International Council on Systems Engineering
A better world through a systems approach

Toward Quantitative Assessments of Cybersecurity Countermeasure Efficacy

Ben Breisch (bbreisch@mitre.org)
Kristin Voss (kevoss@mitre.org)

Today's Agenda

- Problem
- Background
- Efficacy + Efficacy Properties
- Rooting Efficacy in D3FEND™
- Weighting
- Countermeasure Comparison
- Future Work

Problem

How can organizations ***objectively*** prioritize countermeasure investment?

- With limited budgets, organizations must prioritize defensive measures to invest in
  - Efficacy in security is typically subjective, based on Subject Matter Expert (SME) input and system architecture
  - What properties of countermeasures do organizations care about? What metrics can you use?
- A set of countermeasures that works for one system != success in another

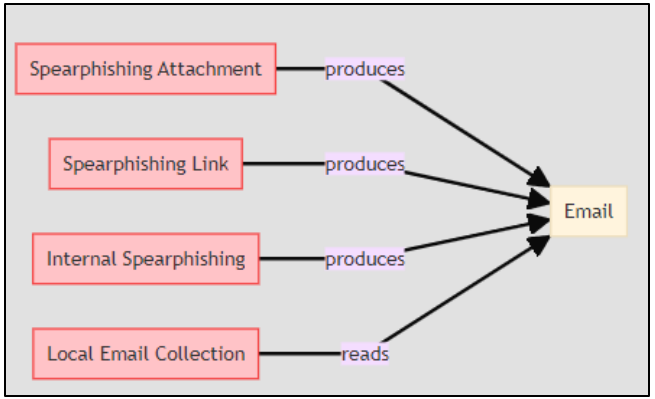


Figure 1: Offensive Technique to Artifact
(*MITRE D3FEND - Email, n.d.*)

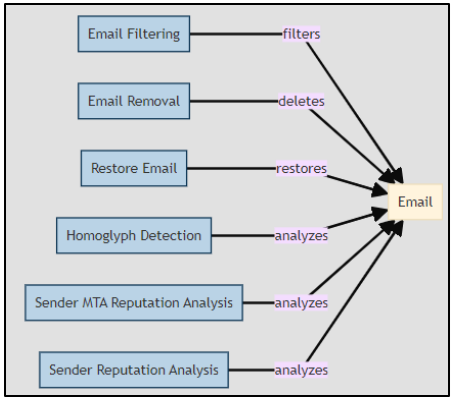
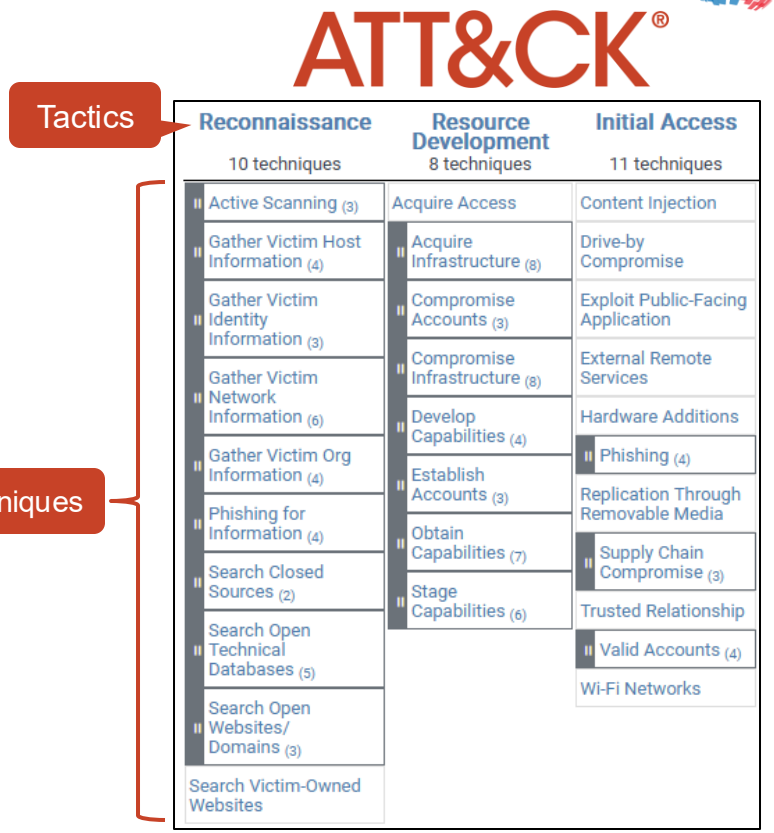


Figure 2: Defensive Technique to Artifact
(*MITRE D3FEND - Email, n.d.*)

Background: MITRE ATT&CK©

What is MITRE ATT&CK?

