



Cyber Risk and Resilience Analytics

Sachin Shetty, Ph.D.

Executive Director, Secure and Intelligent Critical Systems, Virginia Modeling, Analysis and Simulation Center

Professor, Department of Computational Modeling and Simulation Engineering Old Dominion University





About Me

- Executive Director, Center for Secure and Intelligent Critical Systems, Virginia Modeling, Analysis and Simulation Center, Old Dominion University
- Joint Appointment- Professor, Department of Computational Modeling and Simulation and Engineering
- Faculty Appointment, Naval Surface
 Warfare Center, Crane, Indiana (Secret Level Clearance)
- Research supported by AFRL, AFOSR, DHS, DOE, NSF, NEEC, ONR, Sentara, and Boeing

Research Goal

 Modeling and analysis of threats to protect next generation Internet, cloud, mobile systems and networks and critical infrastructures.

Research Interests

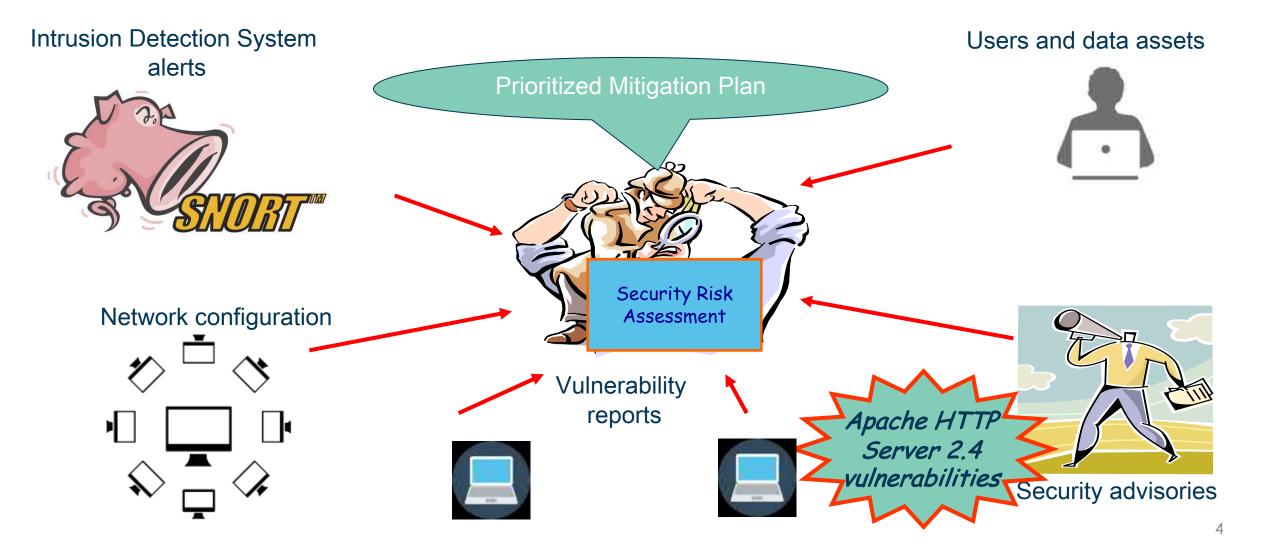
 Cyber security risk and resilience modeling and assessment

Blockchain for distributed systems security
 Machine learning for anomaly detection

Course Overview

- June 17
 - Cyber Risk and Resilience Analytics Overview
 - Modeling Attacker Opportunity
 - Lateral Propagation Analysis
 - Assess Adversarial Effort
 - Infer Adversarial Action and Intent
- June 18
 - Hands on exercise in virtualized environment
 - Learn to generate and analyze attack graphs
 - Computer cyber risk and resilience metrics

Life in the Security Operation Center



<u>()</u>)

TRENDS IN CYBER THREATS

EVOLVING TARGETS

- Data and Knowledge
- Critical Infrastructure
 systems, such as ICS
- Information theft
- Hacking for disruption



- Using human layer (weakest link)
 - Phishing, malicious insiders
- Complex multi-stage (low and slow APTs)

EVOLVING
IMPACTS



- Data still a target
- Theft is not always the outcome
- Data being destroyed
 or changed which generates mistrust

Mining Industry Cyber Threat landscape, IDEFENSE, Accenture Security

Motivation

- The cyber integration with critical infrastructure (CI)
 - Enables high reliability and fast operability
 - Impose risk of disruption of safe and secure operation
- Resilient cyber infrastructure- Ability to anticipate, withstand from deliberate attacks, threats or incidents
- Critical targets often segregated and often deployed away from the perimeter, hard to get into with direct access
- Defense-in-depth architecture forces attackers to conduct lateral propagation.

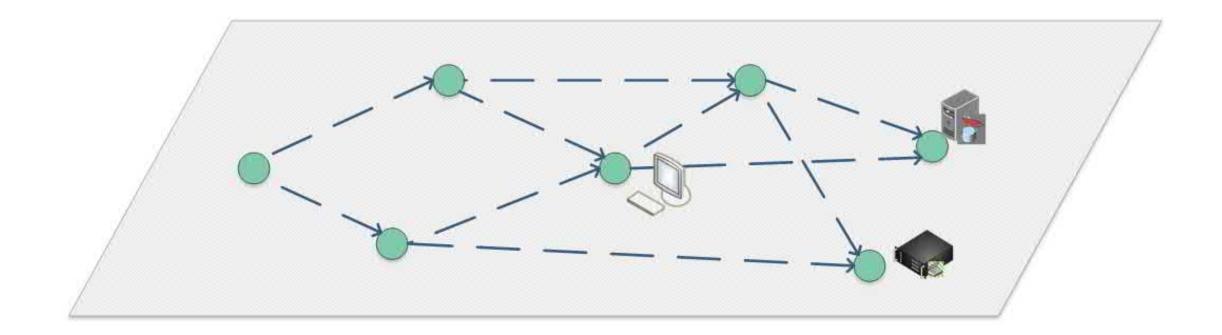
Motivation Cont...

- Adversaries need to propagate a long span of attack surface to reach their goal
- Recently executed more attacks took advantage of this architecture
- Resilience analysis is quite challenging due to—
 - Large scale
 - Heterogeneous network
 - Interdependency
- Adversarial opportunity and behavior is critical to understand the threat cycle
- Detection, prioritization and mitigation is a complex problem in CIs security

Analytics Playground- Attack Surface

- Subset of system resources that can be potentially used by an attacker to launch an attack.
- A larger attack surface-
 - Higher threat landscape
 - Increased risk of compromise
- Different attempts has been taken to dynamically modulate attack surface.
- Some possible challenges:
 - Costly overhead to the legitimate users
 - Maintenance complexity
 - Disruption of service
- Existing work doesn't consider attack life cycle.

Analytics Playground- Attack Surface



Our attack surface analysis aim to protect critical functionality of the systems amid adversarial events

Cyber Resilience Assessment in Critical Infrastructure (CI)

Attack Progression **Opportunity**: Identify stepping stones and lateral propagation Capability:

Adversary skill set to conduct successful exploitation Intent: Attack plans, attacks goals

Defender

Holistic View of the CI. Potential exploitation and impact Technical exposure through attack paths

> Intrinsic Knowledge. Evolving skill

Identify accessible critical assets

Strategic plan driven by motivation and constraint

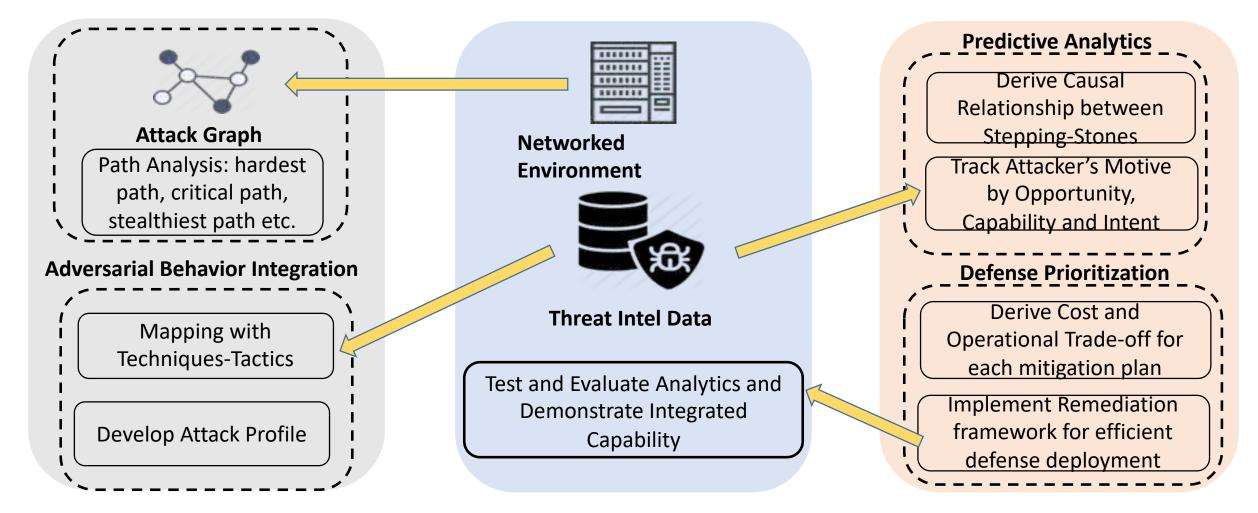


Limited View of the Cl. Increased insight with successful exploitation

Cyber Resilience Analytics



accenturesecurity



Goal 1- Problem Statement

- Existing work model lateral propagation-
 - Graph spectral matrices
 - Bypassing contextual analysis.
- Host in attack surface facilitates attack progression within this.
- Need to conduct susceptibility analysis of hosts along the path to the target
- Heterogeneous network architecture and expands diverse opportunity

How to model attacker's opportunity within each host in the attack surface and how evolves through attack progression

Goal 2- Problem Statement

- Opportunity gives different options to propagate
- Sophisticated attacker's progression dictates by their motivation
- Static cyber defense doesn't provide effective intrusion response
- Existing work doesn't consider different aspects of attack strategy
- Need an efficient situational awareness to restrict persistent threats from deeper penetration

Given attacker's opportunity space, how to model an intrusion response system by considering diverse strategy an attacker could employ in lateral propagation

Goal 3- Problem Statement

- Attacks are diverse in terms of techniques, progression and impacts
- Attacks often follows-
 - Sequence of steps towards target
 - Sequence of actions in each step
- Existing analysis only consider topological connection between stepping stones
- Assuming pre-defined skill set ignoring attack complexity in the propagation path.
- Opportunity gives attackers potential expanse
- Additional Inspection needed to reveal hidden insight of attack step

How to assess attacker's evolving effort by characterizing potential adversarial behavior in the attack surface

Goal 4- Problem Statement

- Followed sequence depends on preference
- Understanding attackers' motive is critical to track the behavior
- Conventional IoC doesn't provide adequate defense against malicious campaign
- Meaningful campaign information could shrink huge attack surface
- Extract attack pattern not IoC

How to infer attackers prioritized action and the respective motive in the attack surface

Modeling Attacker's Opportunity

- Sharif Ullah, Sachin Shetty, Amin Hassanzadeh, "Towards Modeling Attacker's Opportunity for Improving Cyber Resilience in Energy Delivery Systems", Resilience Week, Denver August 2018.
- Ullah, Ullah, Sharif, Sachin Shetty, Amin Hassanzadeh, Anup Nayak and Kamrul Hasan, "On the Effectiveness of Intrusion Response Systems against Persistent Threats." In 2020 International Conference on Computing, Networking and Communications (ICNC), IEEE, 2020.

