A RELEVANCE MODEL FOR THREAT-CENTRIC RANKING OF CYBERSECURITY VULNERABILITIES

BY CORREN MCCOY, ROSS GORE, MICHAEL L. NELSON, AND MICHELE C. WEIGLE

(PHOTO SOURCE: ICONPIX AND ACCOUNTANZ [CANVA])

SUMMARY

The relentless process of tracking and remediating vulnerabilities is a top concern for cybersecurity professionals. The key challenge is trying to identify a remediation scheme specific to in-house, organizational objectives. Without a strategy, the result is a patchwork of fixes applied to a tide of vulnerabilities, any one of which could be the point of failure in an otherwise formidable defense. Given that few vulnerabilities are a focus of real-world attacks, a practical remediation strategy is to identify vulnerabilities likely to be exploited and focus efforts toward remediating those vulnerabilities first.

The goal of this research is to demonstrate that aggregating and synthesizing readily accessible, public data sources to provide personalized, automated recommendations for organizations to prioritize their vulnerability management strategy will offer significant improvements over using the Common Vulnerability Scoring System (CVSS). A framework is provided for vulnerability management specifically focused on mitigating threats using adversary criteria derived from MITRE adversarial tactics, techniques, and common knowledge (ATT&CK). The approach here is tested by identifying vulnerabilities in software associated with six universities and four government facilities. Ranking policy performance is measured using the Normalized Discounted Cumulative Gain (nDCG). Results show an average 71.5%–91.3% improvement toward identifying vulnerabilities likely to be targeted and exploited by cyber threat actors. The return on investment (ROI) of patching using these policies results in a savings of 23.3%–25.5% in annualized costs. The results demonstrate the efficacy of creating knowledge graphs to link large datasets to facilitate semantic queries and create data-driven, flexible ranking policies.

INTRODUCTION

The relentless process of tracking and prioritizing vulnerabilities for patching is a top concern for cybersecurity professionals [1]. Ideally, every organization would apply the security updates for their operating systems and critical applications as soon as possible after updates are released. However, since patches from top vendors are delivered in monthly blocks on "Patch Tuesday," system administrators often find it difficult to select which patches to apply and identify which ones are not applicable [2–4]. Patch Tuesday is the term used to refer to the second Tuesday of each month when Microsoft, Adobe, Oracle, and others regularly release software patches for their software products [5]. Vulnerability prioritization is further hampered when companies delay the automatic installation of security updates in case the patch proves more troublesome than expected [6, 7].

Successful vulnerability management must balance two opposing goals: (1) coverage (fix everything that matters) and (2) efficiency (delay or deprioritize what does not matter) [8]. In industry, the most prevalent vulnerability management strategy identifies the base Common Vulnerability Scoring System (CVSS) scores for all identified vulnerabilities and patches them in descending score order (10 being the highest to 0 being the lowest) [9–11]. Unfortunately, research has shown that CVSS scores are not strongly linked to the emergence of new cyber exploits, and system administrators can be overwhelmed by the volume of vulnerabilities with nearly indistinguishable high scores [12]. While a CVSS score indicates vulnerability severity, it does not predict the exploit potential of the underlying software flaw or the operational impact to the organization.

Aggregating and synthesizing readily accessible, public data sources can provide an automated patch priority ranking by understanding what vulnerabilities and adversaries are relevant to an organization. The proposed relevance-based ranking model enables businesses to adopt a proactive strategy for vulnerability management [13]. Such an approach delivers the most efficient use of people, tools, time, and dollars to address cyber threats that pose the greatest operational risk. Just as search engines provide a better ranking of results based on personalization, so will the ranking of vulnerabilities. Within this context, an approach is sought to define cybersecurity vulnerability mitigation that improves upon rankings employing strategies based on the global CVSS metrics associated with known software vulnerabilities published in the National Vulnerability Database (NVD) [14].

The path to achieve this goal requires gathering, fusing, and analyzing relevant and available data discussed in this article. Specifically, it proceeds as follows. The "Data and Methods" section describes the aggregated public data sources, methods used to synthesize them, and the framework for ranking software vulnerabilities regarding different organizations for patching. The "Evaluation and Results" section evaluates the approach and presents the results. The "Discussions" section examines how the contributions are positioned in the software vulnerability management research landscape and identifies several limitations to the work. Ultimately, the study ends with the "Conclusions" section.

Data and Methods

The goal for this study is to remediate vulnerabilities in the most efficient way possible. This requires leveraging, associating, and analyzing different sources of cyber threat intelligence. The relationships among them need to be understood and organized into a structure for analysis that supports generating prioritized recommendations for effective vulnerability management.

These data sources are used to model software vulnerabilities regarding the skill level of cyber adversaries and their motivation to target a specific industry domain (e.g., national defense, higher education, finance, and health care). The relationships among these data sources and the software vulnerability's life cycle are summarized in Figure 1 to include the following data sets:

1. The Common Weakness Enumeration (CWE) captures data related to the discovery of a software weakness.
2. Data from the Common Vulnerabilities and Exposures (CVE) and CVSS prioritize a vulnerability's severity.
3. The Exploit Database (ExploitDB), Department of Homeland Security's Cybersecurity and Infrastructure Security Agency's Known Exploited Vulnerabilities (KEV) catalog, and Exploit Prediction Scoring System (EPSS) assess the likelihood of a software vulnerability being exploited in the wild.
4. The Common Attack Pattern Enumeration and Classification (CAPEC) and MITRE ATT&CK knowledge base provides data on how to remediate and mitigate published exploits.
5. The NVD catalogs and reports vendor-provided patches to vulnerabilities in commercial or open-source software.

These data sources and their specific leveraged attributes are described in more detail next. Highlighted are how they are synthesized together in a knowledge base to connect data about an adversary's capability to exploit a vulnerability to execute a cyberattack on an organization.

Data

Software Weaknesses Dataset

The Software Weaknesses dataset consists of data from the CWE, which provides a common language for describing security weaknesses in software architecture, design, or code. It is an encyclopedia of hundreds of types of software weaknesses, including buffer overflow, directory traversal, operating system injection,

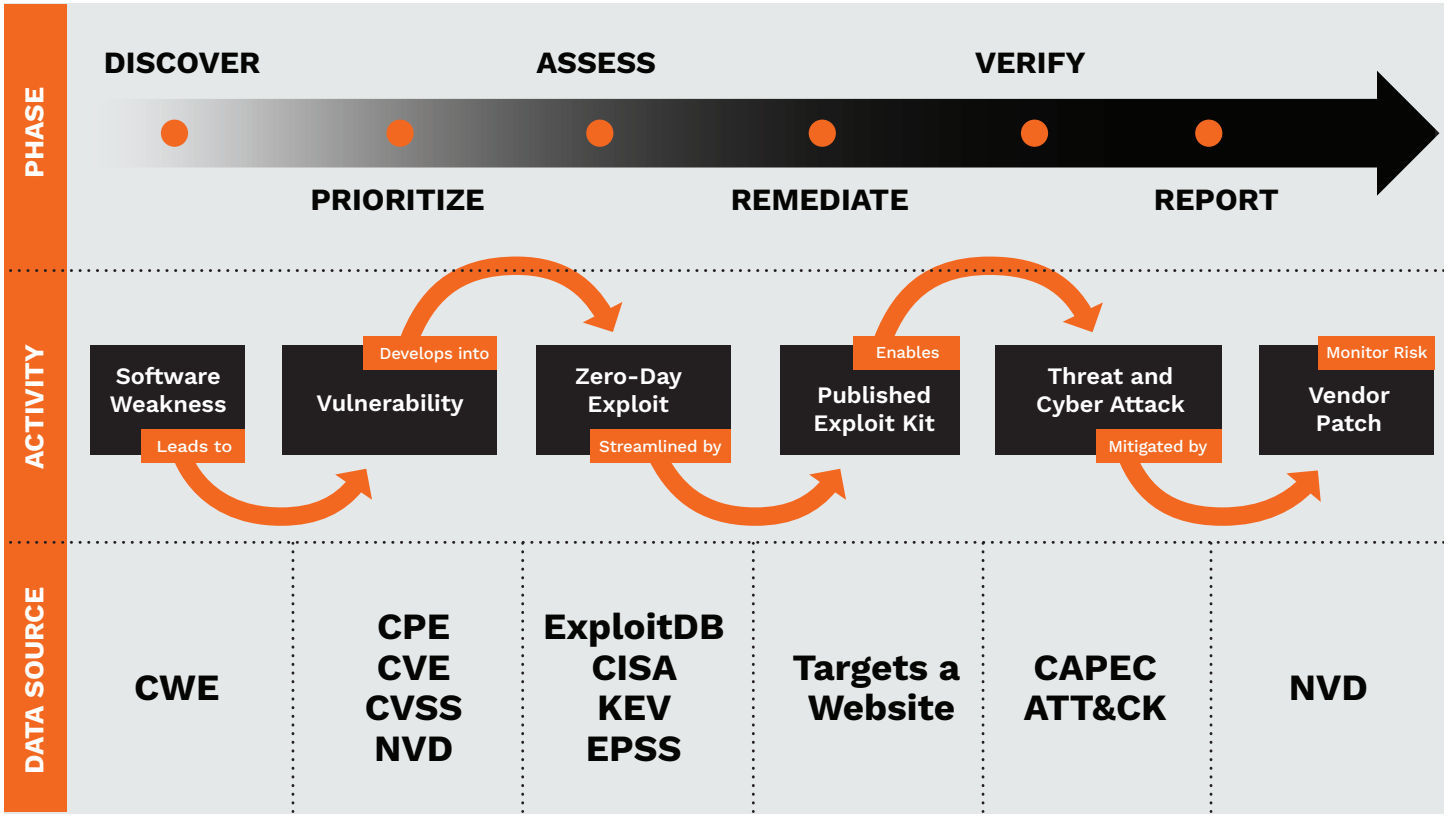

Figure 1. Software Vulnerability Life Cycle Phases and Their Relationships to Public Data Sources *(Source: McCoy [13]).*

race condition, cross-site scripting, hard-coded password, and insecure random numbers. Each software weakness has a technical impact, with eight that lead to failure: (1) read data, (2) modify data, (3–4) deny service - unreliable execution and resource consumption, (5) execute unauthorized code or commands, (6) gain privileges/ assume identity, (7) bypass protection mechanism, and (8) hide activities.