- ATT&CK © is a cybersecurity framework that organizes adversary actions into Tactics, Techniques, and Procedures (TTPs)
- TTPs based on real-world observations of cyber threats
  - Advanced Persistent Threat (APT) groups, ransomware gangs, etc.
- Breaks down cyber offensive actions into **14 high-level tactics**



(Note: 3 of 14 ATT&CK Tactics shown)

Figure 3: ATT&CK Matrix (MITRE ATT&CK, n.d.)

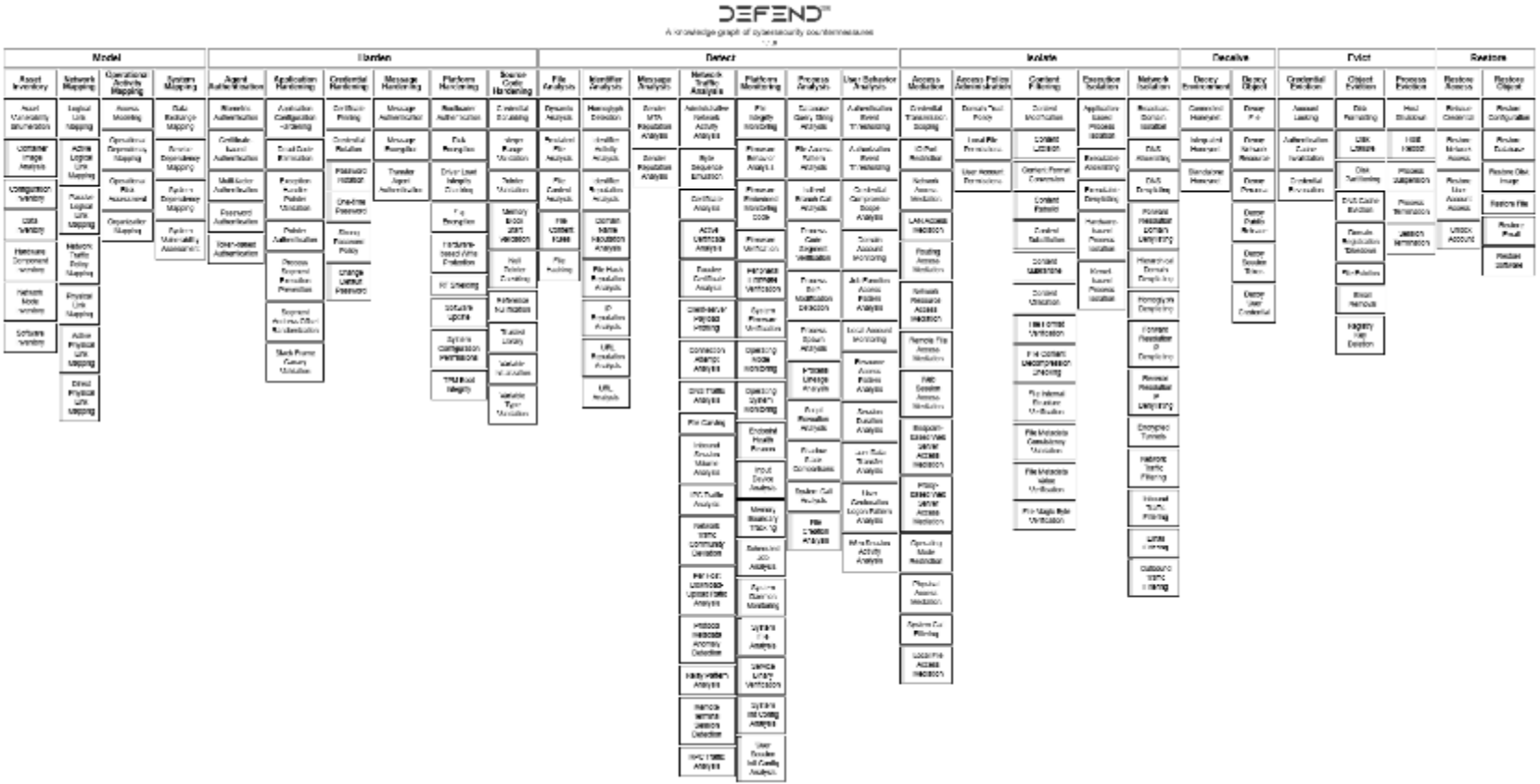
Background: MITRE D3FEND™

What is D3FEND?