Overview of the approach

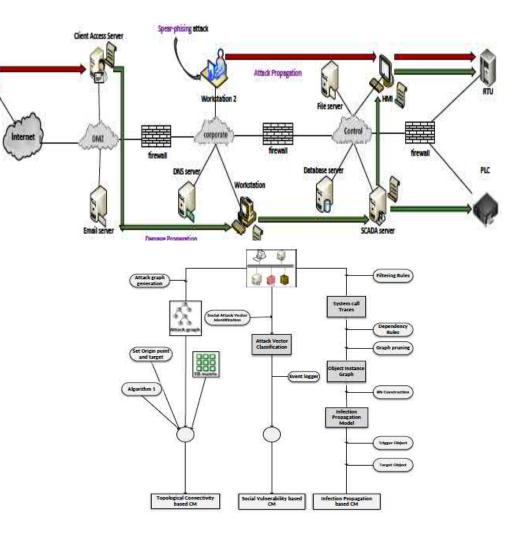
• Goal-Modeling attacker's opportunity for lateral propagation

Opportunity can be categorized as :

- Use case 1 Attack propagation
- Use case 2- Attack origin
- Use case 3- Damage propagation

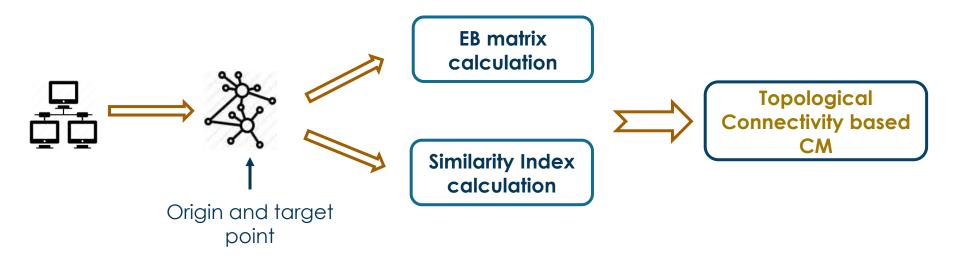
Three criticality metric:

- Topological Connectivity based Criticality Metric (TCCM)
- Social Vulnerability based Criticality Metric (SVCM)
- Infection Propagation base Criticality Metric (IPCS)



TCCM and SVCM

- TCCM captures Attacker's propagation itself by means of connectivity and other resources.
- Along with Global info. the contextual info incorporated to identify correlated risk



- SVCM captures Opportunity to attacker's prior to penetration.
- Assign score to hosts based on the susceptibility of social engineering attack.

TCCM- Parameters

- Model the opportunity each exploitable host provides.
- Attack graph is generated
 - Network connectivity map
 - Attacker's privilege and security condition of host
- Attack path is characterized-
 - Global info- degree of exploitability
 - Contextual Info-
 - Vulnerable Service (VS)
 - Operating System (OS)
 - Isolation Pattern (IP)

Along with global info. the contextual info is incorporated to identify the correlated risk

TCCM- Parameters

- *Similarity Index*: relative abundance of difference instances of contextual parameter in an attack path
- Similarity index of parameter z in attack path p_y is given as

$$S_{index(p_y,z)} = w_z \times r_{p_y,z}$$

Where Effective richness of parameter z

$$\gamma_{p_y,z} = \prod_{j=1}^q \frac{m_j}{|H_y|}$$

~

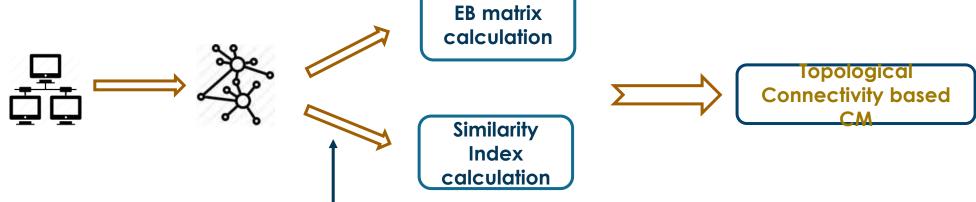
- Each path has q types of instances of a parameter
- m_j : number of instances of type j
- $|\dot{H_y}|$: Total number of host/instances in whole attack path
- w_z : weight factor

TCCM- Parameters

• *Effort – betweenness matrix (EB)*: describe the cost of each host through a path.

• Element of EB: $e_{p_yn}^t = \frac{\text{candidate host cost } (c_{p_y}^t n)}{\text{path cost } (C_{p_y}^t)}$

• TCCM-
$$\mathcal{C}_{top}^{n} = \sum_{t \in \mathcal{T}} \left(\sum_{p_{y} \in \mathcal{G}} \frac{1}{\mathcal{E}_{p_{y}n}^{t}} \times [e_{p_{y}n}]_{t} \times \prod_{j} \frac{1}{1 - S_{index(p_{y},j)}} \right) \times \mathcal{D}^{t}$$



Origin and target point

SVCM- Parameters

- Attack path analysis is not sufficient to capture the opportunity of insider attack.
- This metric assign score to hosts based on the susceptibility of social engineering attack.
- Classify each attack vector (AV) in terms of stages an attack spans.
- Major stages is divided by multiple substages, marked by classification parameter.
- Each sub-stage classification parameters is mutually exclusive.

Table I: Classification parameters for AVs

Orchestration	Target chosen	Explicit target (l_1) Promiscuous target (l_2)
	Method of Distribution	Local (l_3) Remote (l_4)
	Mode of Automation	Manual (l_5) Automatic (l_6)
Exploitation	Deception vector	Cosmetic (l_7) Behaviour (l_8) Hybrid (l_9)
Execution	Attack Persistence	One-off (l_{10}) Continual (l_{11})
	Execution step	Single step (l_{12}) Multi-step (l_{13})

SVCM- Metric

- Importance of classification parameter vary from network to network.
- Define score on each attack vector **social vulnerability score** (SVS)
- Weight Z_k for classification parameter k based on defenders policy and strategy.
- Compute SVS for attack vector *i* given the set of classification parameter L_i

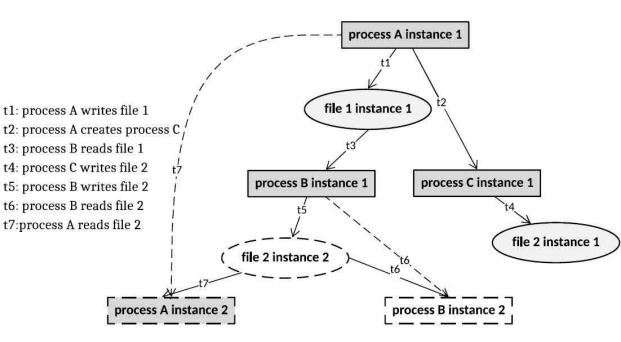
$$SVS_i = \frac{\sum\limits_k (Z_k \times l_k)}{|L_i|}$$

Our second criticality is derived by summing *SVS* and frequency of interaction *f* over all AVs for host *n* and multiply with network diversity

$$\mathcal{C}_{sv}^n = \left(\sum_{i=1}^m SVS_i \times f_i\right) d_{sv}$$

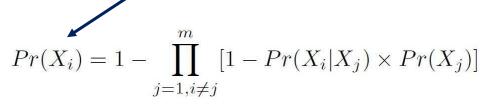
IPCM-Infection Propagation

- Interaction between system objects could be new opportunity for the attacker
- We classify three types of objects : process, files and sockets for our analysis
- Capture the dependency between objects by **control** and **information flow** between them
- Thus intrusion among any object could initiate infection propagation
- We model this propagation by **object instance graph**
 - Host object instance graph $(HOIG) \equiv (V, E, O_v, d_E)$
 - \blacktriangleright $O_{v}: V \rightarrow \sum_{T}$ vertex to syscall trace
 - \blacktriangleright **E**: set of edges \equiv functional dependency
 - V: set of vertices ≡ obejct instances
 - $> d_E : E \rightarrow dep_e \text{ edge with specific source} \\ \rightarrow sink dependency$



IPCM- BN formulation & Metric

- Infection propagation can be classified:
 - Intra-object infection propagation
 - Inter-object infection propagation
- Bayesian Network (BN) is effective tool to incorporate intrusion evidence in order to characterize infection propagation
- Probability of infection can be calculated with the CPT and the given equation

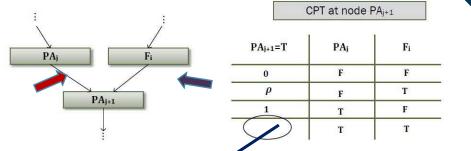


SVCM (maximum damage):

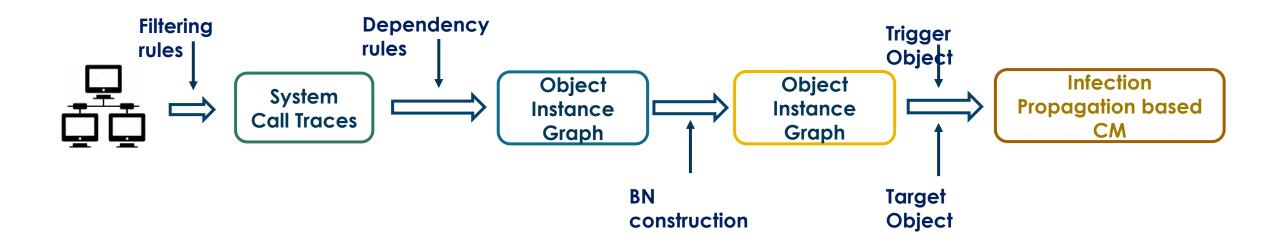
Set of objects related to process control and operation $\mathcal{C}_{inf(max)}^{n}(\eta) = \{Pr[o_{1}, o_{2}, \dots, o_{i} | e_{1}, e_{2}, \dots, e_{j}]\} \prod_{o_{c} \in \{O_{c}\}} Impact on substation operation$ SVCM (minimum damage):

$$\mathcal{C}_{inf(min)}^{n}(\eta) = \{1 - Pr[o_{1}^{c}, o_{2}^{c}, \dots, o_{i}^{c} | e_{1}, e_{2}, \dots, e_{j}]\} \sum_{o_{c} \in \{O_{c}\}} I_{o_{c}}$$

<u>(()</u>)



IPCM- System Model



Ŵ

Simulation Results: TCCM

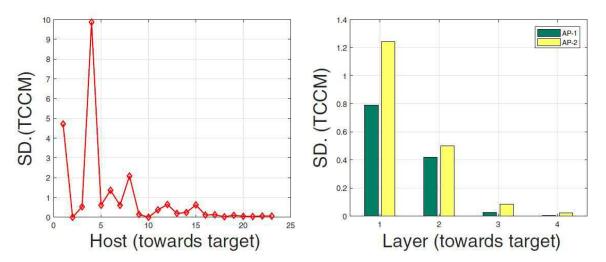


Fig. 2: Standard deviation of TCCM due to multiple initial attack points, for each host (left) & for hosts within layer (right)

