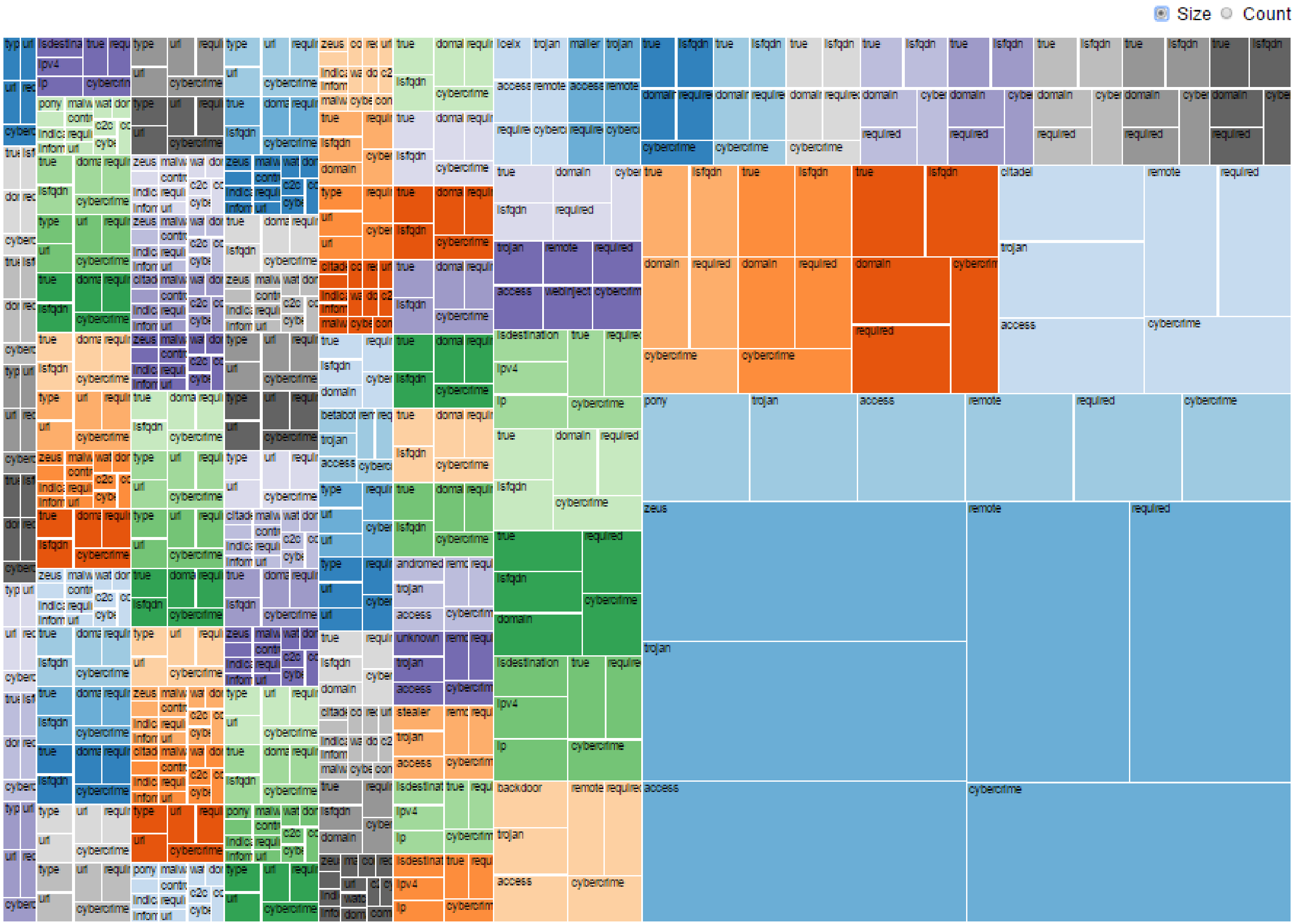
Initial Exploratory Experiments

- Grouping
  - 6 different templates
  - Repeated chunks of text
    - Descriptions
      - [This domain <domain_name> has been identified as a command and control site for <malware_name> malware by cybercrime-tracker.net. For more detailed infomation about this indicator go to [CAUTION!!Read-URL-Before-Click] [http://cybercrime-tracker.net/index.php].</indicator:Description>]
    - Term_of_use
    - Statement