Vulnerability Dataset

The Vulnerability dataset is based on linking entries in the CVE with scoring information from the CVSS. The CVE is the authoritative source of publicly known vulnerabilities. The CVSS is an international standard for measuring the severity of a vulnerability. The CVSS base score is composed of metrics that reflect the intrinsic characteristics of the vulnerability. Each CVE entry includes a unique identifier (CVE number), a short free-text description, and a list of references for additional details of the vulnerability (in the form of URLs). This information is included in the dataset and linked with the CVSS base scores for the vulnerability.

Vendor Product Dataset

The Vendor Product dataset is based on the Common Platform Enumeration (CPE). Each entry (i.e., CPE-ID) defines a specific hardware device, operating system, or application software. Entries marked as deprecated are excluded, and the CPE-IDs of interest restricted to those are written in U.S. English. This dataset contains more than 15,000 CPE entries representing more than 3,000 products from ~200 vendors.

Attack Pattern Dataset

The Attack Patterns dataset includes 545 unique instances of CAPEC identifiers. CAPEC is a comprehensive dictionary and classification taxonomy of known attacks that can be used by analysts, developers, testers, and educators to advance community understanding and enhance defense sponsored by the U.S. Department

of Homeland Security. A CAPEC identifier can be linked to the MITRE ATT&CK enterprise tactics, techniques, and subtechniques. ATT&CK provides a common taxonomy for both offense and defense and has become a standard across many cybersecurity disciplines to convey threat intelligence, perform testing through red teaming or adversary emulation, and improve network and system defenses against intrusions.

ExploitDB Database

ExploitDB is based on a one-to-many mapping between an identified exploit kit (ExploitDB) to the vulnerabilities that are the target of that exploit (CVE). It is updated daily and provided by MITRE. It is augmented with the data from the Cybersecurity and Infrastructure Security Agency's (CISA's) KEV and the EPSS. The CISA KEV provides real-time updates via email alerts when a newly identified CVE-ID is exploited. The EPSS model is based on observations of exploitation attempts against vulnerabilities and analysis of ancillary information about each of those vulnerabilities and then uses historical events to make predictions about future ones. The EPSS score associated with a CVE-ID represents the probability [0–1] of exploitation in the wild in the next 30 days (following score publication) and the percentile of the current score compared to all scored vulnerabilities with the same or lower EPSS score.

> ***ATT&CK provides a common taxonomy for both offense and defense and has become a standard across many cybersecurity disciplines.***

Adversary Tactics and Techniques Dataset

The combination of MITRE ATT&CK and CAPEC datasets forms the adversary Tactics and Techniques dataset. The MITRE ATT&CK matrices are focused on network defense and describe the operational phases in an adversary's life cycle. The matrices also detail the specific tactics, techniques, and procedures that advanced persistent threat (APT) groups use to execute their objectives while targeting, compromising, and operating inside a network. Attack patterns enumerated by CAPEC are employed by adversaries through specific techniques described by MITRE ATT&CK. The dataset is formed by linking the CAPEC attack patterns and related MITRE ATT&CK techniques together, enabling contextual understanding of the attack patterns within an adversary's operational life cycle.

Synthesizing Data Sources Into a Knowledge Graph

The datasets described in the "Data" subsection can be combined to form a knowledge graph. The purpose of this graph is to support queries to effectively rank vulnerabilities for mitigation. This organizational structure is needed as a wealth of information about what vulnerabilities are targeted, who exploits those vulnerabilities, and how they currently exist. However, this information is not organized into a structure that comprehensively defines the relationships among the datasets. The knowledge graph and its schema described in Figure 2 and Table 1 address this deficiency.

Leveraging the Knowledge Graph to Link Vulnerabilities to Sector-Specific Threat Actors

The knowledge graph enables linking the vulnerabilities to APTs that target sectors within and outside the United States. It describes what data processing is required to populate the knowledge graph and how it links the data together once populated.

Defining a Standard Set of Sectors

The critical infrastructure (CI) sectors denoted by the Department of Homeland Security (DHS) reflect assets, systems, and networks that are vital enough to the United States that when incapacitating or destroying them, it would have a debilitating effect on national security, economics, public health, or public safety [15]. Sectors can also be divided into subsectors [16]. The CI sectors and subsectors are

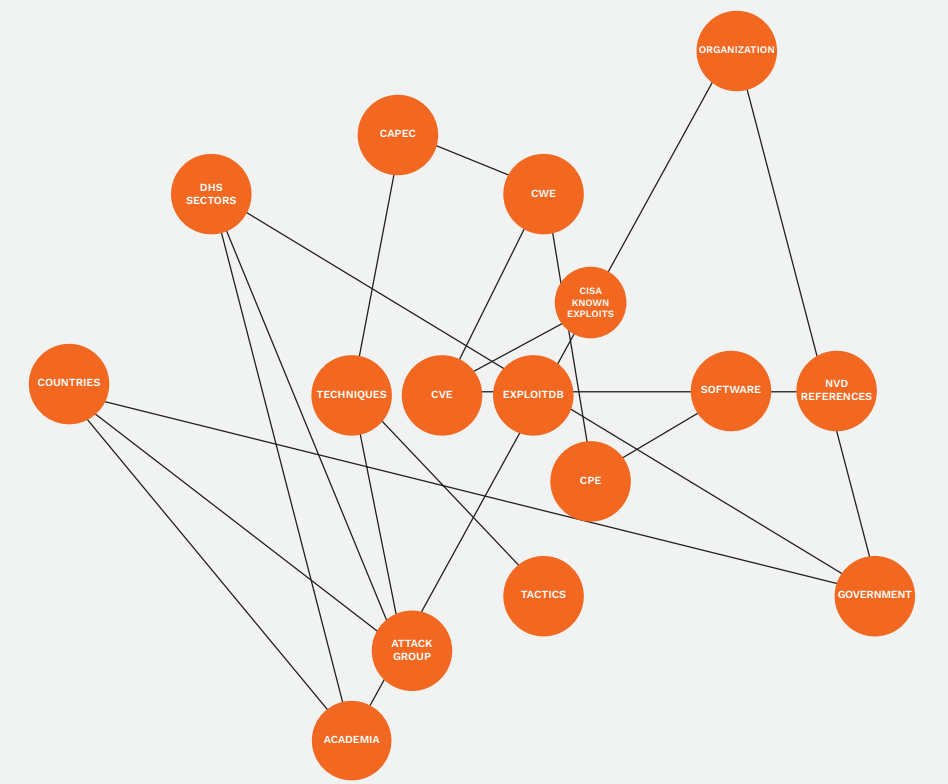

Figure 2. Graph Schema Representing the Entities of the Knowledge Graph and the Relationships Between Them *(Source: McCoy [13])*.

used to provide an affiliation for both threat actors and the organizations they target in the knowledge graph.

Defining Standard Locations

The knowledge graph requires a standard nomenclature to determine the country or region of origin for cyber threat actors and country of residence for organizations they target for attack. To meet this requirement, the U.S. State Department's list of independent states is leveraged. In this list, the term "independent state" refers to people politically organized into a sovereign state, with a definite territory recognized as independent by the United States.

Table 1. Legend for Node Labels and Relationships in Knowledge Graph Schema

LABEL	RELATIONSHIP	LABEL
NVD CVE	Reference exploit	ExploitDB
NVD CVE	Exploits known	CISA Exploit Catalog
NVD CVE	Weakened by	CWE
CWE	Known attack	CAPEC
CAPEC	Employs	Attack enterprise techniques
Attack groups	Achieves goal	Attack enterprise techniques
Attack groups	Originates	Countries
Attack groups	Targets	Countries
Attack groups	Focuses on	DHS sectors
Attack enterprise techniques	Achieved through	Attack enterprise techniques
NVD CVE	Affects	CPE
DHS sectors	Affiliated with	Organizations
Organizations	Operates in	Countries
Organizations	Installs	Software
Software	Has version	CPE
NVD CVE	Informs	NVD references

Assigning Attributes to Adversary Groups

APTs are an extension of nation-states' military forces because of the potential damages and chaos caused by successful critical infrastructure cyberattacks. MITRE keeps track of the APTs. Currently, it lists 129 threat groups [17] in their Enterprise Framework that can be associated with known techniques. Using their defined threat profiles, adversaries or threat groups employing the same tactics and techniques are identified.