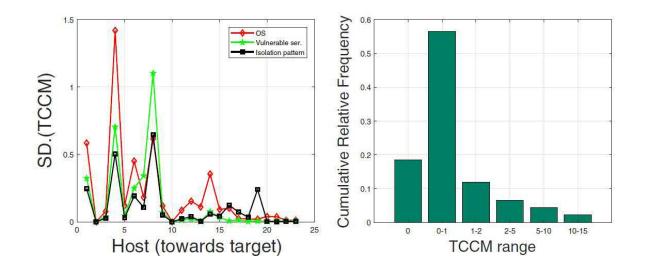
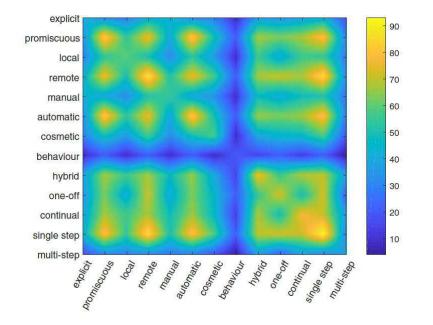


Fig. 3: Evaluation result for SD of TCCM due to OS, VS and IP (left), risk of the network for particular attack points

Simulation Results: SVCM and IPCM





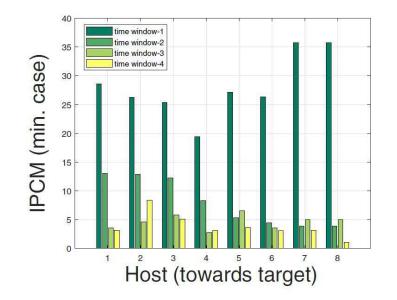


Fig. 4: Network diversity in terms of social attack vectors

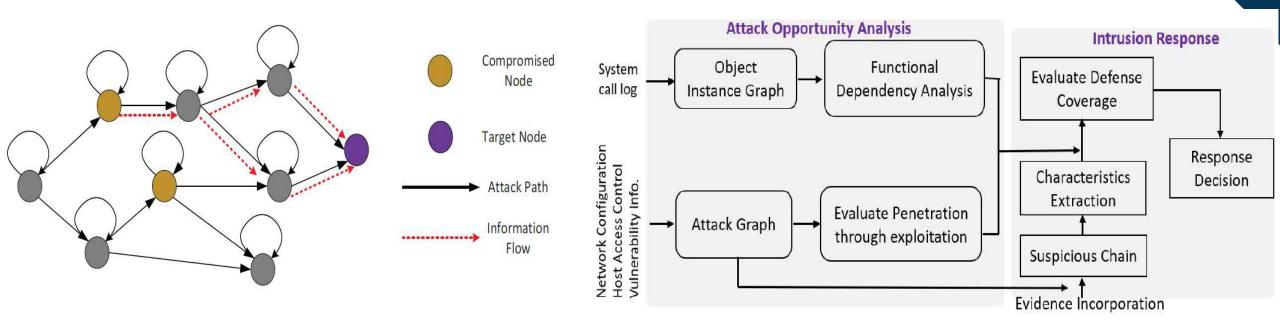
Fig. 5: Infection Propagation based CM for hosts

<u>(Ú)</u>

- Capture and analyze dynamic behavior of attacker within attack path
- Model attacker's opportunity corresponding to each strategy
- Extract post compromise sequence characteristics to inflict next adversarial options.
- An online dynamic intrusion response system
- Deter attack occurrence by increasing the cost and uncertainty in attack planning and execution

• Ullah, Ullah, Sharif, Sachin Shetty, Amin Hassanzadeh, Anup Nayak and Kamrul Hasan, "On the Effectiveness of Intrusion Response Systems against

System Model



- Investigate different options of potential opportunity of adversary.
- Opportunity attack surface explore—
 - Attack graph
 - Object instance graph
- Intrusion response phase examine suspicious chain and enforce appropriate decision

Progression through Exploitation

- Model the opportunity each exploitable host provides
- For node j in path p_k the cost value is calculated as
- Known vulnerability exploitation by CVSS exploitability score
- The cost of target host *t* also determined in a similar way—
- The exploitation score

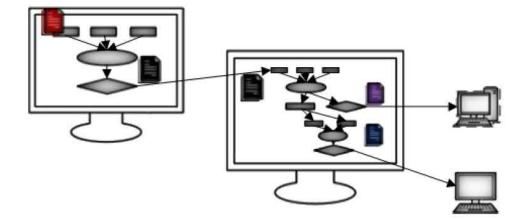
$$ES_j = \mathcal{D}^t \sum_{t \in \mathcal{T}} \sum_{p_k \in P} \left(\frac{1}{\xi_{p_k j}^t} \times \frac{c_{p_k j}^t}{C_{p_k}^t} \times \prod_x \frac{1}{1 - S_{index(p_k, x)}}\right)$$

$$C_{p_k}^t = c_{p_k t}^t = \prod_i^t \frac{y_v}{10}$$

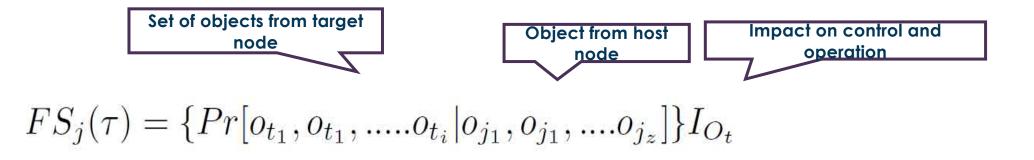
$$c_{p_kj}^t = \prod_i^j \frac{y_v}{10}$$

Functional dependency Estimation

- Functional dependency between hosts could be a stealthy malicious link.
- Capture the dependency between objects by information flow between then and model it through infection propagation.



Functional dependency score:



Intrusion Response Module

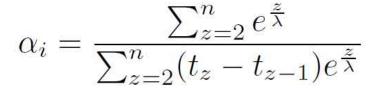
- The IRS module has three stages:
 - Uncertainty of security state
 - Uncertainty of attackers' behavior
 - Response decision making process
- SIEM acquire and analyze real-time information
- Response function triggered after identifying a steeping stone
- Suspicious chain comprised with the set of connected compromised nodes from SIEM events.
- Extract apparent capability for parameter x from chain s_i
- Plugin the weight into exploitation score to model future threat propagation modeling--

$$\begin{split} ES_j^{s_i} &= \mathcal{D}^t \sum_{t \in \mathcal{T}} \sum_{p_k \in P} (\frac{1}{\xi_{p_k j}^t} \times \frac{c_{p_k j}^t}{C_{p_k}^t} \times \\ &\prod_x 1 + (\frac{1 - w_x^{s_i}}{1 - S_{index(p_k, x)}}) \end{split}$$

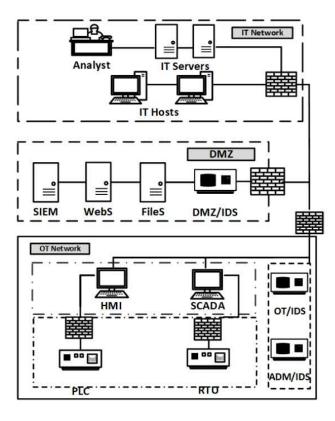
$$w_x^{s_i} = \frac{\sum_{l=1}^{q_x^{s_i}} \left(\frac{1}{\phi_l} \sum_{l} e^{\frac{t_l}{\lambda}}\right)}{\sum_{z=1}^{n} e^{\frac{t_z}{\lambda}}}$$

Intrusion Response Module

- Capture two behavior from suspicious chain:
 - Diverse capability
 - Aggressiveness
- Aggressiveness is tracked by using temporal information associated with each evidence
- How much penetration deviates from most aggressive attacks
- Candidate node has two scores:
 - Exploitation Aggressiveness
 - Functional dependency Stealthy
- Attacker constrained to take one strategy in single time-slot
- Finding optimum node *j* for response--



Implementation and Results



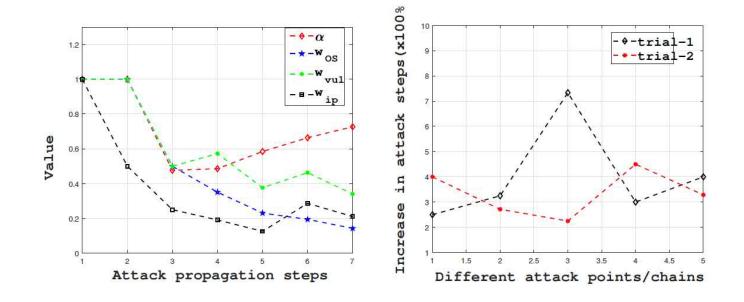


Fig 1: Attacker's evolving characteristics in penetration steps(left) & Performance evaluation of IRS with respect to attack propagation delay(right)

- IEC-62443 architecture including IT, OT and DMZ zones.
- Set PLC and RTU as critical target
- Randomly take initial attackers' position from IT zone

<u>(</u>

Modeling Attacker's Capability (SecureComm'19)

Sharif Ullah, Sachin Shetty, Anup Nayak, Amin Hassanzadeh and Kamrul Hasan "Cyber Threat Analysis based on Characterizing Adversarial Behavior for Energy Deliver System", Securecomm, 2019

Kamrul Hasan, Sachin Shetty, Sharif Ullah, Amin Hassanzadeh, Ethan Hadar, "Towards Optimal Cyber Defense Remediation in Energy Delivery Systems", IEEE Globecom, Hawaii, 2019





"An ounce of prevention is worth a pound of cure." Benjamin Franklin

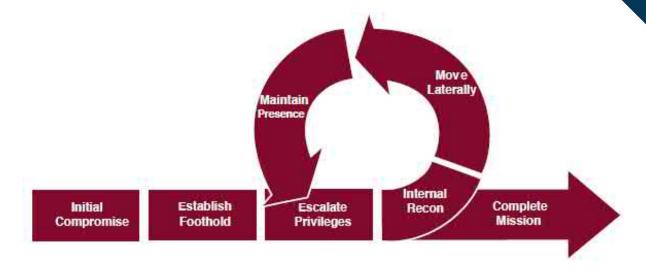
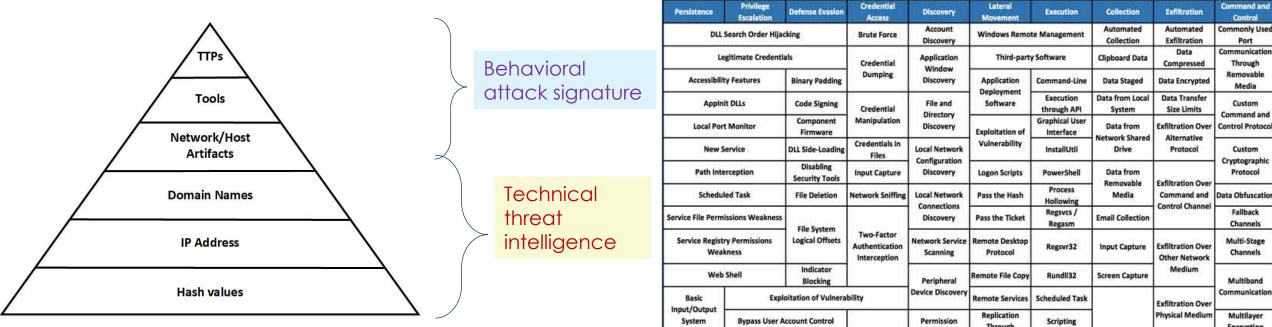