Initial Exploratory Experiments

- Visualization of Related domains
  - https://public.tableau.com/profile/hu4869#!/vizhome/stix/Sheet2

Treemap view Identifies which attacks are Important



'* each color represents an attack

OBJECTIVE REPUTATION OF CYBER THREAT INTELLIGENCE SOURCES–

TOWARD PURIFICATION AND CLASSIFICATION

Motivation and Goals

Time-consuming process

- STIX will include noisy and possible malicious sources
- How do you know which CTI sources to consider:
  - Removing noise: duplication, bogus etc
  - Priority-based classification
- Creating community self-awareness and accountability
- Allow customers to narrow their search and act faster
- Proposed Ranking Service is based on:
  1. **Threat-source profiling** based on time-series and information theatric analysis
  2. **Multi-Source correlation** using clustering and visualization for STIX inter-relationship and source inter-dependency analysis
  3. Sentiment Analysis and **Consumer Reports**
  4. Integrating **Cyber Intelligence** information to enrich the reputation analysis

Key Features in selection of Reputable CTI sources

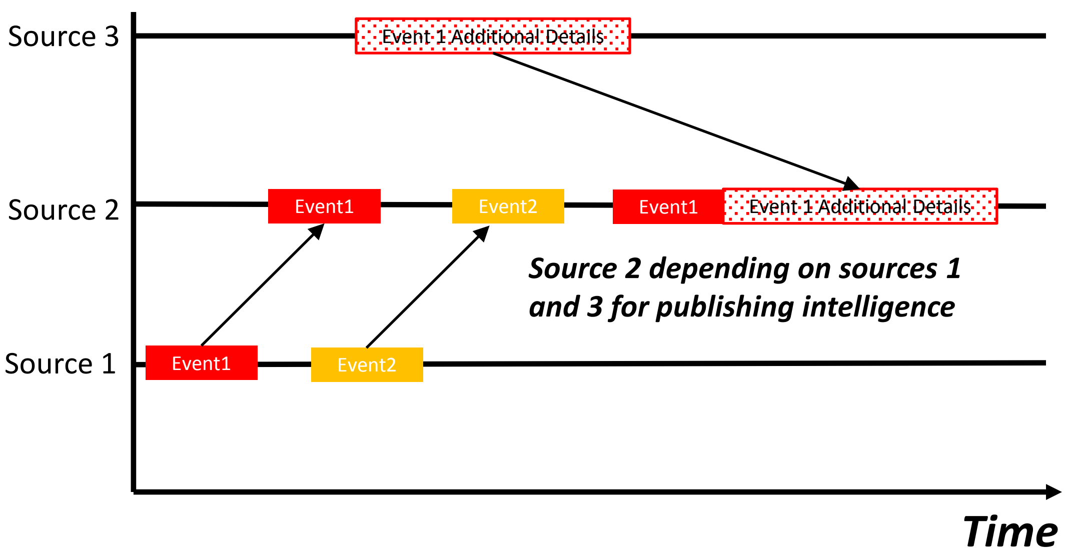
- Number of entries (signal/noise)
- Certainty (blind aggregation, lack of context)
- Type of badness (only certain types e.g. C&C)
- Standards followed (direct input to network FW?)
- Update Frequency (daily, hourly, real-time)
- Varying level of detail
- Frequency of false positives
- Threat Querying by application and features