Where Attacks Originate

For each APT group description provided by MITRE, natural language processing is used to extract keywords to determine the country or independent state from which the group operates. For example, a North Korean state-sponsored threat group would be assigned to North Korea with the mapping. The descriptions were also mined to determine year of origin (e.g., 2008) to ascertain each group's potential longevity. If a year was not explicitly stated in the description, the creation date of the MITRE description (e.g., has been active since at least 2009) was used.

Who Attacks Each Sector

Adversarial groups relevant to organizations based on who they target for attacks were identified next by mapping APTs and their country to DHS critical infrastructure sectors. To accomplish this, the subject of the term "targets," "targeted," or "targeting" was extracted in the group description from MITRE. The knowledge graph includes those where the United States is a targeted country, thus focusing on those attacks. The attribution of APTs to sectors is shown in Table 2. Note that some groups target more than one sector.

Relevance-Based Ranking Model

The goal here is to define an approach to cybersecurity vulnerability mitigation that improves upon rankings that employ strategies based on the global CVSS metrics associated with known software vulnerabilities published in the NVD. The outcome is a relevance-based ranking model that can be employed before an adversary takes advantage of a particular vulnerability. The model requires the following components:

> ***For each APT group description provided by MITRE, natural language processing is used to extract keywords to determine the country or independent state from which the group operates.***

Table 2. DHS Sectors Ranked by the Number of Attack Groups Targeting Those Sectors Based on Mentions in MITRE ATT&CK

SECTOR	GROUPS TARGETING
Government facilities	50
Information technology	33
Financial services	19
Healthcare and public health	17
Defense industrial base	14
Energy	14
Critical manufacturing	10
Communications	9
Transportation systems	7
Chemical	2
Water and wastewater systems	1
Nuclear reactors, materials, and waste	1
Emergency services	0
Dams	0
Commercial facilities	0
Food and agriculture	0

- Profiles that describe the organization under evaluation in terms of the DHS sector and country in which they operate.
- Collection and normalization of a complete software inventory for each organization.
- Threat-centric ranking policy definitions based on attack groups of interest and their skill levels.
- Scoring method for each ranking policy.

Creating Organizational Profiles

A vulnerability ranking policy needs to consider the installed software for the organization under evaluation. A representative set of organizations is identified and defined in government and education facilities to serve as organizational benchmarks for evaluating the vulnerability management approach.

Software Used in the Education Subsector

CollegeSimply [18] provides a list of Virginia colleges and sources public domain college data from the U.S. Department of Education National Center for Education Statistics. Using the list, six universities of varying sizes and funding sources (public and private) were chosen. The public universities were the University of Virginia (UVA), Virginia Tech (VT), Old Dominion University (ODU), and William & Mary University (W&M). The private universities were Regent University (REGENT) and Washington and Lee University (WLU). For each university, a published list of supported academic software was located on the university's website, and CPE-IDs were assigned to each piece of software. The full academic software listing is provided in McCoy [13]. A summary of the number of vulnerabilities found in the academic software associated with each university is shown in Table 3. A "size designation" (small [S], medium [M], large [L], and extra-large [XL]) was assigned based on the number of software products publicly listed. However, this did not reflect the size of the university or the number of software products used by the university.

Software Used by Government Facilities

Government facilities do not routinely publish the software they use. However, the "National Security Telecommunications and Information Systems Security Policy (NSTISSP) No. 11" requires government agencies to purchase only commercial security products that have met specified third-party assurance requirements and have been tested by an accredited national laboratory [19]. The list of certified products is available at https://www.commoncriteriaportal.org/products/. In accordance with NSTISSP, the "Common Criteria" is an internationally recognized set of guidelines (International Organization for Standardization [ISO] 15408) that defines a common framework for evaluating security features and capabilities of information technology (IT) security products against functional and assurance requirements [20].

The Common Criteria was reduced to the set of products certified for use in the United States. CPE-IDs across all categories were then searched based on the vendor and product name. The software list shown in Table 4 consists of applications and operating systems. It was generated by randomly selecting software from the Common Criteria with assigned CPE-IDs in groups of 14, 24, 30, and 47 to approximately

Table 3. Academic Software Associated With Vendor Product CPE-ID

UNIT OF MEASURE	W&M	ODU	VT	REGENT	UVA	WLU
CPEs assigned	24	47	12	23	30	13
Software listed	33	69	22	31	49	23
Size designation	M	XL	S	M	L	S

Table 4. Government (GOV) Facility Software Associated With Vendor Product CPE-ID

NUMBER OF SOFTWARE PRODUCTS	GOV-S	GOV-M	GOV-L	GOV-XL
Software assigned	14	24	30	47
Common criteria	57	57	57	57

match the cardinality of the S, M, L, and XL university software lists.

Ranking Policy Definitions

Deciding which vulnerabilities to remediate is a daunting task. In a perfect world, all vulnerabilities would be remediated as they were discovered; unfortunately, this does not happen. An exploit observed in the wild is the most relevant proxy for the probability that an exposed vulnerability can be used to compromise an organization's network. To that end, the predictive ranking policies evaluated identify candidate vulnerabilities that fit the pattern of known attack groups. Formally, this is the intersection of vulnerabilities in the software used by an organization and vulnerabilities being actively targeted by threat actors.

> ***An exploit observed in the wild is the most relevant proxy for the probability that an exposed vulnerability can be used to compromise an organization's network.***

The criteria for the ranking policies using the attacker characteristics and targets is discussed in the "Synthesizing Data Sources Into a Knowledge Graph" subsection. Each policy leverages data points in the knowledge graph to provide a scoring methodology that considers the following:

- Which threat actors use the same technique to initiate an attack?
- Given an industry, which threat actors target it?
- Given a type of attack, which vulnerabilities does it exploit?
- At present day, what is the probability of exploit?
- Given an organization, which vulnerabilities are present in the installed software?

Four different ranking policies were created to answer these questions. Each policy prioritizes different information based on organizational information preferences regarding specific threats. The policies also include knowledge on whether an exploit for the CVE-ID has been observed.

- **Policy 1: CVSS Base Score Ranking –** Vulnerabilities are remediated based on the assigned CVSS base score ranking from most severe ("critical") to least severe ("low").
- **Policy 2: APT Threat Ranking –** Vulnerabilities are remediated based on the likelihood of present-day exploit and the existence of a technique employed by an attack group that targets the industry in the country where the organization operates.
- **Policy 3: Generalized Threat Ranking –** Vulnerabilities are remediated based on the likelihood of exploit by a low-skilled or highly-skilled adversary that has high impact on the organization.
- **Policy 4: Ideal Ranking –** The ideal ranking employs the same criteria as the APT and generalized threat rankings, Policies 2 and 3, but has the foreknowledge that a vulnerability has already been exploited using information from the ExploitDB and CISA KEV databases.

Ranking Policy Implementations

For each CVE-ID, 16 features using the cyberintelligence data sources are examined. The features, which inform each policy and create a set

of relevance scores for ranking CVE-IDs as they are published, are as follows: (1) CVE-ID, (2) CVSS base score metrics, (3) publication date, (4) modification date, (5) CAPEC-ID, (6) CAPEC skill level, (7) ATT&CK technique name, (8) MITRE ATT&CK group ID, (9) MITRE ATT&CK group country of operation, (10) risk appetite, (11) EPSS probability of exploit, (12) CISA known exploit catalog, (13) ExploitDB, (14) organization identifier, (15) critical infrastructure sector, and (16) organization's country of residence. The source code used in implementing the ranking policies is available in McCoy [21].

Based on the policy definitions, the CVSS V3.1 base score is the only feature needed to implement Policy 1. The features needed to implement Policy 2 and its ideal version in Policy 4 are listed in Table 5.

For Policies 2–4, a binary weighting [0,1] is used for each feature to determine its existence as applicable to a specific CVE-ID. The sum of the categorical values is presented as the relevance score to rank the associated CVE-IDs using the logic shown in the algorithm provided in McCoy [13]. The minimum assigned relevance score is set to 1 using this algorithm to avoid a long tail of nonrelevant CVE-IDs and ensure only relevant CVE-IDs associated with the organization's installed software are candidates for ranking.

When determining what to patch, the setup and business disruption costs must be considered and weighed against the potential exploitation cost and when and how often to patch an enterprise system or application decided. The total costs of a vulnerability are the sum of its direct costs (level of effort employed by human resources) and indirect costs (productivity losses and interruption of production processes after patching). Previous research has established that these costs can be measured in nonmonetary units based on the severity of the vulnerability where low = 0.25, medium = 1, high = 1.5, and critical = 3 units [22]. The economic cost of remediating vulnerabilities is evaluated using these established units.

EVALUATION AND RESULTS

Candidate Generation

In this study, 55,939 CVE-IDs published between 2019 and 2021 were used as the corpus from which to identify a much smaller subset of candidate vulnerabilities for ranking. The CVE modification date was used to simulate examining the vulnerabilities as they were published. A total of 3,079 unique CVE-IDs applied across all the government facilities and education subsector software lists. The data and source code used in this evaluation are available in McCoy [21].