- An ontology of **defensive countermeasures and techniques**, with mappings to **ATT&CK TTPs** via **Digital Artifacts**
- Breaks down cyber defensive countermeasures into **7 high-level tactics**
- Tactics used for this problem: **Harden, Detect, Isolate, Evict, and Restore**
- Countermeasures can be grouped into D3FEND tactics based on techniques

Tactics
Model
Harden
Detect
Isolate
Deceive
Evict
Restore

Background: MITRE D3FEND™



https://d3fend.mitre.org

Figure 4: D3FEND Matrix (MITRE D3FEND, n.d.)

Tactic

Harden

Hierarchical Techniques

Agent Authentication	Application Hardening	Credential Hardening	Message Hardening	Platform Hardening	Source Code Hardening
Biometric Authentication	Application Configuration Hardening	Certificate Pinning	Message Authentication	Bootloader Authentication	Credential Scrubbing
Certificate-based Authentication	Dead Code Elimination	Credential Rotation	Message Encryption	Disk Encryption	Integer Range Validation
Multi-factor Authentication	Exception Handler Pointer Validation	Password Rotation	Transfer Agent Authentication	Driver Load Integrity Checking	Pointer Validation
Password Authentication	Pointer Authentication	One-time Password		File Encryption	Memory Block Start Validation
Token-based Authentication	Process Segment Execution Prevention	Strong Password Policy		Hardware-based Write Protection	Null Pointer Checking
	Segment Address Offset Randomization	Change Default Password		RF Shielding	Reference Nullification
	Stack Frame Canary Validation			Software Update	Trusted Library
				System Configuration Permissions	Variable Initialization
				TPM Boot Integrity	Variable Type Validation

Figure 5: Harden Tactic (MITRE D3FEND, n.d.)

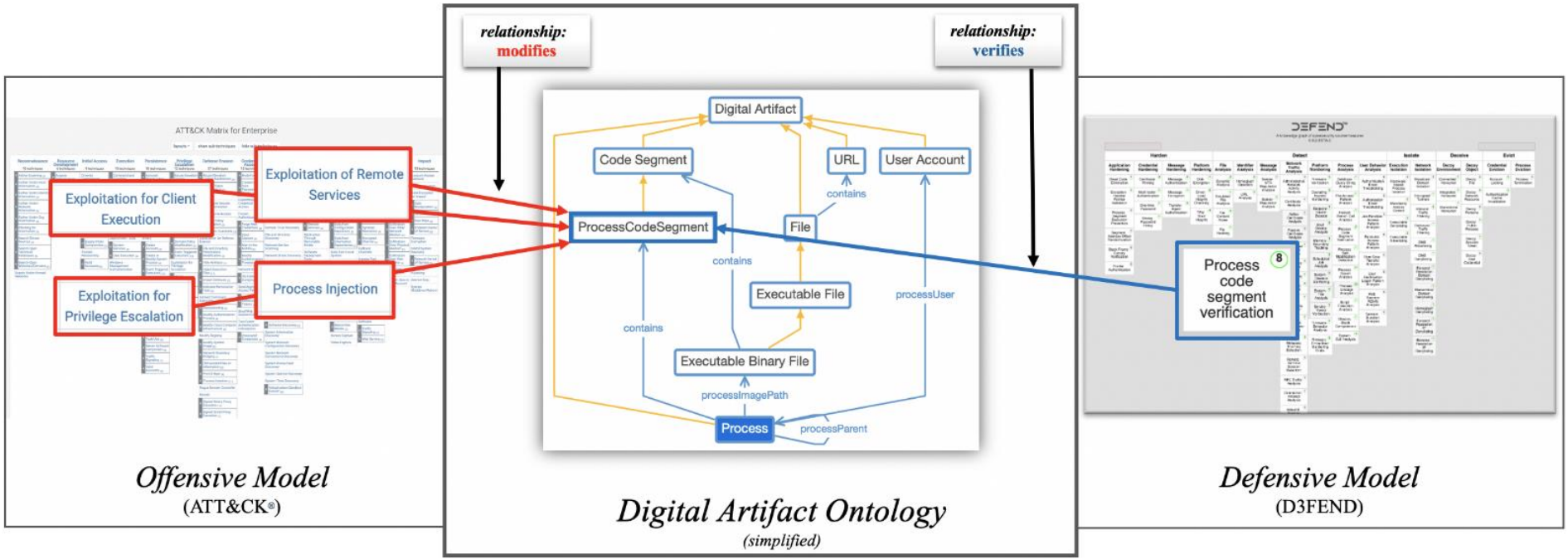


Figure 6: D3FEND ontology (Kaloroumakis & Smith, 2021)