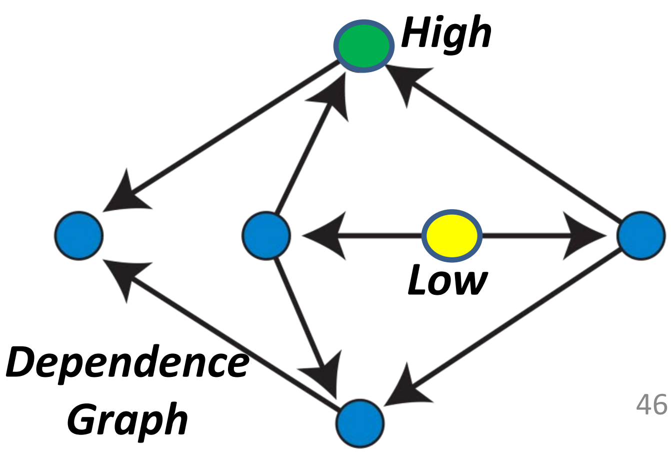
Consumer Reports Best & Worst CTI Sources 2015
Recommended Sources
Exclusive Ratings
Rating of over xx Sources

	Coverage	Dependency	Standards	Trustworthiness
Source 1				
Source 2				

Development of scores/metrics

- ***Detect faster***: How much will feed reduce time to detect?
- ***Detect Better:*** How much will feed enable me to detect what I would otherwise miss?
- ***Dependency Score:*** for decision making about independence of source





Value Proposition

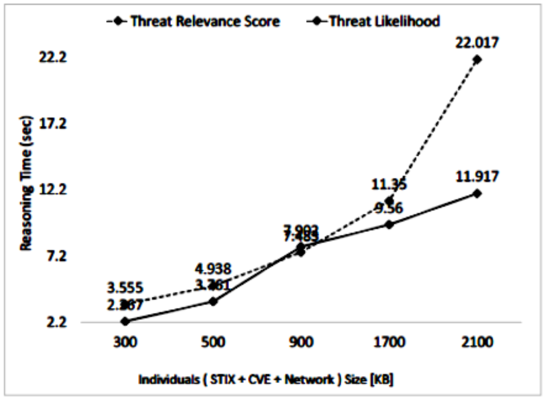
- Ability to rank source reputation and purchase source based on:
  - The specialization of the threat source
  - Quantitative/qualitative scoring
  - Feature wish-list search
  - Partially ordered (ranked) lists
  - Coverage
  - Suggestions based on user requirements
    - Single best source of threat intelligence
    - Customization of services from multiple sources

PRELIMINARY RESULTS

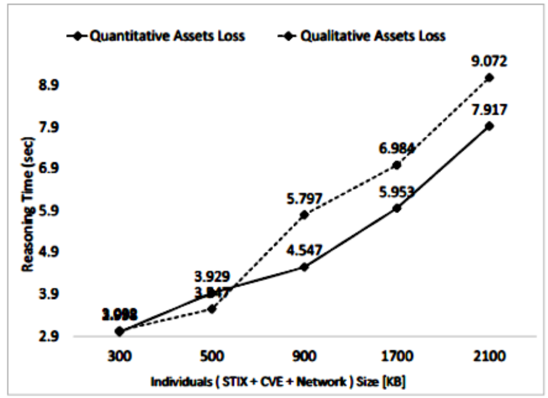
Validation Experiments of our Preliminary Results

- We used 20 different case studies of various attack represented in STIX
  - X are already made ones including Red October APT attack
  - Y are created by our team based on CTI sources such as ThreatConnect including Ashley Madison attack
- Validation Methodology