Table 5. Policies 2 and 4 Scoring Features Using MITRE ATT&CK Data Feed to Characterize the Threat to the Organization

FEATURE	SPECIFIC THREAT RELEVANCE RANK	IDEAL RANK VALUE
CVSS base metric (attack vector)	Network	Network
DHS sector	Government facilities education	Government facilities education
Organization's country	United States	United States
Attack group's country	China, Russia, Iran	China, Russia, Iran
Risk appetite	[0, 100]	[0, 100]
EPSS probability	0.876	NA
CISA KEV or ExploitDB entry exists	NA	True
Software affected	True	True
Scoring range	[1–6]	[1–6]

For the government facilities shown in Table 6, low annual vulnerability counts for three of the four proxy organizations were less than 2% of all CVE-IDs analyzed. Even the largest government organization, GOV-XL, which was designed to mirror the breadth of software (i.e., 47 products) of its counterpart ODU in the education subsector, experienced less than 4% of all CVE-IDs analyzed. The low number of vulnerabilities in the sector may be attributed to the selection process for software products assigned to government facilities in this study, which were selected exclusively from the certified product list approved by the Common Criteria [19]. This outcome may provide an indication that the rigor imposed upon these products in terms of security requirements and ongoing evaluation may potentially reduce their exposure to published vulnerabilities.

For the education subsector shown in Table 7 vulnerability counts of less than 2% were observed for organizations with small amounts of reported software, such as VT and WLU. Conversely, it was noted that universities who reported more software in use such as ODU, REGENT, and WM need to evaluate hundreds of vulnerabilities as candidates for remediation during any given year.

Figures 3 and 4 show the accumulated vulnerabilities by month and year for each organization in this study. It is important to note the unpredictable way newly published and modified CVE-IDs can present themselves for analysis and remediation to an organization. Similarly, Tables 8 and 9 show the vulnerabilities for the government and education subsectors. Note that WM, ODU, and REGENT experienced a steady stream of vulnerabilities across all three years of this study. They also experienced an increase in the number of weeks per year during which a continuous remediation policy would be advantageous. For ODU, note an increase from 42 weeks per year in 2019 to 50 weeks per year in 2021.

Normalized, Discounted Cumulative Gain

Within the field of cybersecurity,

Table 6. Total Vulnerabilities by Year for Government (GOV) Facilities Sector

YEAR	GOV-S	GOV-M	GOV-L	GOV-XL
2019	8	41	51	102
2020	11	34	55	144
2021	16	84	140	285
Total vulnerabilities	35	159	246	531
Percentage of all vulnerabilities	0.25%	1.15%	1.77%	3.85%

Table 7. Total Vulnerabilities by Year for Education Subsector

YEAR	VT	WLU	REGENT	WM	UVA	ODU
2019	14	3	1,396	1,370	188	1,457
2020	6	57	565	556	279	751
2021	15	144	1,721	1,704	639	2,026
Total vulnerabilities	35	204	3,682	3,630	1,106	4,234
Percentage of all vulnerabilities	0.25%	1.45%	26.56%	26.19%	7.98%	30.54%

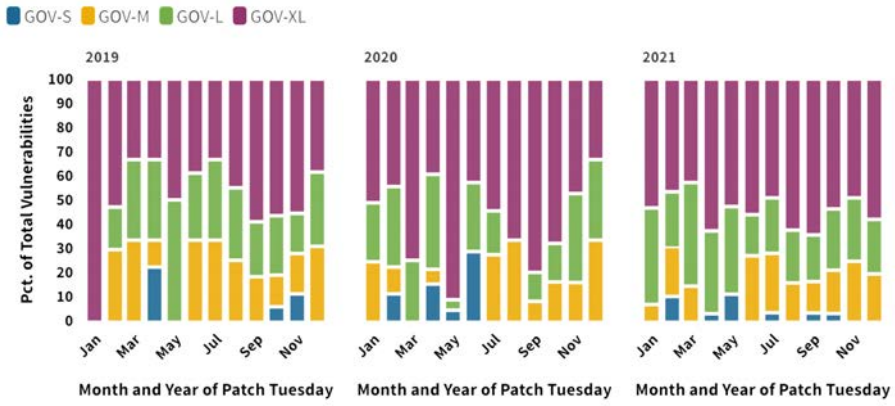

Figure 3. Vulnerabilities by Month and Year for CVE-IDs Between 2019 and 2021 for the Government Facilities Sector *(Source: McCoy [13])*.

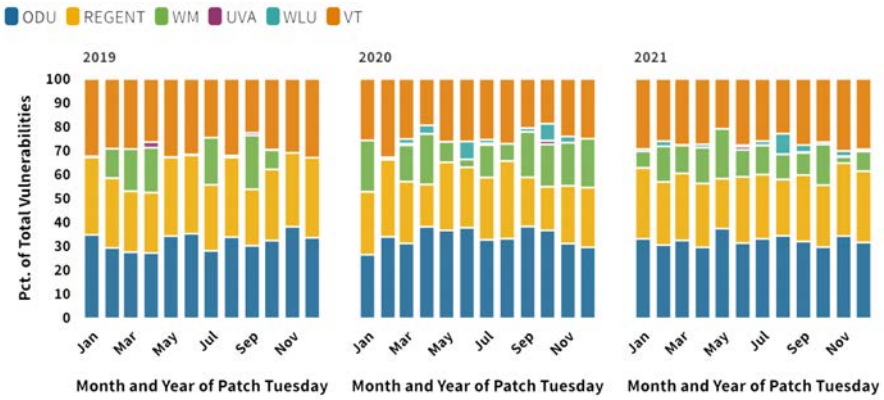

Figure 4. Vulnerabilities by Month and Year for CVE-IDs Between 2019 and 2021 for the Education Sector *(Source: McCoy [13])*.

Table 8. Weekly Vulnerability Traffic by Year for the Government (GOV) Facilities Subsector

AVERAGE VULNERABILITY			MINIMUM VULNERABILITY	MAXIMUM VULNERABILITY	WEEKS
Year	Organization	Per Week	Per Week	Per Week	Per Year
2019	GOV-S	4	1	20	32
2019	GOV-M	3	1	9	24
2019	GOV-L	2	1	4	23
2019	GOV-XL	2	1	2	5
2020	GOV-S	4	1	13	40
2020	GOV-M	3	1	10	23
2020	GOV-L	3	1	10	16
2020	GOV-XL	2	1	3	6
2021	GOV-S	7	1	25	43
2021	GOV-M	4	1	14	40
2021	GOV-L	3	1	11	29
2021	GOV-XL	2	1	3	10

> ***Within the field of cybersecurity, there is no consensus approach for measuring, testing, and comparing the accuracy of a ranking model.***

there is no consensus approach for measuring, testing, and comparing the accuracy of a ranking model. Therefore, this research, like others discussed in the "Discussion" section, builds upon measurements derived from the information retrieval (IR) field. Evaluation measures for IR assess how well the search results from a recommender satisfy a given query. Specifically, recommender systems use the nDCG [23] score to evaluate the ranking of items (e.g., individual vulnerabilities) in a collection (e.g., NVD).

The nDCG varies from 0.0 to 1.0, with 1.0 representing the ideal ranking order. The nDCG is commonly used to evaluate search engine result pages (SERPs), where the position of an entry indicates its search result relevance. Higher ranked pages are more likely to gain the consumer's attention. The same approach is applied toward creating a ranking list for patching vulnerabilities. Order is important to ensure higher ranked CVE-IDs are considered first. The main difficulty encountered when using nDCG is

Table 9. Weekly Vulnerability Traffic by Year for the Education Subsector

AVERAGE VULNERABILITY			MINIMUM VULNERABILITY	MAXIMUM VULNERABILITY	WEEKS
Year	Organization	Per Week	Per Week	Per Week	Per Year
2019	WM	2	1	4	7
2019	ODU	1	1	1	3
2019	REGENT	35	1	442	40
2019	UVA	43	1	441	32
2019	VT	11	1	44	18
2019	WLU	35	1	444	42
2020	ODU	1	1	1	6
2020	REGENT	5	1	20	12
2020	WM	15	1	57	40
2020	UVA	15	1	58	39
2020	WLU	9	1	34	33
2020	VT	18	1	59	43
2021	ODU	3	1	4	7
2021	REGENT	7	1	23	21
2021	WM	36	1	264	48
2021	UVA	36	1	258	48
2021	WLU	15	1	120	43
2021	VT	41	1	315	50

the availability of an ideal ordering of results when feedback (e.g., human judgment) is unavailable. This shortcoming was faced by SERPS with Policy 4, introduced in the "Ranking Policy Implementations" subsection as a data-driven proxy of an ideal ordering of vulnerabilities.

To compare the results of rankings between each relevance policy and the ideal ranking (Policy 4), the nDCG of each CVE-ID for every organizational interaction was calculated with the ranking system. The nDCG values were averaged for each weekly collection of CVE-IDs to obtain a measure of the average performance of the ranking algorithms. The application of nDCG in this study is interpreted as follows:

1. "G" is for gain – it corresponds to the magnitude of each vulnerability's relevance.
2. "C" is for cumulative – it refers to the cumulative gain, or summed total, of every vulnerability's relevance score.
3. "D" is for discounted – it divides each vulnerability's scored relevance by the scored relevance of the associated ideal policy to reflect the goal of having the most relevant vulnerabilities ranked toward the top of the mitigation lists.
4. "n" is for normalized – it divides discounted cumulative gain (DCG) scores by ideal DCG scores calculated for a ground truth data set, as represented by the relevance scores and ranking resulting from the ideal policy (i.e., Policy 4), which used foreknowledge of exploited vulnerabilities contained within historical ExploitDB and CISA KEV intrusion detection reports.