Fig. 1: APT Life-cycle²

- Understanding the motivation and operation of APT actors play a vital role.
- Kill chain provides a reference to understand and map APT actors-
 - Targets
 - Motivations
 - Actions

Pyramid of Plain Model



Pyramid of plain model¹

Tactics: why an adversary performs an action

Techniques: how they take the action

 Described from both the offensive and defensive points of view

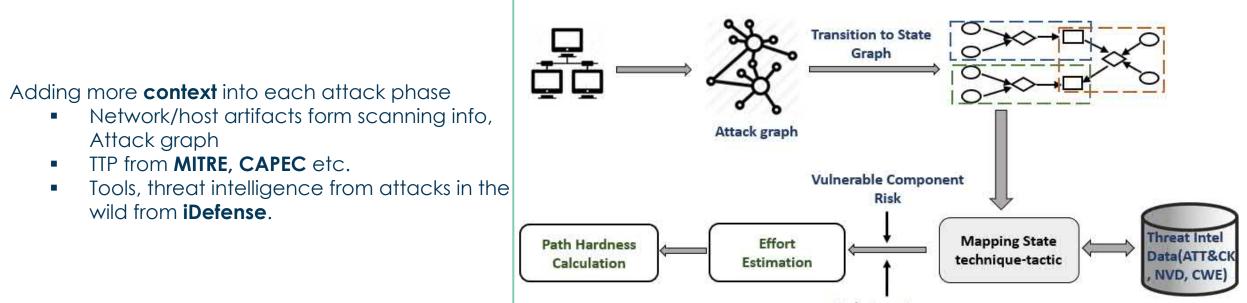
Port Communicatio Through Removable Media Custom Command and ontrol Protoco Custom Cryptographic Protocol Data Obfuscation Fallback Channels Multi-Stage Channels Multiband Communication Multilayer Through Encryption Groups Removable Service Scheduled Bootkit **DLL Injection** Discovery eer Connection Media Execution Transfer Process Indicator Shared Webroot Windows Remote File Copy **Change Defaul** Discovery Removal from Management **File Association Taint Shared** Tools Query Registr nstrumentation Content Standard **Application Lave** Component Remote System indows Admin Protoco Indicator Firmware Discovery Shares Removal on Host Standard Security Hypervisor Cryptographic Software Logon Scripts InstallUtil Discovery Modify Existing Masquerading System Service Information Redundant Discovery Modify Registry Access **NTFS Extended Registry Run** System Attributes Owne

Contro

1. D. Bianco. (2014) The pyramid of plain. [Online]. Available: http://detect-respond.blogspot.com/2013/03/thepyramid-of-pain.html dossier.pdf

 \checkmark

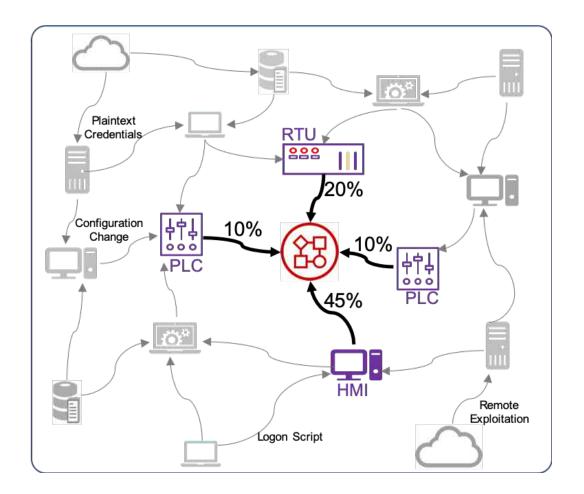




Technique Score

Attack Graph

- 1. Choose business process with the highest monetized risk
- 2. Start with highest value critical asset
- 3. Examine <u>easiest</u> attack paths to the asset
- 4. Fix according to ease of attack and cost of remediation



ATTACK GRAPH ANALYTICS ENGINE

- What are the possible impacts on asset X?
- What are the possible paths to a target?
- What is the most probable attack path from outside the network to asset X?
- Given a path, what are all the configuration issues across it?
- What asset if compromised, provides more lateral movement options (TTPS) for attacker to proceed?
- How to avoid all possible impacts on a given asset?



Copyright © 2019 Accenture. All rights reserved.

Background: Attack Graph

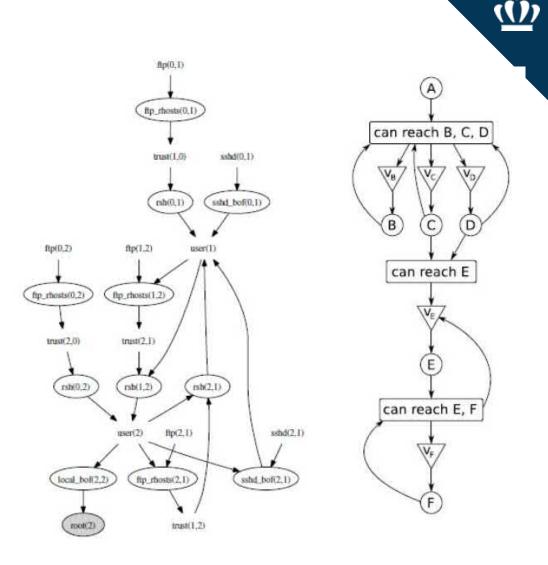
- An attack graph is the mathematical abstraction of details of potential attacks leading to a specific target
- Major two parameters:
 - Node: Probable states
 - Edge: Corresponding changes of states
- Different attack graph model comprised of:
 - Different system parameters
 - Application behaviors
- Multiple use cases in security and risk analysis.

Background: Attack Graph

- Different representation of attack graph is proposed
- State enumeration attack graph:
 - Emerged from model checking technique
 - Node represents entire network state
 - Shows all possible attack paths to particular goal attacker state
 - Suffers from state explosion problem
- Dependency attack graph:
 - First comes with exploit dependency attack graph
 - Node \rightarrow State condition
 - Edge \rightarrow Causal relationship between conditions
 - # of nodes scales linearly

Background: Attack Graph Models

- Topological Vulnerability Analysis(TVA):
 - Model attacker's exploit as transition between security conditions
 - Exploit and security condition nodes
- MulVal reasoning engine:
 - Derivation and fact nodes
 - Directed graph
- NetSPA attack graph:
 - Multi-prerequisite attack graph
 - State node, prerequisite node and vulnerability instance node



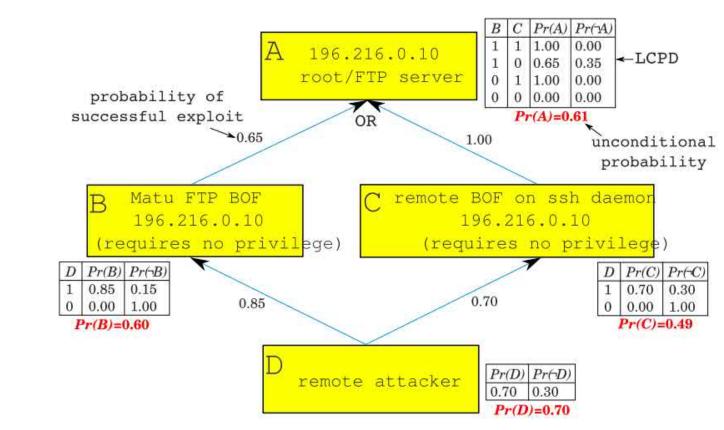
TVA Attack Graph¹

NetSPA Attack Graph²

- 1. Wang, Lingyu, Anoop Singhal, and Sushil Jajodia. "Measuring the overall security of network configurations using attack graphs." In IFIP Annual Conference on Data and Applications Security and Privacy, pp. 98-112. Springer, Berlin, Heidelberg, 2007.
- 2. Ingols, Kyle, Richard Lippmann, and Keith Piwowarski. "Practical attack graph generation for network defense." In 2006 22nd Annual Computer Security Applications Conference (ACSAC'06), pp. 121-130. IEEE, 2006.

Background: Bayesian Network (BN)

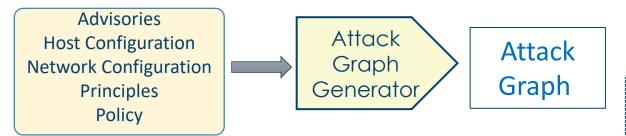
- Probabilistic graphical model representing variables and relationship between them
- Demonstrate causal dependency between exploits
- Quantify the likelihood of attack goals and predict potential attacks.
- Bayesian attack graph is a directed acyclic graph
- A great analyzer for security under uncertainty



Simple Bayesian Attack Graph¹

1. Poolsappasit, Nayot, Rinku Dewri, and Indrajit Ray. "Dynamic security risk management using bayesian attack graphs." IEEE Transactions on Dependable and Secure Computing 9, no. 1 (2011): 61-74.

Attack Graph and Action State



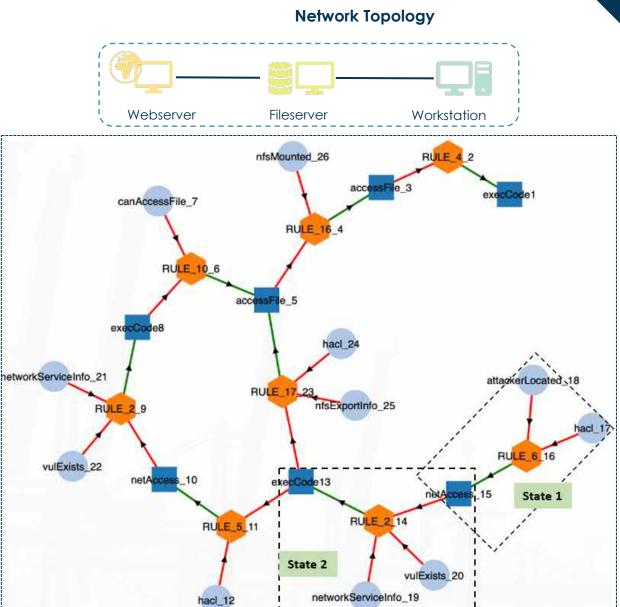
Configuration: Condition, provide possibilities of action by adversary

Rule: Attack methodology attacker can leverage

Impact: Sub-goal achieved by the former action

 Incorporate conditional dependency to transfer attack graph to state graph

Action State

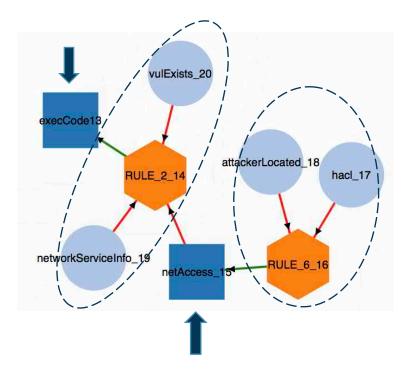


Mapping to Techniques-Tactics

<u>(i)</u>

- ✓ We map each action state to distinct Technique-Tactic(TT) pair
- \checkmark Unfolds the current phase of attack strategy
- ✓ Each attack path eventually exposes a sequence of TT

- Mapping Attack State to Techniquetactic:
 - Use rule and configuration information to map state to technique
 - Impact information to map state to tactic
 - Correlated technique pre-requisite in a sequence to improve mapping accuracy