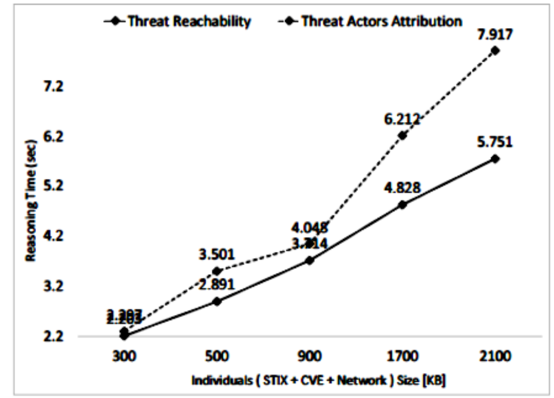
Reasoning Time and CPU Utilization



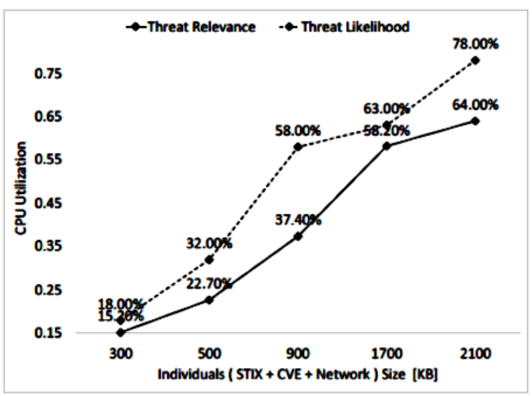
(a) Relevance Score (S_i) & Threat Likelihood (L)



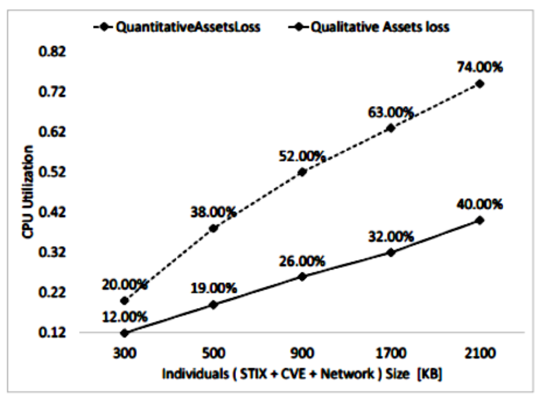
(b) Quantitative Assets Loss (A_n) & Qualitative Assets Loss (A_l)



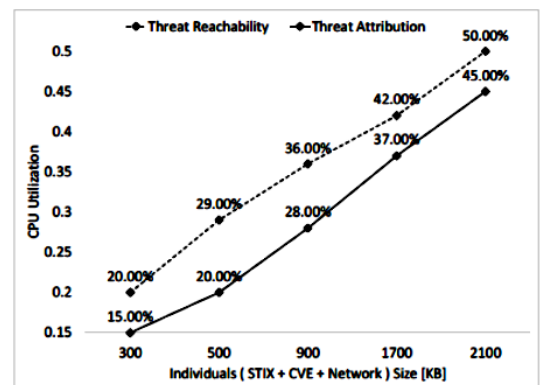
(c) Threat Reachability (R) & Threat Actors Attribution



(a) Relevance Score (S_i) & Threat Likelihood (L)

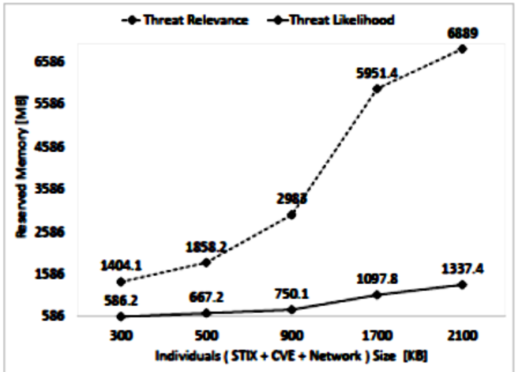


(b) Quantitative Assets Loss (A_n) & Qualitative Assets Loss (A_l)