Once the relevance value is computed for each CVE-ID, each entry is ranked based on the relevance value and the nDCG is computed using the following formulas:

$$DCG_k = \sum_{i=1}^{k} \frac{2^{rel_i - 1}}{\log_2(i+1)} . \quad (1)$$

The cumulative gain at K is the sum of gains of the first K items recommended. *iDCGk* is the maximum possible (ideal) *DCG* for a given set of queries, vulnerabilities, and relevance scores.

$$nDCG_k = \frac{DCG_k}{iDCG_k} . \quad (2)$$

The chart in Figure 5 illustrates the average values of nDCG for each position K based on weekly vulnerability collections. K reflects the number of CVE-IDs to remediate. The number of observations ranges from 383 when K = 1 to 16 when K = 100.

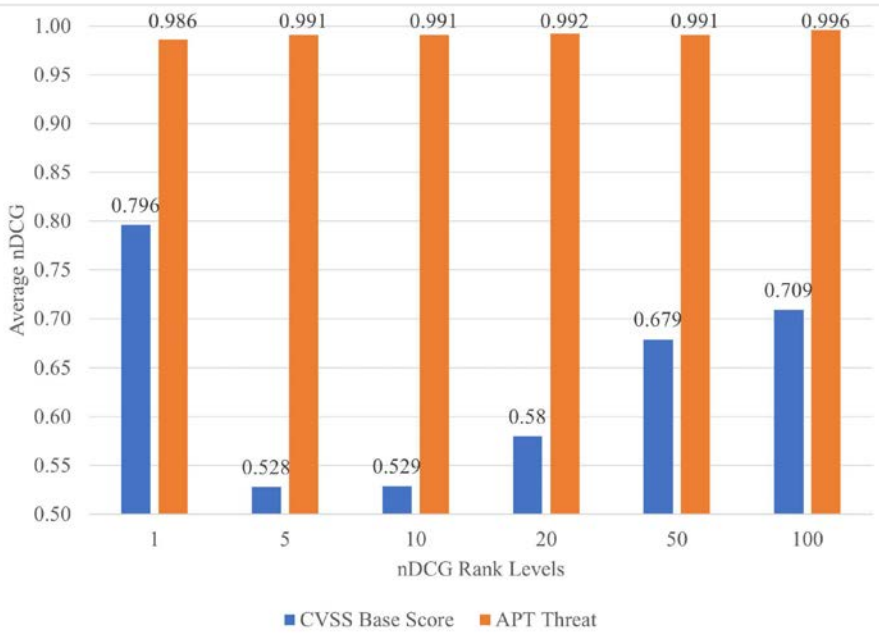

Figure 5. Average Value of nDCG at Different Rank Levels (K) for CVSS Base Score vs. APT Threat Policy for the ODU, REGENT, and WM Organizations *(Source: McCoy [13])*.

The x-axis reports the rank (from 1 to 100), while the y-axis displays the respective value of nDCG@K. Figure 5 shows that the CVSS base score performs moderately well at the ends of the spectrum when K = 1 and K = 100. However, the performance decreases when $5 \leq K \leq 50$. Policy 2 is not impacted by the number of weekly CVE-IDs; it performs at a consistent level regardless of the number of CVE-IDs encountered.

Testing and Evaluating the Policies

Within the evaluation, the number of CVE-IDs to be evaluated each week can vary for each organization. Therefore, to calculate nDCG, the cumulative gain needs to be normalized at each ranking position for a chosen number of vulnerabilities. Tables 8 and 9 show that the average weekly vulnerability traffic across all organizations establishes a natural threshold of 20 CVE-IDs during a given week as the minimum number needed to apply a relevance ranking policy.

The GOV-XL, ODU, REGENT, UVA, WLU, and WM organizations consistently met this threshold. However, GOV-XL, UVA, and WLU were excluded from further examination in this section, as there were numerous weeks where no published CVE-IDs applied to the organization's installed software.

For the remaining organizations with more than 50 weekly observations (ODU, REGENT, and WM), the necessary features were collected using the cyberintelligence data sources identified in the "Data" subsection to compute a relevance score, rank the CVE-IDs, and calculate nDCG using Policy 4 as the ideal ranking. Only the CVSS V3.1 base score was needed to evaluate Policy 1. For all ranking policies, the set of applicable CVE-IDs was ranked in descending order by relevance score and then subsequently ordered by CVE-ID to avoid ties. The performance of Policy 1 was evaluated against the threat-centric policies (Policies 2 and 3). Finally, the patch cost (in nonmonetary units) for the top 20 CVE-IDs was determined, where low = 0.25, medium = 1, high = 1.50, and critical = 3.00 [22].

Measuring Ranking Quality

For the threat-centric policies (Policies 2 and 3), the average performance was measured across all three years of the evaluation period using nDCG@20. China was chosen as the APT group of interest for vulnerabilities impacting ODU, REGENT, and WM since it contained the most frequent origin of APT threats against the United States [24].

The nDCG is measured on a scale of 0.0 to 1.0, and a score of 1.0 indicates the ideal ranking order has been achieved. The goal is to obtain an nDCG score close to 1.0 for each threat policy. Table 10 shows the average nDCG@20 for each organization. The average nDCG@20 of 0.99 indicates Policy 2 performs better than Policy 1. The average

Table 10. Average Performance of Policy 1 vs. Policy 2, Where China Is the Source Region of Interest (nDCG@20)

SCHOOL	YEAR	CVSS BASE SCORE	APT THREAT CHINA	AVG. DIFF. IN nDCG	KNOWN EXPLOITS
ODU	2019	0.601	0.996	0.394	4
ODU	2020	0.557	0.998	0.441	2
ODU	2021	0.571	0.986	0.415	12
REGENT	2019	0.592	0.999	0.407	2
REGENT	2020	0.557	0.998	0.441	1
REGENT	2021	0.585	0.985	0.399	12
WM	2019	0.598	0.998	0.400	3
WM	2020	0.565	0.998	0.433	1
WM	2021	0.585	0.985	0.399	12

difference in nDCG@20 of 0.41 indicates that Policy 2 performs 71.5% better than Policy 1 as an indicator of vulnerabilities that might be targeted by an APT group.

In the results, the nDCG@20 measures for Policy 1 were in the range of [0.343, 1], as shown in Figure 6. Lower values for nDCG@20 were observed with Policy 1 when the number of vulnerabilities collected exceeded the minimum threshold (i.e., 20) by more than 1,000% (e.g., 200+). Higher nDCG@20 values were observed when the number of vulnerabilities were closer to the threshold (e.g., 20 to 30). Policy 2 was minimally impacted by the number of vulnerabilities and was in the range of [0.878, 1].

Table 11 shows similar results for Policy 3. The average difference in nDCG@20 of 0.35 indicates Policy 3 performs 91.3% better than Policy 1 as an indicator of vulnerabilities that might be targeted by a highly skilled cyber threat actor. This is highlighted as well in Figure 7.

Using all the weekly observations (n = 163) across organizations, a paired t-test was performed to compare the mean of the nDCG for Policy 1 against Policy 2 [25]. Results of this test indicated that there was a significantly large difference between Policy 1 [mean = 0.58, STDEV = 0.1] and Policy 2 [mean = 0.992, STDEV = 0.02], and the p-value equaled 0. The Policy 2 population's nDCG@20 average was greater than the Policy 1 population's average, and the difference

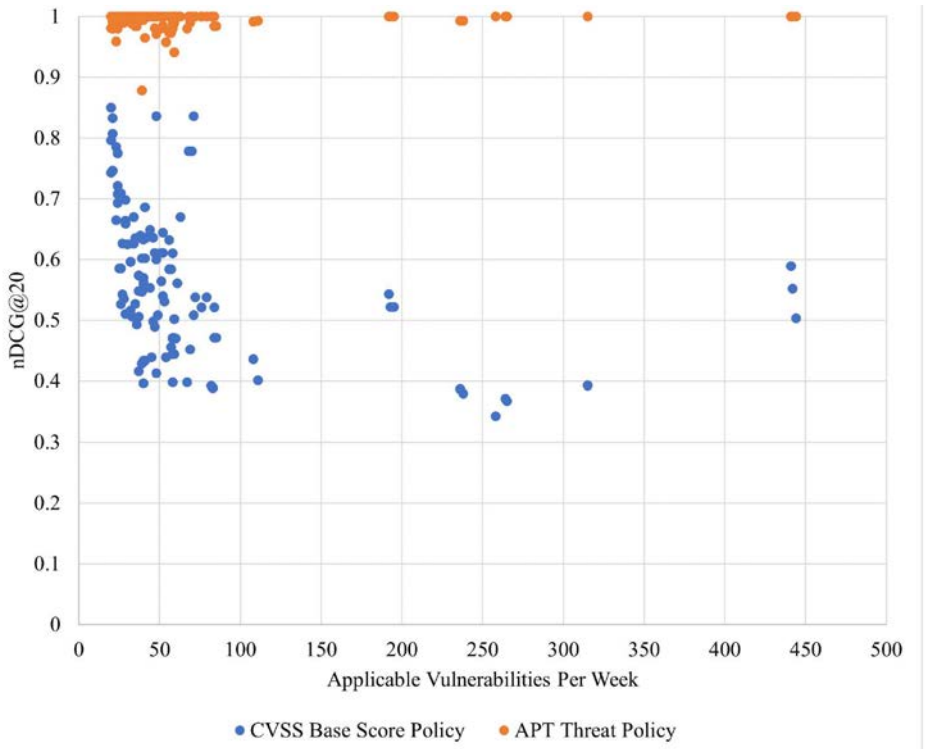

Figure 6. nDCG@20 for Policy 1 vs. Policy 2 for the ODU, REGENT, and WM Organizations *(Source: McCoy [13])*.