Mapping to Techniques-Tactics

<u>(Ú)</u>

rule_desc('multi-hop access')

execCode('192.168.15.124', someUser):vulExists('192.168.15.124', 'CVE-2015-2808',
safari,remoteExploit,privEscalation)
networkServiceInfo('192.168.15.124', safari,tcp,
'1433', someUser)
petAccess('102.168.15.124', tep.'1422')

netAccess('192.168.15.124',tcp,'1433')

rule_desc('remote exploit of a server program')

netAccess('192.168.15.124',tcp,'1433'): attackerLocated(internet)
 hacl(internet,'192.168.15.124',tcp,'1433')
rule_desc('direct network access')

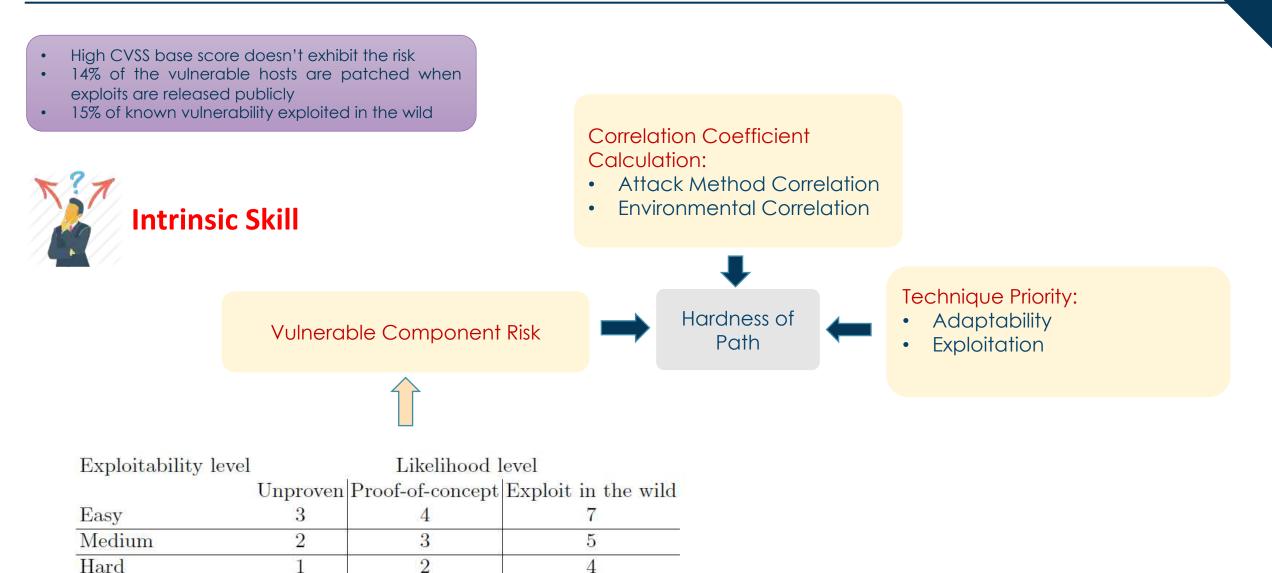
Tactic: lateral movement

Tactic: execution, Technique-Exploitation for Client Execution (T1203)

Tactic: lateral movement, Technique-Exploitation of Remote Service (Ti 210)

Tactic: initial access

Path Complexity and Effort Estimation



Path Complexity and Effort Estimation

Vulnerable Component Risk:

- Unproven
- Proof of concept
- Exploited in the wild

Exploitability level	Likelihood level					
	Unproven	Proof-of-concept	Exploit in the wild			
Easy	H	MH	\mathbf{VH}			
Medium	М	H	MH			
Hard	VL	L	М			

✓ Technique Priority Score:

□ Two factors considered.

 Adaptability: depends on the environment and conditions allowing a technique to be exercised.

$$ASc(ta_t) = pl_t \times \sum_{i=1}^p pr_i^t \times \tau_t$$

 $ExSc(ta_t) = sf_t \times gr_t$

$$TSc(ta_t) = \beta ASc(ta_t) + (1 - \beta) ExSc(ta_t)$$

Path Complexity and Effort Estimation

Ŵ

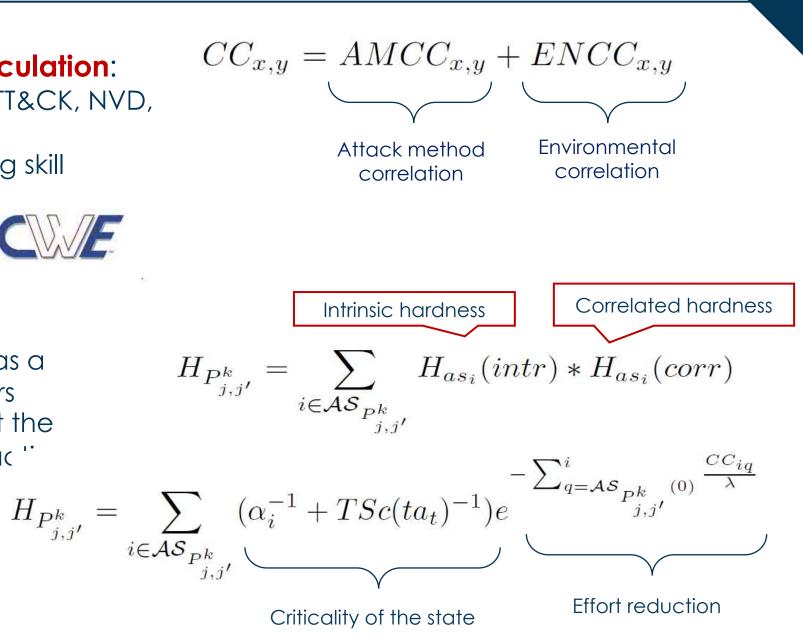
Correlation Coefficient Calculation:

- State integrated with ATT&CK, NVD, CWE
- Track attacker's evolving skill

ATT&CK[™]

✓ Hardness of Path:

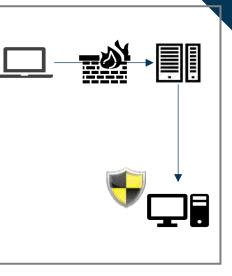
- State hardness is defined as a function of two parameters
- Decay factor (λ) represent the effort reduction in similar ac '.



Path Stealthiness

- Hypothesis behind the stealthiness: The more isolation a path introduces –
 - More detectable by the defender
 - Less exploitable by the attacker
- Adversary Categorization:
 - Persistent or goal specific adversary (tailored attack)
 - Different constraint (time, resource etc.) bounded adversary (commodified attack)
- **Strategic** plan through attack graph help to find the motive of attacker.

- Data Quality parameters (DeTTECT):
 - Data Field Completeness: Indicates to what degree the data has the required information/fields as well data in the field
 - Timeliness: Indicate how accurate the timestamp of the data corresponding to the actual time an event occurred.
 - Consistency: Indicate the correlation with other data sources in terms of data field names and types.







Monitored data source

Data Quality

Measurement

Path Stealthiness Calculation

m

- Operation Pipeline:
 - Generate Attack graph (AG) from network
 - Identify security control
 - Identify data sources monitored by security control
 - Map data sources to AG analytics
 - Path stealthiness calculation

Stealthiness of Path:
$$S_{P_{j,j'}^k}^{th} = S_{as_1}^{th} + S_{as_2}^{th} + \dots + S_{as_m}^{th} = \sum_{i=1}^m S_{as_i}$$

$$S_{as_i}^{th} = \frac{\sum_k \{ (z_{as_i}^k \times \omega_{as_i}^k) + dQ^k \}}{|SC_{as_i}|} \quad S_{P_{j,j'}^k}^{th} = \sum_{i \in \mathcal{AS}_{P_{j,j'}}} S_{as_i}^{th}$$

 $\omega_{as_i}^k$ =Security Control k deployed in action state as_i SC_{as_i} =Required Security Control in as_i

 $dQ_{as_i}^k$ =Data quality of Security Control k deployed in action state as_i

Additional Analytics:

Additional analytics using the Framework



Investigate critical data sources:

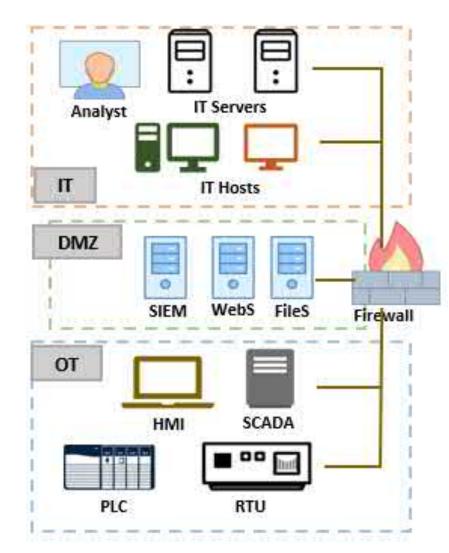
- Detect additional exploited techniques
- Determine new attack paths
- Monitored frequently in the attack graph and determine mitigation plan for attack paths Map critical data sources into security control to effective security control

Cyber Analytics Repository (CAR)



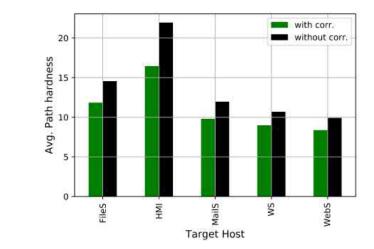
- Impose more granular information from data sources
- Each data model is comprised of {object/action/ field} e.g., {driver, file/ load, create / md5_hash, pid }
- Extract unique data model which should be monitored to prevent all ATT&CK techniques from exploitation Prioritize the defense.

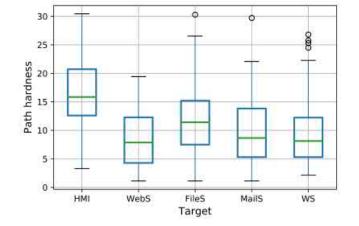
Validation on ICS



□ Active scanning with Nessus for IT network

- Passive scanning with Grassmarlin and ClarOty for OT network.
- Extract attack paths terminating into multiple targets



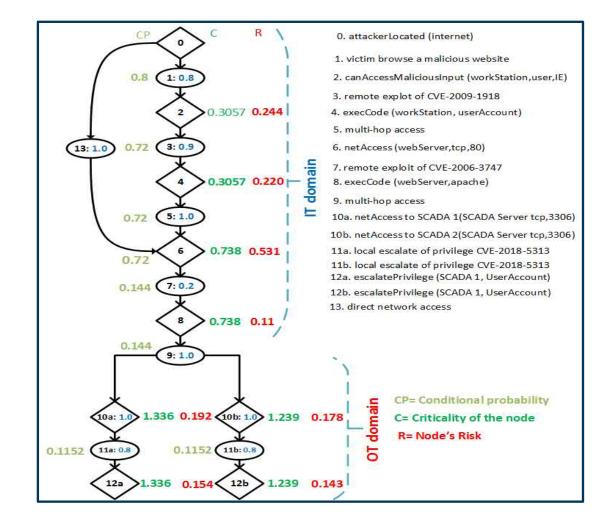


Deviation of Path Hardness

Distribution of Attack path

Accenture ICS Testbed

Attack Graph and Criticality Analysis



- Attack Paths: $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$ $\rightarrow 8 \rightarrow 9 \rightarrow 10a \rightarrow 11a \rightarrow 12a; 0 \rightarrow 13 \rightarrow 6 \rightarrow 7$ $\rightarrow 8 \rightarrow 9 \rightarrow 10a \rightarrow 11a \rightarrow 12a$ - SCADA1 (target).
- Attack Paths: $0 \rightarrow 1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$ $\rightarrow 8 \rightarrow 9 \rightarrow 10b \rightarrow 11b \rightarrow 12b; 0 \rightarrow 13 \rightarrow 6 \rightarrow 7$ $\rightarrow 8 \rightarrow 9 \rightarrow 10b \rightarrow 11b \rightarrow 12b$ SCADA2

(target).