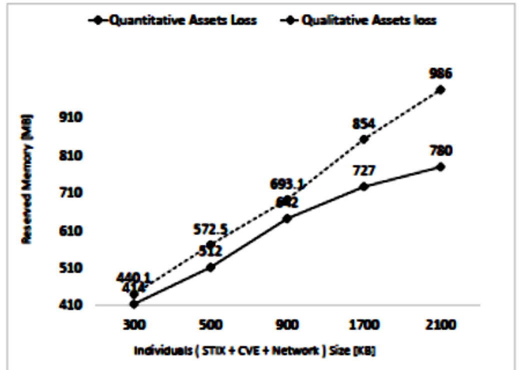


(c) Threat Reachability (R) & Threat Actors Attribution

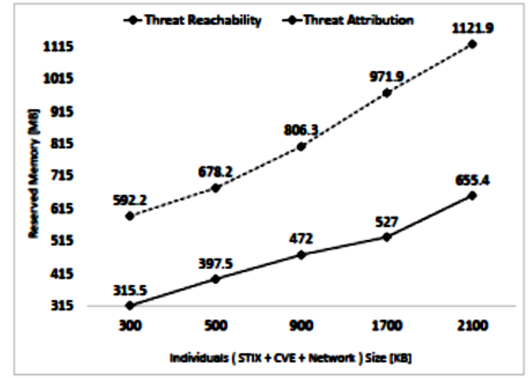
Memory Consumption



(a) Relevance Score (S_i) & Threat Likelihood (L)



(b) Quantitative Assets Loss (A_n) & Qualitative Assets loss (A_l)



(c) Threat Reachability (R) & Threat Actors Attribution

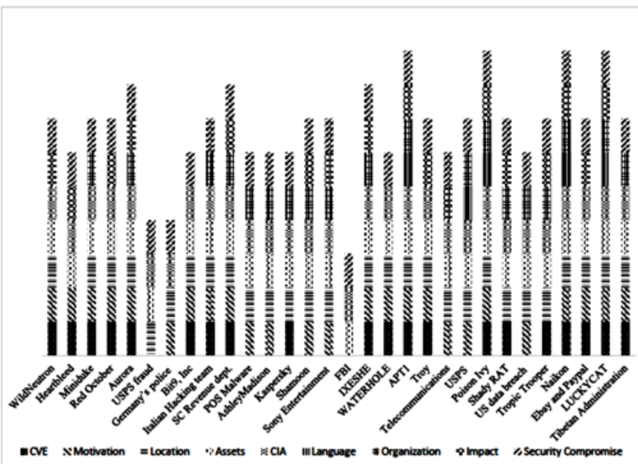


Fig. 10: Relevance Factors (F) found in STIX

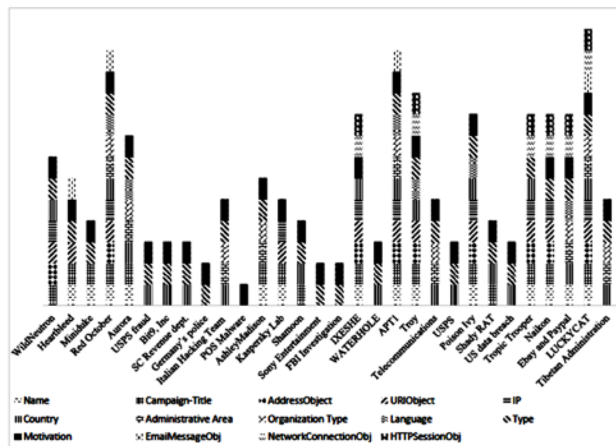


Fig. 11: Threat Actor's Attributes found in STIX

Conclusion

- STIX Threat Information Sharing is the right step in the right direction for cybersecurity automation
- But many others steps have to follow to create incentive (usability and effectiveness) of STIX-based CTI.
- Our experience shows both formal- and data-driven approaches to address critical challenges and bridge this gab between CTI sharing and usability/effectiveness
- This is the tip of the iceberg: More research and development is needed in this direction …
- Relevance: Invitation visit and join the NSF Center on [Security] Configuration Analytics and Automation (www.ccaa-nsf.org);

Questions