Background: MITRE D3FEND™

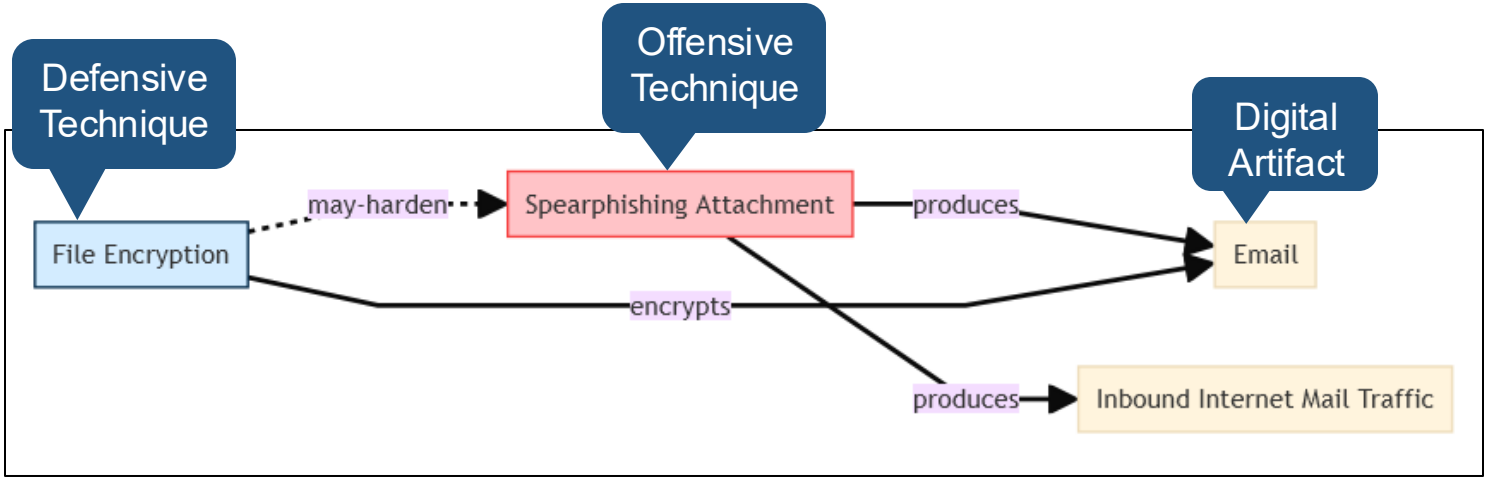


Figure 7: Ontology Example (*MITRE D3FEND - Email, n.d.*)

Relationships between offensive and defensive techniques are inferred through artifacts

Background: Efficacy

Medicine

- Medical trials quantify treatment **efficacy** using **quantitative metrics**
- Metrics include comparisons of Key Performance Indicators (KPIs) before and after treatment across placebo/non-placebo groups
- Treatment performance isn't skewed by individual doctor or patient biases
- Efficacy = how treatment performs against placebos and alternatives

(Zwarenstein, et al., 2008)

Agriculture

- Pesticide efficacy is measured with multi-dimensional metrics
- Pest density before/after application across different target species
- Validity of metrics highly determined by physical environment / region

(Sudo, et al., 2019)

Many industries measure efficacy using quantitative multi-dimensional metrics

Existing Security Metrics

Return on Security Investment (ROSI)

$$ROSI = \frac{Cost\ Avoided - Cost\ of\ Investment}{Cost\ of\ Investment} \times 100\%$$

- What losses would be expected by a breach or incident?
- Does the cost to implement the countermeasure outweigh the potential loss?
- **What is the monetary cost of mission failure?**

(Sonnenreich, Albanese, & Stout, 2006)

Risk Reduction

- How much risk is mitigated by deploying a countermeasure?
- Risk is often measured subjectively on SME input
- Risk Management Framework (RMF) checkbox mentality
  - **Compliant** vs. **Secure**

(Ross, 2018)