- Though paths have identical exploitation probability from attacker starting node to SCADA1/SCADA2, the damages along the paths are different.
- Attacker has opportunity to analyze' options and select the path that can make the most damage to the target

Modeling Attacker's Intent

Charles Kamhoua, Alexander Kott, Laurent Njilla, Sachin Shetty, "Modeling and Design of Secure Internet of Things", John Wiley & Sons, 1 edition, 2020, ISBN 978-1-119-59336-2

Goal





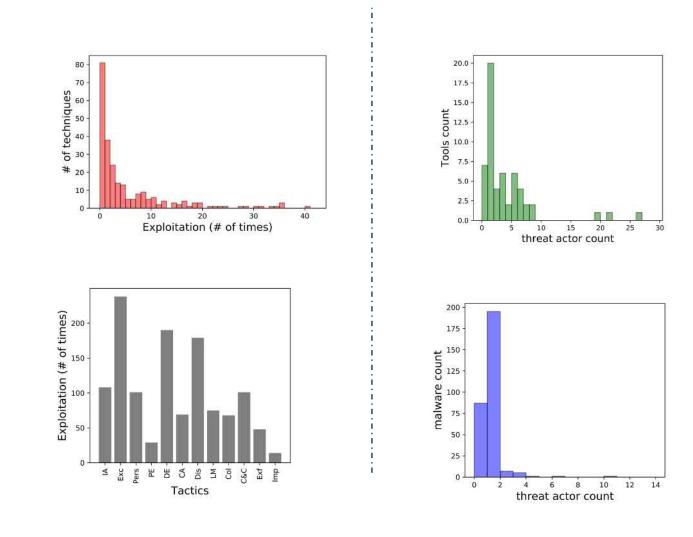
Understand adversary Strategy

Predict attackers' movement into the system using TTP Chain



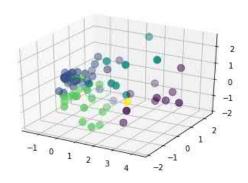
Observe suspicious activity and predict future action

Threat Actor Insight from MITRE

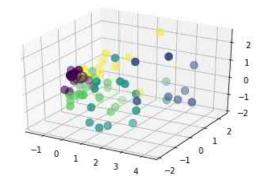


PCA + K-means clustering

<u>Ú)</u>

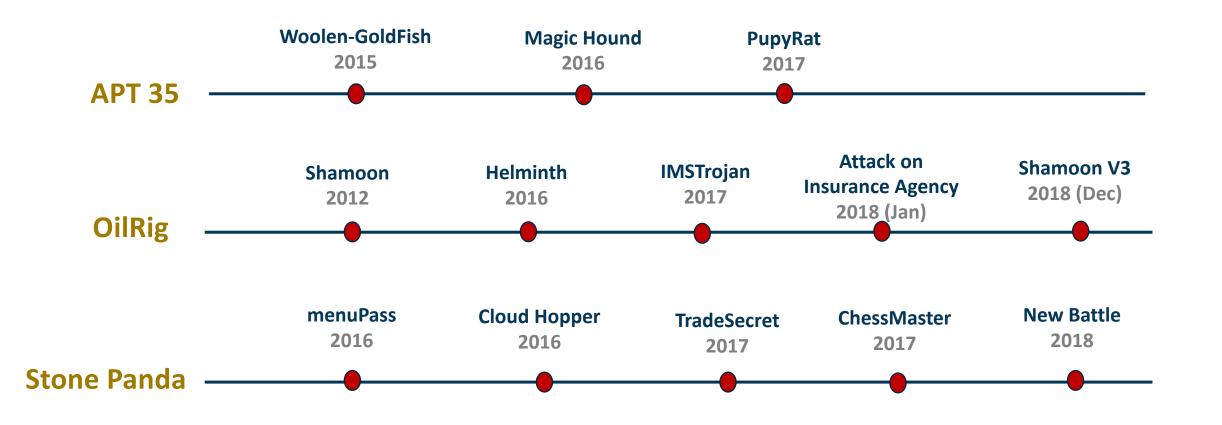


Threat Actor Cluster based on Technique



Threat Actor Cluster based on Tactic

Threat Campaign Information



Ŵ

Ongoing efforts

- The Threat Report ATT&CK Mapper (TRAM) is a web-based tool from MITRE to analyze report and extracting ATT&CK technique.
- It takes the procedure example from ATT&CK to train the model.
- Use logistic regression with tokenized data to match the technique to the report

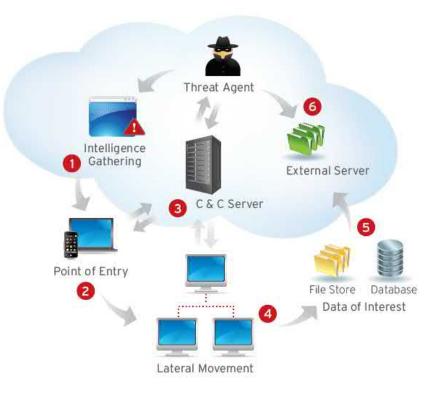
Limitation--

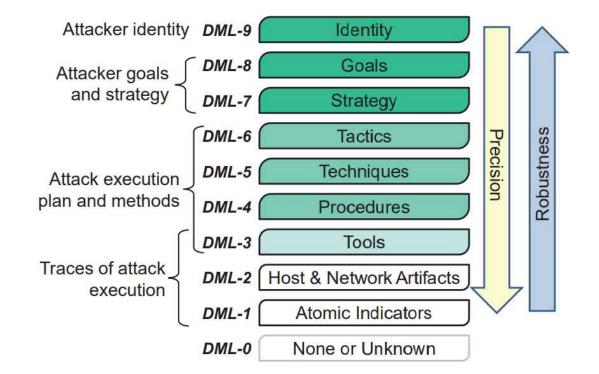
- A beta release. Needs analyst review for better accuracy.
- It only used count vectorizer for feature extraction which could embed lot's of noisy data during training.
- It failed to extract the context of each technique could turn out error prone result.

and Slove Photos			
Enter New Report	NEEDS REVIEW	ANALYST REVIEWING	COMPLETE
Enter URL			
nsert Title	Example Report		
Enter the article title	Source Analyze		



Attack life cycle and defense strategy

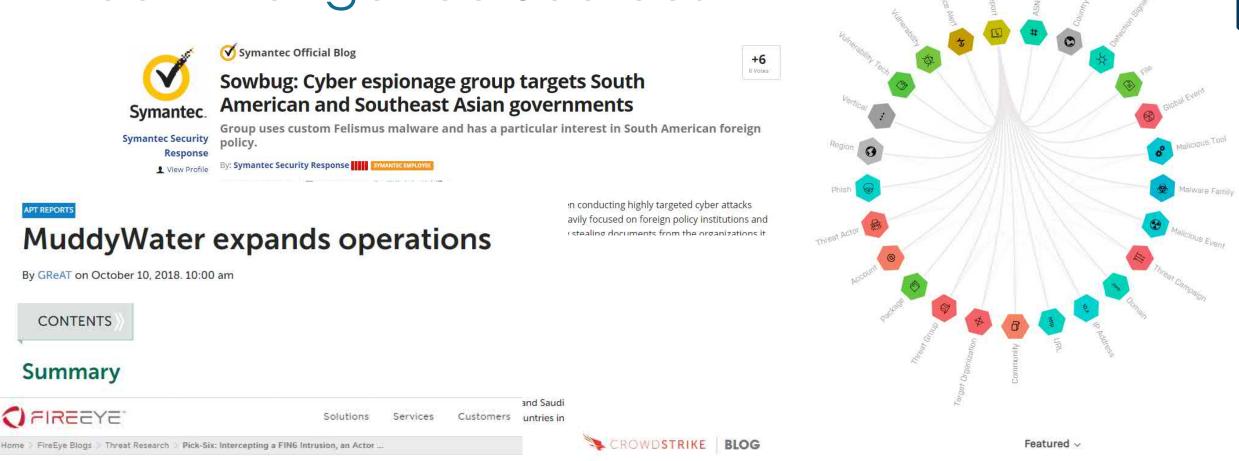




Attack Life Cycle in APT Progression

Detection Maturity Level model

Threat Intelligence Sources



Threat Research

Pick-Six: Intercepting a FIN6 Intrusion, an Actor Recently Tied to Ryuk and LockerGoga Ransomware

April 05, 2019 | by Brendan McKeague, Van Ta, Ben Fedore, Geoff Ackerman, Alex Pennino, Andrew Thompson, Douglas Bienstock