Table 11. Average Performance of Policy 1 vs. Policy 3, With a Highly Skilled Adversary (nDCG@20)

SCHOOL	YEAR	CVSS BASE SCORE	GENERAL THREAT HIGHLY SKILLED	AVG. DIFF. IN nDCG	KNOWN EXPLOITS
ODU	2019	0.543	0.988	0.444	4
ODU	2020	0.548	0.998	0.450	2
ODU	2021	0.474	0.986	0.511	12
REGENT	2019	0.528	0.995	0.467	2
REGENT	2020	0.512	0.999	0.487	1
REGENT	2021	0.500	0.984	0.484	12
WM	2019	0.538	0.992	0.454	3
WM	2020	0.520	0.999	0.478	1
WM	2021	0.499	0.984	0.485	12

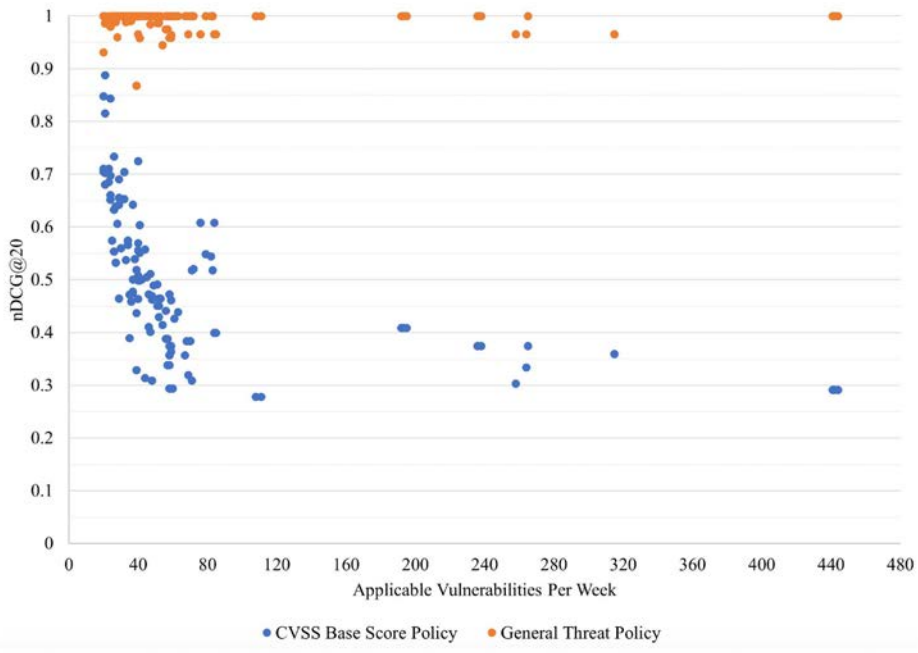

Figure 7. nDCG@20 for the Policy 1 vs. Policy 3 for the ODU, REGENT, and WM Organizations *(Source: McCoy [13])*.

was large enough to be statistically significant.

A similar test to compare the mean of the nDCG for Policy 1 against Policy 3 was performed. Results of the paired t-test indicated that there was a significantly large difference between Policy 1 [mean = 0.512, STDEV = 0.139] and Policy 3 [mean = 0.99, STDEV = 0.022], and the p-value equaled 0. The Policy 3 population's average nDCG@20 was greater than the Policy 1 population's average, and the difference was large enough to be statistically significant.

These results showed that CVSS base score metrics did not contain a data element or scoring component that allowed enumeration of a specific threat. The paired t-test indicated that the difference in the recommended ranking positions of CVE-IDs between policies was statistically significant (p-value equaled 0). Therefore, any relevance ranking based solely on the CVSS base score would fall short of the organization's specified goals. These results also provided another indication that the severity of a vulnerability, as measured by its CVSS base score, might not be the optimal ranking approach for every organization.

Cost of Patch Prioritization

Past research has shown that organizations cannot fix all their known vulnerabilities. Instead, they can fix 5%–20% of known vulnerabilities per month [26]. Here, the annualized cost of remediating the top 20 vulnerabilities produced by the different ranking Policies 1–3 was examined. Defined by Fruhwirth et al., the nonmonetary units were used

with patching [22]. The results of this analysis are shown in Table 12. In all cases, there is a decreased average cost of 23.3% when Policy 2 is used for prioritizing CVE-IDs for remediation. Specifically, Policy 2 realizes decreases 498 units for ODU, 390.5 units for REGENT, and 455.75 units for WM over the three-year evaluation period when compared to Policy 1.

Table 13 shows increased savings in patch costs using Policy 3. The cost of patching remains the same across all organizations using the CVSS base score. However, for each organization, there are additional savings over using Policy 2. Decreases of 548.25 units for ODU, 500.75 units for REGENT, and 499.75 for WM represent an average 25.6% improvement over the CVSS base score approach. Policy 2 only provided a 23.3% improvement.

Using all the weekly observations (n = 163) across organizations, a paired t-test was performed to compare the mean of the patch costs for Policy 1 against Policy 2 [25]. Results of this test indicated that there was a significantly large difference between Policy 1 [mean = 37.025, STDEV = 10.291] and Policy 2 [mean = 28.362, STDEV = 5.475], and the p-value = 7.45e-27. The population of Policy 2's average patch cost was less than Policy 1's, and the difference was large enough to be statistically significant.

Similarly, a paired t-test to compare the mean of the patch costs for Policy 1 against Policy 3 was performed [25]. Results of this test indicated that there was a significantly large difference between Policy 1 [mean = 37.025,

Table 12. Difference in the Cost of Patching the Top 20 CVE-IDs for Policy 1 vs. Policy 2, Where China Is the Source Region of Interest

SCHOOL	YEAR	CVSS BASE SCORE	APT THREAT CHINA	AVERAGE SAVINGS
ODU	2019	631.50	449.25	185.25
ODU	2020	531.00	439.00	92.00
ODU	2021	994.50	770.00	244.50
REGENT	2019	604.75	422.25	182.50
REGENT	2020	375.50	308.50	67.00
REGENT	2021	960.00	752.00	208.00
WM	2019	603.75	424.75	179.00
WM	2020	374.00	308.50	65.50
WM	2021	960.00	748.75	211.25

Table 13. Difference in the Cost of Patching the Top 20 CVE-IDs for Policy 1 vs. Policy 3 From a Highly Skilled Adversary

SCHOOL	YEAR	CVSS BASE SCORE	GENERAL THREAT COST	AVERAGE SAVINGS
ODU	2019	631.50	438.50	193.00
ODU	2020	531.00	424.50	106.50
ODU	2021	994.50	745.75	248.75
REGENT	2019	604.75	412.00	192.75
REGENT	2020	375.50	294.50	81.00
REGENT	2021	960.00	733.00	227.00
WM	2019	603.75	412.50	191.25
WM	2020	374.00	296.50	77.50
WM	2021	960.00	729.00	231.00

STDEV = 10.291] and Policy 3 [mean = 27.523, STDEV = 4.905], and the p-value = 9.989e-30. The population of Policy 3's average patch cost was less than Policy 1's, and the difference was large enough to be statistically significant.

Predicting Exploits

Only a small subset (2%–7%) of published vulnerabilities are exploited in the wild [26]. Given that such a small number of CVE-IDs are exploited, it is advantageous for organizations to leverage as much insight as possible to identify potential threats. How Policy 2 can be used to prioritize a vulnerability with a known exploit is demonstrated here. The ODU organization identified 39 CVE-IDs to mitigate during the week of 23 November 2021.

In this case study, the top 20 are ranked according to Policy 2, as shown in Table 14. Note that three CVE-IDs in this group, CVE-2021-38000, CVE-2021-30632, and CVE-2021-30633, have known exploits. The CISA known exploits entry for CVE-2021-38000, which impacts Google Chrome, is shown in Figure 8. The entries in Table 14 show that all three CVE-IDs are identified as relevant using Policy 2. However, CVE-2021-38000 is ranked at position 29 using Policy 1 based on its CVSS base score of 6.1 (medium severity). This highlights that when using Policy 1, CVE-2021-38000 falls outside the top-20 range for remediation by IT administrators at ODU. In contrast, Policy 2 elevates this CVE-ID to position no. 3 because of its high relevance score.