Efficacy Properties

- Other industries fuse data from multiple metrics for cohesive analysis
- What kinds of properties/metrics do security teams care about?
- Defined and categorized potential properties of efficacy a countermeasure could have
- Can sum metrics of properties to reach an efficacy “score”

$$E_{countermeasure} = \sum_{property \in E} E_{property}$$

$$= E_{speed} + E_{accuracy} + \cdots + E_{usability}$$

Background: Efficacy Properties

(Note: 3 out of X properties shown)

Property	Definition
Accuracy	How close a countermeasure's measurement is to the ground truth. (Picus Labs, 2023)
Deployment Speed	How quickly a countermeasure can be implemented within an infrastructure and available to users.
Response Time	How quickly a countermeasure produces a result. Examples: Mean Time to Detection (MTTD), Mean Time to Respond (MTTR)

Table 1: Selected Efficacy Properties

Accuracy

How close a countermeasure's measurement is to the ground truth. (Picus Labs, 2023)

- Considerations:
  - Easiest property to quantify, but a variety of analysis approaches
  - What kind of weight should be given to false positives? false negatives?
- Metrics:
  - Binary Classification: **Did you detect X or not?**
    - Example: F1 Score, Precision, Recall, etc.
  - Regression: **How close is your measurement to the ground truth?**
    - Example: Mean-square Error, Confidence Intervals, etc.

$TP = \#\text{ of true positives}$
$TN = \#\text{ of true negatives}$
$FP = \#\text{ of false positives}$
$FN = \#\text{ of false negatives}$

$$\text{Precision (P)} = \frac{TP}{TP + FP}$$

$$\text{Recall (R)} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2\,(P*R)}{P+R}$$

Deployment Speed

How quickly a countermeasure can be implemented within an infrastructure and available to users.

- Considerations
  - How much configuration is required?
  - Built-in to system or added on later?
  - Is the countermeasure deployed to every device or to a central location?
- Metrics: **Mean**, **median**, **max**, **min**, etc.

$$\text{mean} = \frac{Speed_1 + Speed_2 + \cdots Speed_n}{n}$$

Response Time

How quickly a countermeasure produces a result.

- Considerations
  - Once a countermeasure is deployed, how long to start producing results?
- Metrics to measure response time may vary across D3FEND tactics
  - **Detection - Mean Time to Detect** – time between an event occurring and its detection
  - **Isolate - Mean Time to Contain** – time between discovering an event and isolating an attacker
  - **Restore - Mean Time to Restore** – time to restore a system after an incident
- Other Metrics:
  - **Time to Baseline Performance** – time between countermeasure deployment and reaching some baseline performance
  - **Attacker Dwell Time** – average time that an attacker has access to a system or environment
- Combining multiple metrics into a single property score is user-dependent

Rooting Efficacy Properties in D3FEND

- Group these properties into D3FEND tactics
- Some properties only apply to some tactics
- Organizations may not care about every property equally

Property \ D3FEND Tactic	Harden	Detect	Isolate	Evict	Restore
Deployment Speed	H_{speed}	D_{speed}	I_{speed}	E_{speed}	R_{speed}
Response Time (RT)	H_{RT}	D_{RT}	I_{RT}	E_{RT}	R_{RT}
Scalability	$H_{scalable}$	$D_{scalable}$	$I_{scalable}$	$E_{scalable}$	$R_{scalable}$
Flexibility	$H_{flexible}$	$D_{flexible}$	$I_{flexible}$	$E_{flexible}$	$R_{flexible}$
Preventative	$H_{preventative}$	$D_{preventative}$	$I_{preventative}$	-	-
Accuracy	-	$D_{accuracy}$	-	-	-
Resilience	$H_{resilience}$	-	-	$E_{resilience}$	$R_{resilience}$
Integration	$H_{integration}$	$D_{integration}$	-	-	-
Usability	$H_{usability}$	$D_{usability}$	$I_{usability}$	$E_{usability}$	$R_{usability}$

Table 2: Property – Tactic Mapping

Weighting Properties

- Organizations have a wide variety of architectures, risk tolerances, etc.
- Need a **traceable** and **repeatable** way to fuse quantitative metrics with organizational priorities
- Weight matrix for each property grouped by D3FEND tactic

Property \ D3FEND Tactic	Harden
Deployment Speed	w_{speed}
Response Time	w_{RT}
Scalability	$w_{scalable}$
Flexibility	$w_{flexible}$
Preventative	$w_{preventative}$
Accuracy	-
Resilience	$w_{resilience}$
Integration	$w_{integration}$
Usability	$w_{usability}$