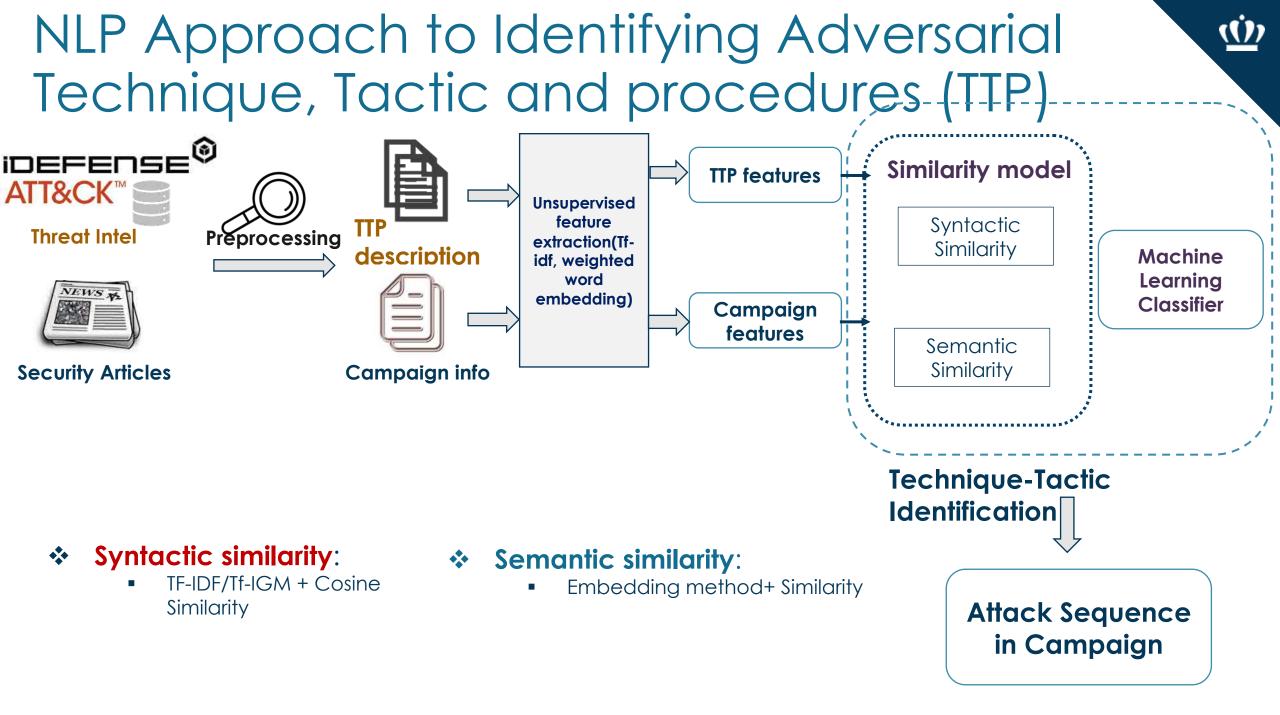
RANSOMWARE MANAGED DEFENSE FIN GROUP

Deep in Thought: Chinese Targeting of National Security Think Tanks

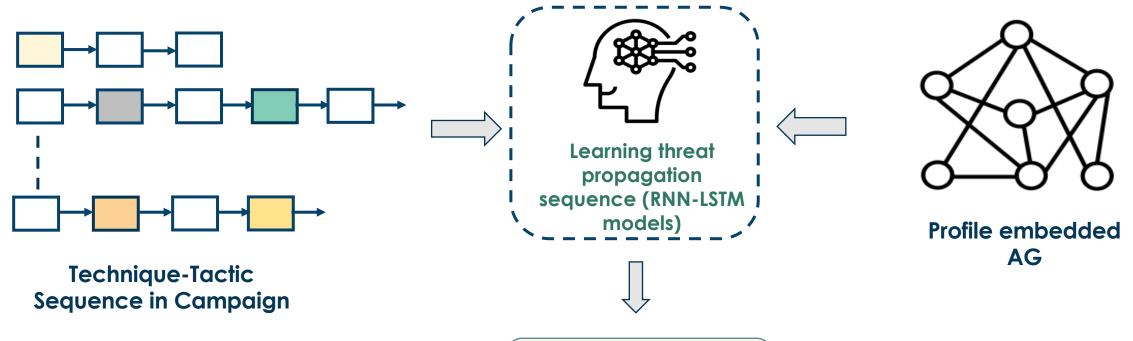
July 7, 2014 Dmitri Alperovitch Executive Viewpoint



1



Threat Propagation Sequence Analysis



Attack Pattern Identification in AG <u>()</u>)

Data Preprocessina

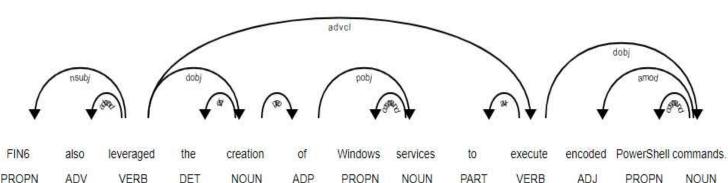
The encoded payload was a Cobalt Strike httpsstager that was injected into the PowerShell process that ran the command. The Cobalt Strike httpsstager was configured to download a second payload from hxxps://176.126.85[.]207:443/7sJh. FireEye retrieved this resource and determined it was a shellcode payload configured to download a third payload from hxxps://176.126.85[.]207/ca. FireEye was unable to determine the final payload due to it no longer being hosted at the time of analysis.

Second technique: FIN6 also leveraged the creation of Windows services (named with a random 16-character string such as IXiCDtPbtGWnrAGQ) to execute encoded PowerShell commands. The randomly named service is a by-product of using Metasploit, which creates the 16-character service by default. The encoded command contained a Metasploit reverse HTTP shellcode payload stored in a byte-array like the first technique. The Metasploit reverse HTTP payload was configured to communicate with the command and control (C2) IP address

Data Scraping and Preprocessing:

- Extract unstructured text data from web (threat reports) and MITRE (threat intel)
- Initially remove noise from the text like advertisement and other unnecessary information by regular expression
- Perform Stemming and Lemmatization: the process of reducing inflection of words in their roots form belong to the dictionary form as well
- Remove Stop words
- Initially extract features by dependency

Dependency Parsing



Feature Indicator of compromise (IoC) Noise + Stop words

Techniques Extraction

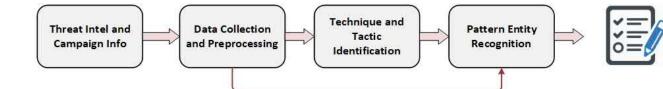
- We extract potential adversary techniques from the threat reports
- TF-IDF is used to put weight on each feature we previously extracted by dependency parsing.
- *TF IDF* = *Term frequency* * *Inverse Document Frequency*
- This process signifies the importance of words in the document and corpus.
- In our model, we investigate which TTP features are more important for a particular technique than others.
- Then cosine similarity is used to measure cosine angle of two vectors
 - TTP feature vectors from threat intel
 - Campaign feature vectors from threat report
- Result shows some of the techniques probability based on our analytics on a specific threat report.

 $tf - idf = tf(t, d) * \log(N/(df + 1))$

 $df = Occurance \ of \ t \ in \ docuemnts$ t - term(word), d $- document(set \ of \ words), N$ $- count \ of \ corpus, corpus$ $similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$

Application Window Discovery	0.079472
Binary Padding	0.249444
Fallback Channels	0.253859
System Service Discovery	0.121529
File System Logical Offsets	0.039606
Data from Local System	0.118763
Winlogon Helper DLL	0.133888
Credential Dumping	0.199588
Data Compressed	0.106594
Data Obfuscation	0.280427

Tool Architecture



- Multistage learning pipeline to mine threat intel resources to unfold attack pattern
- Convert every letter to lowercase an stopwords like 'for', 'the', 'to' tc.
- Reduce noise-, ; : .
- Trun the word in root form
- Multi-word frequently co-occur togethertokenize together

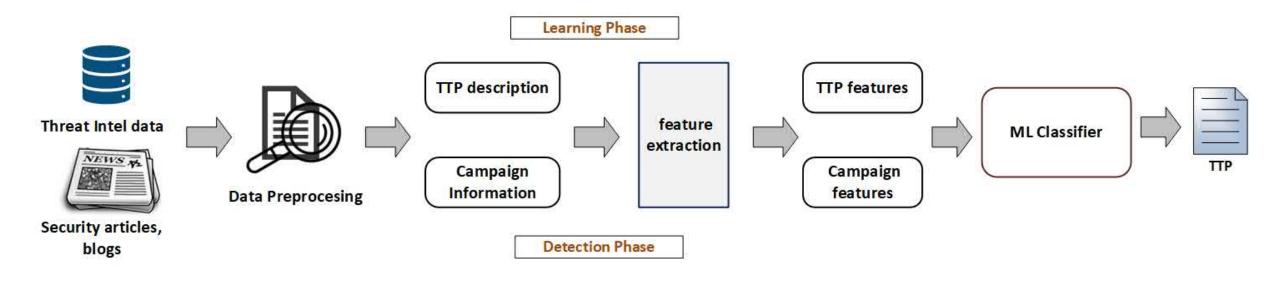
Adversarial Attack Pattern Learning Pipeline



Data Collection and Preprocessing

Multi-word Expressions					
white lambe	rt, peppy trojan, moonwind rat, royal dns, metasploit stager,				
black lambe	rt, sakula rat, googledrive rat, apt3 keylogger, havex rat,				
poison ivy,	byebye shell, blue lambert, cobra carbon system				

Technique-tactic Identification



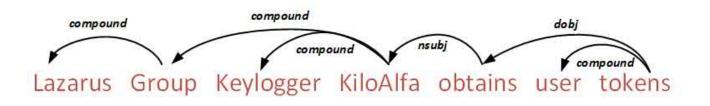
<u>()</u>)

Feature Extraction-Bag of words

- Form One-hot vector corresponding to every term
 - Term frequency Inverse document frequency (Tf-IDF): $f_{t,d}$ is number $TF IDF_{t,d} = tf_{t,d}.log \frac{N}{df_t}$ of times the term appear t in document d
 - Term frequency Inverse gravity Moment (TF-IGM): $f_{k,r}$ frequency of term occurring t_k in different class, which sorted in descending order, r is the rank

$$igm(t_k) = \frac{f_{k1}}{\sum\limits_{r=1}^n f_{kr}.r}$$

Feature Extraction-Sentence Embedding



- Normalize the word vectors in the sentence
- Two types of word embedding is used-
- Glove Embedding- Global word-word cooccurrence matrix
- Dependency based embedding Use Skip Gram model. Linear context to arbitrary context

Words	Contexts
Lazarus	$Group/compound^{-1}$
Group	KiloAlfa/compound ⁻¹
Keylogger	KiloAlfa/ $compound^{-1}$
KiloAlfa	Group/compound, Keylogger/compound, obtains/nsubj ⁻¹
obtains	KiloAlfa/nsubj, tokens/dobj ^{-1}
user	$tokens/compound^{-1}$
tokens	obtains/dobj, user/compound

Technique and Tactic Classification

		Accuracy	Precision	Recalls	F1-Score
Model	Features	LR/SVM	LR/SVM	LR/SVM	LR/SVM
Bag of Words	Tf-IDF	51.72(cos sim)	C2	<u>2</u>	<u></u>
Bag of words	Tf-IGM	53.83(cos sim)	-	ā.	
	Universal	58.63/54.28	58.24/56.41	58.63/54.28	56.39/53.9
Word Embedding (Glove)	Universal(Tf-IDF Weighted)	72.65/65.57	74.62/69.95	72.65/65.57	71.68/65.23
	Pre-trained	54.69/53.60	53.50/54.55	54.69/53.60	52 31/52 17
	Pre-trained(Tf-IDF Weighted)	74.01/66.12	74.60/70.56	74.01/66.12	72.38/65.72
Dapandanay basad Embaddina	Pre-trained	49.79/45.17	48.95/47.24	49.79/45.17	46/44.46
Dependency based Embedding	Pre-trained(Tf-IDF Weighted)	47.75/46.53	45.82/46.69	47.75/46.53	<mark>43.66/44.5</mark> 9

- BoW formed a very sparse vector- cosine similarity used
- Universal trained on general corpus
- Pre-trained trained on our corpus
- 6600 sentence for training and 1050 for testing
- Topical dependency is more relevant than functional dependency
- High dimension turns out more distinct feature

Dimensions	Accuracy	Precision	Recalls	F1-Score
50	65.03	69.77	65.03	64.77
100	69.25	72.72	69.25	68.40
300	72.65	74.62	72.65	71.68

Performance with different word embedding dimension (Glove)

<u>(</u>)

Pattern Entity Recognition-Annotation

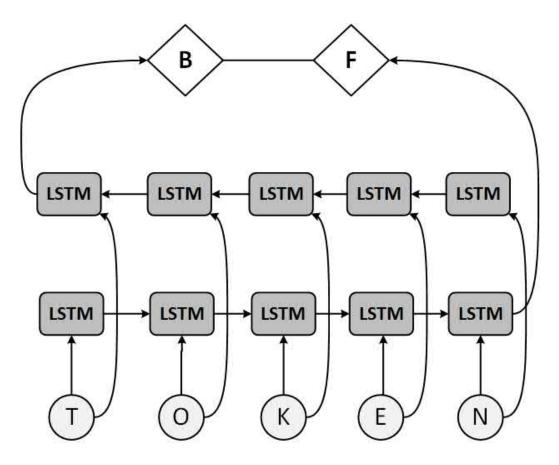
Label	# Description Does not contain useful information				
0					
Action	Action performed in cyber campaign				
Intent	Motive of an action				
Tool	Utilities and tools used in cyber campaign				
Conf	Configuration facilitate malicious action				
Action Object	Surface of action				
Intent Object	Surface of potential action or objective				