Table 14. Application of Ranking Policies by ODU for Vulnerabilities Published During the Week of 23 November 2021 (Known Exploits Are Bolded and Highlighted in Grey)

CVE-ID	CVSS BASE SCORE	RELEVANCE SCORE	POLICY 1 RANK	POLICY 2 RANK	EXPLOIT
CVE-2021-37966	4.3	6	34	1	—
CVE-2021-37999	6.1	6	28	2	—
CVE-2021-38000	**6.1**	**6**	**29**	**3**	**Yes**
CVE-2021-30542	8.8	2	5	4	—
CVE-2021-30543	8.8	2	6	5	—
CVE-2021-30626	8.8	2	7	6	—
CVE-2021-30627	8.8	2	8	7	—
CVE-2021-30628	8.8	2	9	8	—
CVE-2021-30629	8.8	2	10	9	—
CVE-2021-30630	4.3	2	31	10	—
CVE-2021-30632	**8.8**	**2**	**11**	**11**	**Yes**
CVE-2021-30633	**9.6**	**2**	**2**	**12**	**Yes**
CVE-2021-34423	9.8	2	1	13	—
CVE-2021-34424	7.5	2	26	14	—
CVE-2021-37956	8.8	2	12	15	—
CVE-2021-37957	8.8	2	13	16	—
CVE-2021-37958	5.4	2	30	17	—
CVE-2021-37959	8.8	2	14	18	—
CVE-2021-37961	8.8	2	15	19	—
CVE-2021-37962	8.8	2	16	20	—

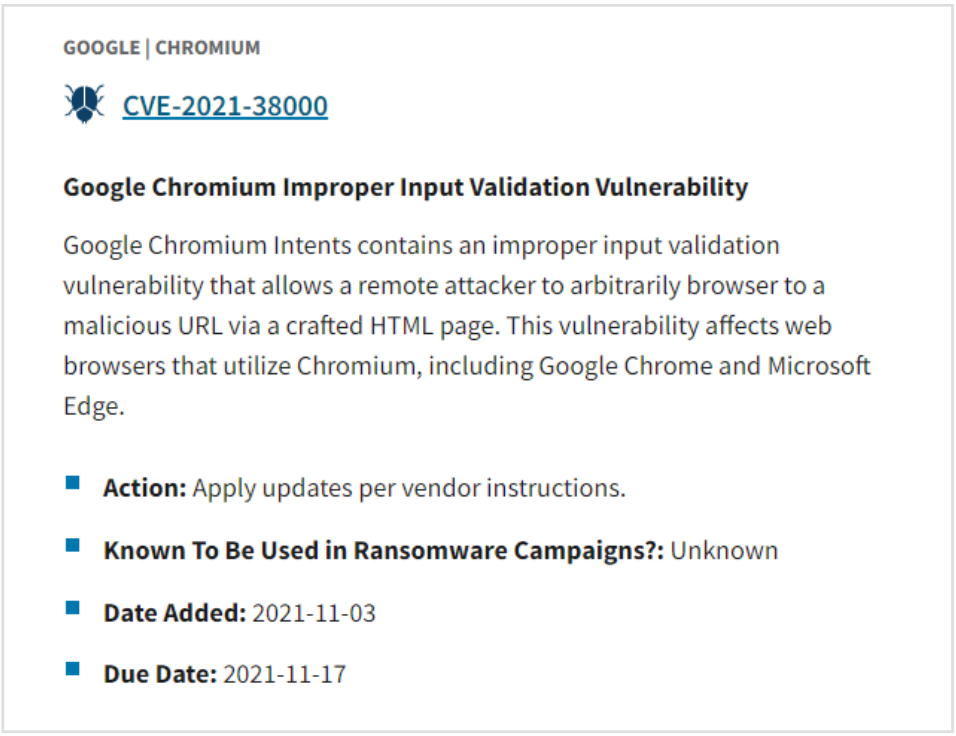

Figure 8. ACISA Known Exploits Catalog Entry for CVE-2021-38000 *(Source: CISA [27])*.

DISCUSSION

There is a myriad of existing research that falls within the scope of this work. Related research is discussed, limitations of the work identified, and provided contributions highlighted.

Related Research

Multiple researchers have created ontologies to represent the cybersecurity domain by aggregating multiple sources of information [28–33]. This work provides the foundation for building automated tools, which reduce the scope, complexity, and volume of security data that must be managed by security professionals leveraged in this approach. However, this research differs from these efforts in that more information and sources are extracted to achieve completeness in the knowledge graph. In addition, categorization is a necessary precursor to the ranking policies for vulnerability management. Multiple research efforts have shown that identifying and categorizing additional metadata about vulnerabilities, exploits, attacks, and targets can be beneficial [22, 34–36]. More recently, applying text mining to extra additional data about these entities has led to models which predict the severity of a vulnerability using only text-based data [37–40].

Even with an organized understanding of the cyber threat domain, understanding how to minimize the cost of managing and protecting information assets is a challenge. A core component of this challenge is adopting a vulnerability management process that can detect and remediate known vulnerabilities [12]. A common approach is to remediate all vulnerabilities above a certain severity score. However, this approach has been found to be suboptimal [41] and, in some cases, no better than randomly choosing vulnerabilities to remediate [42]. Furthermore, in many cases, it is infeasible to patch all the CVEs with the highest CVSS base scores due to the time and resources required for remediation actions. This is because 13.5% of the NVD vulnerabilities are scored between 9 and 10 [43].

This has led to extensive work in evaluating if the CVSS score can be a good predictor for vulnerability exploitation [44] and whether it can be improved by additional information [45–47]. Machine-learning approaches have been explored [48, 37] as well as exploit prediction models that leverage data from online sources generated by the white-hat community (i.e., ethical hackers) [39]. Vulnerability exploitation can also be modeled as a transition between system states [49–55]. However, these graphs often tend to be unwieldy as network size grows, making the identification of realistic paths to compromise difficult to achieve [56]. Customized and target specific ranking approaches also exist [43, 12, 57, 42]. However, these approaches assume the existence of site-specific threat intelligence information.

Contributions of the Approach

Prior research has demonstrated the ability to examine adversary capabilities and vulnerability management and exploit prediction at a particular point in time or with isolated threat scenarios. However, little research has been done to create an end-to-end prioritization approach that encompasses the entire vulnerability management life cycle. This gap is addressed by the following:

- Extracting dozens of essential features about the vulnerability, including its potential for harm, the degree to which it is exploitable, and how frequently the vulnerability is targeted by adversaries.
- Leveraging the ability of property graphs to offer a flexible schema where attributes can be added to extend the data model, creating hierarchies with different levels of granularity, and combining multiple dimensions to better manage big data.
- Performing an assessment of current and predicted future attacker activity based on known tactics and techniques.
- Correlating threat and exploit intelligence from publicly available authoritative sources.
- Devising an approach to convert raw data about threat indicators into contextual risk scores.
- Identifying how important the affected asset is to an organization in any industry.
- Inferring indirect facts and hidden relationships, which can further inform the results.

Parsing real-time, open-source cyber threat intelligence data cannot be accomplished by a human analyst. Therefore, its correlation and analysis are automated using a knowledge graph. Application programming interfaces (APIs) and data feeds maintained by the National Institute of Standards & Technology (NIST) can also be leveraged to provide awareness of the changing threat landscape while allowing dynamic and continuous assessment of the underlying network architecture. This research provides benefits to organizations seeking to create high-level strategies to examine cybersecurity posture in a manner that is predictive and not just reactive.

Known Limitations

This work is not without limitations. To apply the approach here, organizations must have a methodology to accurately construct a software inventory that can be correlated with an entry in the CPE database. Vulnerabilities cannot be allocated without a CPE-ID, and low fidelity inventory reporting may result in residual cyber risk. The relevance ranking policies identified can only be effectively applied to a known software architecture. Furthermore, it is important to note that the attack group list in MITRE ATT&CK is not all encompassing. A Google search will identify emerging APT groups that are not included in the MITRE's enterprise matrices. In addition, the proof-of-concept code entries collected via ExploitDB do not include a time component indicating when the POC entry was made. As a result, it is not possible to discretely link the CVE-ID's publication or modification date with the subsequent appearance of an intrusion report. The inclusion of a timestamp would have allowed evaluating the predictive portion of the policies based on a timeline of events. The approach here is naive regarding exploitation and does not consider the publication date for exploit code maturity using ExploitDB. The ExploitDB to CVE mapping webpage is also not well covered in the internet archives.

Time lapse dynamics related to data sources also exist. The EPSS probability scores and percentiles are dynamic and should be collected near the time of the CVE publication date. To maintain consistency in the

> ***Vulnerabilities cannot be allocated without a CPE-ID, and low fidelity inventory reporting may result in residual cyber risk.***

dataset, all cyberintelligence data was collected and frozen for analysis as of 31 December 2021. Future work can utilize the API provided by the EPSS team to dynamically collect the scores and percentiles in real-time. This is a candidate for future work. Finally, the prescribed optimum ordering approach may not ease patch hesitancy or prevent a culture of "wait and see" regarding patching vulnerabilities. The policies also cannot control the quality of vendor patch distributions on Patch Tuesday that, in some cases, can lead to recalls later in the month. These scenarios are outside the scope of this research. However, the ranking policies here can reduce the amount of unnecessary work spent patching CVE-IDs that are neither applicable nor associated with a known cyber threat actor.

CONCLUSIONS

The process of tracking and remediating vulnerabilities is relentless. The key challenge is trying to identify a remediation scheme specific to in-house, organizational objectives. Without a strategy, the result is a patchwork of fixes applied to a tide of vulnerabilities, any one of which could be the point of failure in an otherwise formidable defense. The goal of this research is to demonstrate that aggregating and synthesizing readily accessible, public data sources to provide personalized, automated recommendations for organizations to prioritize their vulnerability management strategy will offer significant improvements over the current state-of-the art solutions. Results showed an average 71.5%–91.3% improvement toward identifying vulnerabilities likely to be targeted and exploited by cyber threat actors. The ROI of patching using the policies results in a savings in the 23.3%–25.5% range for annualized costs. A paired t-test demonstrates these findings are statistically significant and offer an improvement over the industry standard approach to vulnerability management.