Table 3: Property Weights

Weights Example

Property	Rank	Harden Tactic
Deployment Speed	1	w_{speed}
Integration	1	$w_{integration}$
Response Time (RT)	2	w_{RT}
Flexibility	3	$w_{flexible}$
Scalability	4	$w_{scalable}$
Resilience	5	$w_{resilience}$
Preventative		$w_{preventative}$
Usability		$w_{usability}$

Table 4: Weight Example

Notional weighting scheme prioritizing deployment speed, integration between tools

$$H_{countermeasure} = \sum_{property \in H} C_{property} * w_{property}$$

Comparison

Weight matrix × Property Value matrix

Property	Rank	Harden Tactic
Deployment Speed	1	w_{speed}
Integration	1	$w_{integration}$
Response Time (RT)	3	w_{RT}
Flexibility	3	$w_{flexible}$
Scalability	4	$w_{scalable}$
Resilience	5	$w_{resilience}$
Preventative	-	$w_{preventative}$
Usability	-	$w_{usability}$

Table 5: Weight Example

Property	Capability 1	Capability 2
Deployment Speed	$C1_{speed}$	$C2_{speed}$
Integration	$C1_{integration}$	$C2_{integration}$
Response Time (RT)	$C1_{RT}$	$C2_{RT}$
Flexibility	$C1_{flexible}$	$C2_{flexible}$
Scalability	$C1_{scalable}$	$C2_{scalable}$
Resilience	$C1_{resilience}$	$C2_{resilience}$
Preventative	$C1_{preventative}$	$C2_{preventative}$
Usability	$C1_{usability}$	$C2_{usability}$
Total	$C1_{total}$	$C2_{total}$

Table 6: Property Value + Weighted Sum

Impacts

***Begins* to answer the question of objective countermeasure prioritization**

- Limits qualitative input (human bias) to the weight matrix and metrics chosen
- Can trace and repeat how a particular score was reached
- Need additional thoughts and work on this problem
- **Not** a complete solution

Potential to be a useful framework for analysis of alternatives (AoA) of cyber countermeasures

Future Work

- Expand list of properties to incorporate more industry standards
- Test & validate property measurement methods
- Create a countermeasure case study using each property calculation
  - Provides an example on how to use each property and its weighting
- Test results of efficacy framework on real system
  - Comparative analysis with SME based recommendations
  - Do the results differ from SME recommendations? Why?

References

Kaloroumakis, P. E., & Smith, M. J. (2021). *Toward a knowledge graph of cybersecurity countermeasures*. The MITRE Corporation. Retrieved from https://d3fend.mitre.org/resources/D3FEND.pdf

MITRE ATT&CK®. MITRE ATT&CK. (n.d.). https://attack.mitre.org/

MITRE D3FEND. D3FEND Matrix. (n.d.). https://d3fend.mitre.org/

MITRE D3FEND - Email. Artifact Details | MITRE D3FEND. (n.d.). https://d3fend.mitre.org/dao/artifact/d3f:Email/

Picus Labs. (2023, June 23). *What Is Security Control Effectiveness?* Retrieved from Picus Security: https://www.picussecurity.com/resource/glossary/what-is-security-control-effectivess

Ross, R. (2018). *Risk Management Framework for Information Systems and Organizations: A System Life Cycle Approach for Security and Privacy.* Gaithersburg: Special Publication (NIST SP), National Institute of Standards and Technology. doi:https://doi.org/10.6028/NIST.SP.800-37r2

Sonnenreich, W., Albanese, J., & Stout, B. (2006). Return on security investment (ROSI)-a practical quantitative model. *Journal of Research and practice in Information Technology, 38*, 45-56. doi:10.3316/ielapa.937199632104879

Sudo, M., Yamanaka, T., & Miyai, S. (2019). Quantifying pesticide efficacy from multiple field trials. *Population Ecology, 61*(4), 450-456. doi:10.1002/1438-390X.12019

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Hayes, B., . . . Moher, D. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ, 337*, a2390. doi:10.1136/bmj.a2390

Questions?

35th Annual INCOSE international symposium

hybrid event

Ottawa, Canada
July 26 - 31, 2025

The Team

Ben Breisch

Systems Security Engineer

bbreisch@mitre.org

The MITRE Corporation

Kristin Voss

Cybersecurity Engineer

kevoss@mitre.org

The MITRE Corporation

Will Barnum

Principal Systems Security Engineer

wbarnum@mitre.org

The MITRE Corporation