Systemd services can be used to establish persistence on a Linux system. The systemd service manager is commonly used for managing background daemon processes and other system resources. Systemd is the default initialization system on many Linux distributions starting with Debian 8, Ubuntu 15.04, CentOS 7, RHEL 7, Fedora 15, and replaces legacy init systems including SysVinit and Upstart while remaining backwards compatible with the aforementioned init systems. Systemd utilizes configuration files known as service units to control how services boot and under what conditions. By default, these unit files are stored in the /etc/systemd/system and /usr/lib/systemd/system directories and have the file extension .service. Each service unit file may contain numerous directives that can execute system commands ExecStart, ExecStartPre, and ExecStartPost directives cover execution of commands when a services is started manually by systemctl or on system start if the service is set to automatically start. ExecReload directive covers when a service restarts ExecStop and ExecStopPost directives cover when a service is stopped or manually by systemctl. Adversaries have used systemd functionality to establish persistent access to victim systems by creating and/or modifying service unit files that cause systemd to execute malicious commands at recurring intervals, such as at system boot. While adversaries typically require root privileges to create/modify service unit files in the /etc/systemd/system and /usr/lib/systemd/system directories, low privilege users can create/modify service unit files in directories such as ~/.config/systemd/user/ to achieve user-level persistence. The adversary has established persistence using a systemd service. The adversary has a hardcoded location under system that it uses to achieve persistence if it is running as root. The adversary can be used to establish persistence using a systemd service.



Pattern Entity Recognition – Charembedding

- Threat Intel often has out-of-vocabulary token
- Mostly software, malware, threat actors
- Concatenation of forward and backward representation



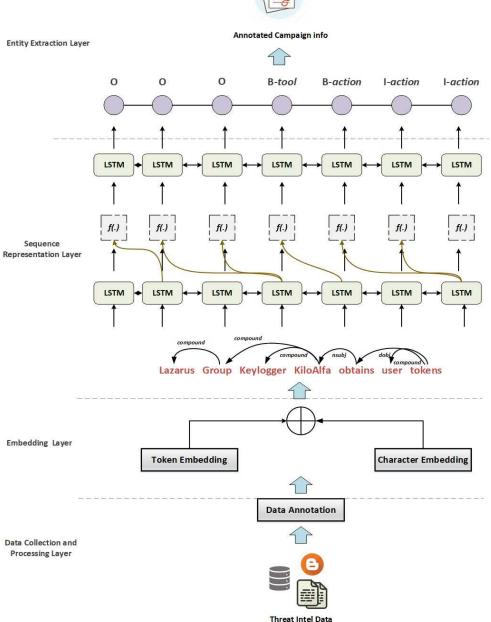
Neural Network Architecture for PER



Word vector is concatenated with character vector

- Input embedding $e = [w_i, w_h, d_r]$
- BiLSTM captures contextual information
- Interaction function captures the interaction between word and parents
- For higher layer BiLSTM $H^{l+1} = BiLSTM(H^l)$
- Score from BiLSTM $score(y, x) = \sum_{i=1}^{n} p_i[y_i] + \sum_{i=0}^{n} T[y_i, y_{i+1}]$
- $CRF P(y|x) = \frac{\exp(score(y,x))}{\sum_{y} \exp(score(y,x))}$

Interaction Function	$f(\mathbf{h_i}, \mathbf{h_{pi}})$
Self connection	hi
Concatenation	$\mathbf{h_i} \oplus \mathbf{h_{pi}}$
Addition	$h_i + h_{pi}$



Data Preparation

- Transform the annotated data into BIO 'Begin', 'Inside', 'Outside' schema.
- Then transform into CONLL-X format

Adversaries may also _ compromise shared _ network _ directories	_ _ _ _	 	4 - - -	_ 4 nsubj 0 aux 0 4 advmod 0 _ 0 ROOT B-intent 7 amod I-intent 7 compound I-intent 4 dobj I-intent
	_	-	-	
through _	-	-	-	
binary _	_	-	-	10 amod B-tool
infections	_	_	4	_ 8 pobj I-tool
by	_	_	4	
appending	_	_	_	_ 11 pcomp B-action
or	_	_	12	
prepending	_	_	_	_ 12 conj I-action
its	_	_	16	poss I-action
code _	_	_	_	14 dobj I-action
to	_	_	14	prep I-action
the	_	_	20	det I-action
healthy _	_	_	_	20 amod I-action
binary				17 pobj I-action
on	_	_	20	prep 0
the	_	_	25	det 0
shared	_	_		25 amodB-action_object
network	-	-	-	25 compound I-action_object
directory	-	-	-	_ 21 pobj I-action_object
all cecory	-	-	4	punct 0
·	_	-	-	Punce 0

<u>(</u>

PER- results



Hyper-parameter	value	Test Set				
		Accuracy	Precision	Recalls	F1-Score	
Word Embedding	Glove (100d)	88.61	60.5	61.84	61.17	
word Embedding	Cyber-embedding (100d)	88.14	63.79	59.88	61.77	
L CTM L	Layer-1	88.61	60.5	61.84	61.17	
LSTM Layer	Layer-2	89.23	60.55	62.70	61.60	
Ontimizar	SGD	88.61	60.5	61.84	61.17	
Optimizer	Adam	88.66	58.34	60.49	59.40	

			Test Set				
		Dimensions	Accuracy	Precision	Recalls	F1-Score	
	22	100	88.61	60.55	61.84	61.17	
Use universal word embedding Use concatenation as interaction function	Word Embedding	200	89.41	62.24	59.88	61.04	
		300	88.71	60.62	55.34	57.86	
	station of the second s	100	89.04	61.68	59.63	60.64	
• Unlike TTp adding more learning state doesn't	nelest unit	150	88.05	62.67	53.87	57.76	
for pattern entity recognition		200	88.61	60.55	61.84	61.17	

CYBER RISK MANAGEMENT

PROACTIVE AND REACTIVE APPROACHES

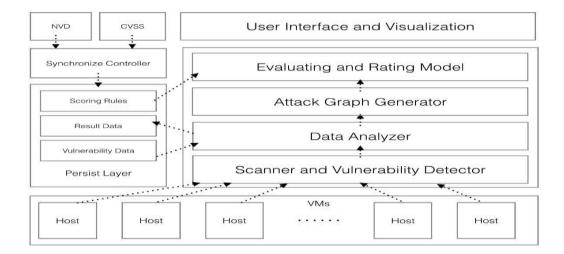


RECOMMENDATION ENGINE CONTEXTUALIZATION IMPACT ANALYSIS THREAT SECURITY ATTACK GRAPH GENERATOR EVENTS INTELLIGENCE NETWORK ANALYSIS **VULNERABILITY ANALYSIS** ASSET DISCOVERY

<u>(</u>

accenturesecurity

Cyber RIsk Scoring and Mitigation (CRISM©)



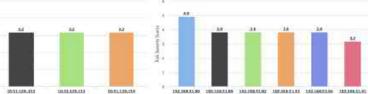


10.11.129 131

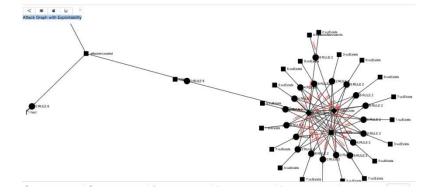
Risk scores of the servers in your cloud environment



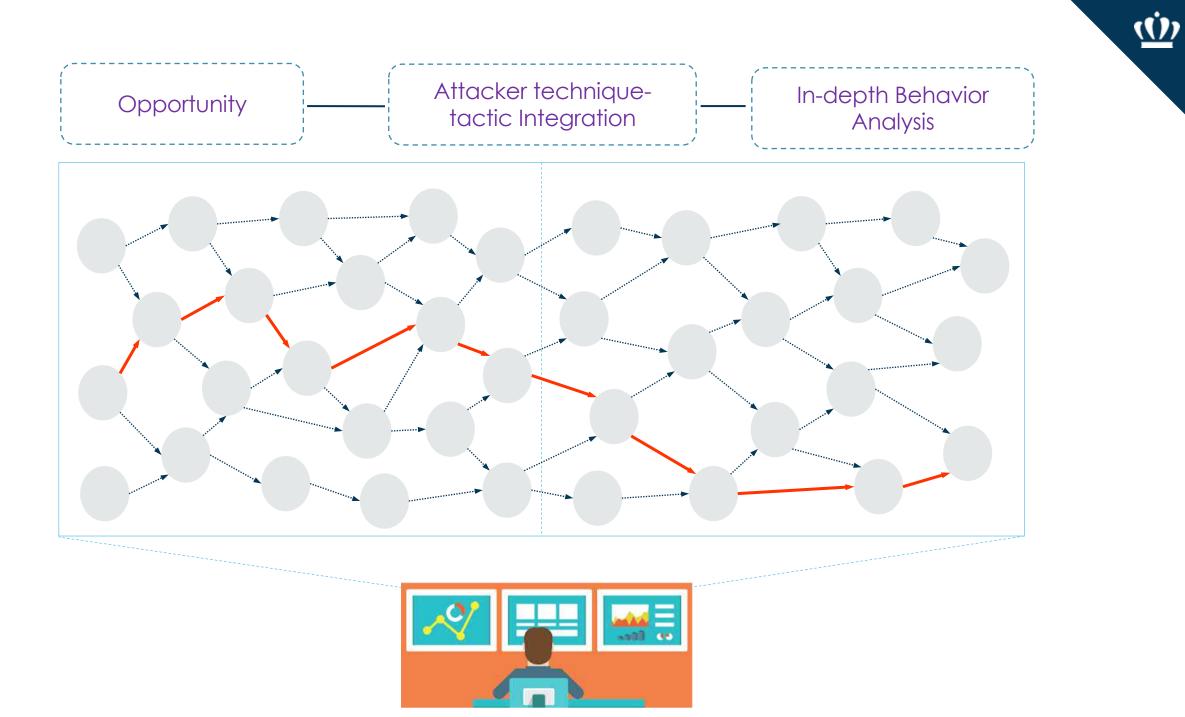
Risk scores of the servers in your cloud environment City the planets to slow extendedly deals.







Sachin Shetty, Michael McShane, Linfeng Zhang, Jay Kesan, Charles A. Kamhoua, Kevin Kwiat, Laurent Njilla, "<u>Reducing Informational</u> <u>Disadvantages to Improve Cyber Risk Management</u>", Geneva Papers on Risk and Insurance, April 2018, Volume 43, Issue 2, pp 224–238 Marco Gamarra, Sachin Shetty, Oscar Gonzalez, David Nicol, Charles A. Kamhoua, Laurent Njilla, "Analysis of Stepping Stone Attacks in Dynamic Vulnerability Graphs," IEEE International Conference on Communications (ICC) 20-24 May 2018, Kansas City, MO



Thank You

Sachin Shetty, Ph.D. Email – <u>sshetty@odu.edu</u> Web – <u>www.odu.edu/~sshetty</u>