Overall, the relevance ranking strategy described in this study emphasizes the capability of threat-centric scenarios for ranking and prioritizing vulnerabilities with due consideration to the threat environment. A network defender, who typically must address thousands of exposed vulnerabilities, can spend fewer resources to patch more vulnerabilities that are much more likely to be exploited and of interest to a specific set of cyber threat actors. The automated data aggregation within the knowledge graph allows the user to submit queries to identify new vulnerabilities that affect the most important software and servers. This ability to differentiate among vulnerabilities and how they might be targeted by an adversary enhances the state of the art in vulnerability management.

NOTE

This work was unfunded and performed as part of Corren McCoy's Ph.D. work at ODU in Norfolk, VA. ■

REFERENCES

[1] Sheyner, O., J. Haines, S. Jha, R. Lippmann, and J. M. Wing (editors). "Automated Generation and Analysis of Attack Graphs." Proceedings of the 2002 IEEE Symposium on Security and Privacy, 2002.

[2] Krebs, C. C. "Emergency Directive 20-03 (ED 20-03), Mitigate Windows DNS Server Vulnerability From July 2020 Patch Tuesday." https://www.cisa.gov/news-events/directives/ed-20-03-mitigate-windows-dns-server-remote-code-execution-vulnerability-july-2020-patch-tuesday, 2020.

[3] Coble, S. "CISA Issues Emergency Vulnerability Warning." https://www.infosecurity-magazine.com/news/cisa-issues-emergency/, 2020.

[4] Dulaunoy, A. "CVE-Search Vulnerability Lookup." Social media post, *Mastodon*, https://infosec.exchange/@adulau/112013472555767369, 2024.

[5] Wikipedia. "Patch Tuesday." https://en.wikipedia.org/wiki/Patch_Tuesday, 2020.

[6] Schmeidler, N. "Microsoft Patch Tuesday: All or Nothing Patches." https://blog.morphisec.com/microsoft-patch-tuesday-all-or-nothing-patching, 2016.

[7] O'Donnell, L. "Microsoft Pulls Bad Windows Update After Patch Tuesday Headaches." https://threatpost.com/microsoft-windows-update-patch-tuesday/163981/, 2021.

[8] Foreman, P. "Vulnerability Management: Taylor and Francis Group," 2010.

[9] Forum of Incident Response and Security Teams (FIRST). "Exploit Prediction Scoring System v2022.01.01." https://www.first.org/epss/, 2022.

[10] Romanosky, S., and J. Jacobs. "Probability, Percentiles, and Binning – How to Understand and Interpret EPSS Scores." FIRST, https://www.first.org/epss/articles/prob_percentile_bins, 2022.

[11] Mell, P., K. Scarfone, and S. Romanosky (editors). "A Complete Guide to the Common Vulnerability Scoring System Version 2.0." FIRST-Forum of Incident Response and Security Teams, 2007.

[12] Allodi, L., and F. Massacci. "Comparing Vulnerability Severity and Exploits Using Case-Control Studies." *ACM Transactions on Information and System Security* (*TISSEC*), vol. 17, no. 1, pp. 1–20, 2014.

[13] McCoy, C. G. "A Relevance Model for Threat-Centric Ranking of Cybersecurity Vulnerabilities." Ph.D. dissertation, Old Dominion University, Norfolk, VA, 2022.

[14] NIST. "National Vulnerability Database. https://nvd.nist.gov/, 16 November 2022.

[15] CISA. "Critical Infrastructure Sectors." https://www.cisa.gov/topics/critical-infrastructure-security-and-resilience/critical-infrastructure-sectors, 2020.

[16] Commission USEA. "CI Scoop: What Are Sectors and Sub-Sectors?" https://web.archive.org/web/20181105225720/https://www.eac.gov/ci-scoop-what-are-sectors-and-sub-sectors/, 2017.

[17] Infoblox. "An Introduction to MITRE ATT&CK." https://blogs.infoblox.com/security/an-introduction-to-mitre-attck/, 2019.

[18] CollegeSimply. "Virginia Colleges Ranked by Largest Enrollment." https://www.collegesimply.com/colleges/rank/colleges/largest-enrollment/state/virginia/, 2021.

[19] CNSS. "National Security Telecommunications and Information Systems Security Policy (NSTISSP) No. 11." https://www.hsdl.org/?view&did=487791, 2003.

[20] ISO. "Information Technology — Security Techniques — Evaluation Criteria for IT Security — Part 1: Introduction and General Model." https://www.iso.org/standard/50341.html, 2022.

[21] McCoy, C. "oduwsdl/CyberThreatRelevanceRank: CyberThreatRelevanceRank." https://github.com/oduwsdl/CyberThreatRelevanceRank, 2024.

[22] Fruhwirth, C., and T. Mannisto (editors). "Improving CVSS-Based Vulnerability Prioritization and Response With Context Information." Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, 2009.

[23] Järvelin, K., and J. Kekäläinen. "Cumulated Gain-Based Evaluation of IR Techniques." *ACM Transactions on Information Systems* (*TOIS*), vol. 20, no. 4, pp. 422–446, 2002.

[24] Utterback, K. "An Analysis of the Cyber Threat Actors Targeting the United States and Its Allies." Utica College, 2021.

[25] Wikipedia. "Student's T-Test." https://en.wikipedia.org/wiki/Student%27s_t-test, 2022.

[26] Jacobs, J., S. Romanosky, B. Edwards, M. Roytman, and I. Adjerid. "Exploit Prediction Scoring System (EPSS)." arXiv preprint arXiv:190804856, 2019.

[27] CISA. "Known Exploited Vulnerabilities Catalog." https://www.cisa.gov/known-exploited-vulnerabilities-catalog, 16 November 2022.

[28] Hemberg, E., J. Kelly, M. Shlapentokh-Rothman, B. Reinstadler, K. Xu, N. Rutar, et al. "Linking Threat Tactics, Techniques, and Patterns with Defensive Weaknesses, Vulnerabilities and Affected Platform Configurations for Cyber Hunting." arXiv preprint arXiv:201000533, 2020.

[29] Bridges, R. A., C. L. Jones, M. D. Iannacone, K. M. Testa, and J. R. Goodall. "Automatic Labeling for Entity Extraction in Cyber Security." arXiv preprint arXiv:13084941, 2013.

[30] Jones, C. L., R. A. Bridges, K. M. T. Huffer, and J. R. Goodall (editors). "Towards a Relation Extraction Framework for Cyber-Security Concepts." Proceedings of the 10th Annual Cyber and Information Security Research Conference, 2015.

[31] Iannacone, M., S. Bohn, G. Nakamura, J. Gerth, K. Huffer, R. Bridges, et al. (editors). "Developing an Ontology for Cyber Security Knowledge Graphs." Proceedings of the 10th Annual Cyber and Information Security Research Conference, 2015.

[32] Bizer, C., T. Heath, and T. Berners-Lee. "Linked Data: The Story So Far." *Semantic Services, Interoperability and Web Applications: Emerging Concepts: IGI Global*, pp. 205–227, 2011.

[33] Wang, J. A., and M. Guo (editors). "OVM: An Ontology for Vulnerability Management." Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies, 2009.

[34] Frei, S., M. May, U. Fiedler, and B. Plattner (editors). "Large-Scale Vulnerability Analysis." Proceedings of the 2006 SIGCOMM Workshop on Large-Scale Attack Defense, 2006.

[35] Alberts, C. J., and A. J. Dorofee. *Managing Information Security Risks: The OCTAVE Approach*." Addison-Wesley Professional, 2003.

[36] Tsukerman, E. "Cybersecurity Threat Modeling With OCTAVE," September 2020.

[37] Jacobs, J., S. Romanosky, I. Adjerid, and W. Baker. "Improving Vulnerability Remediation Through Better Exploit Prediction." *Journal of Cybersecurity*, vol. 6, no. 1, 2020.

[38] Chen, T., and C. Guestrin (editors). "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[39] Almukaynizi, M., E. Nunes, K. Dharaiya, M. Senguttuvan, J. Shakarian, and P. Shakarian (editors). "Proactive Identification of Exploits in the Wild Through Vulnerability Mentions Online." Proceedings of the 2017 International Conference on Cyber Conflict (CyCon US), 2017.

[40] Khazaei, A., M. Ghasemzadeh, and V. Derhami. "An Automatic Method for CVSS Score Prediction Using Vulnerabilities Description." *Journal of Intelligent & Fuzzy Systems*, vol. 30, no. 1, pp. 89–96, 2016.

[41] Dey, D., A. Lahiri, and G. Zhang. "Optimal Policies for Security Patch Management." *INFORMS Journal on Computing*, vol. 27, no. 3, pp. 462–77, 2015.

[42] Allodi, L., F. Massacci, and J. Williams. "The Work-Averse Cyberattacker Model: Theory and Evidence From Two Million Attack Signatures." *Risk Analysis*, vol. 42, no. 8, pp. 1623–1642, 2022.

[43] Alperin, K., A. Wollaber, D. Ross, P. Trepagnier, and L. Leonard (editors). "Risk Prioritization by Leveraging Latent Vulnerability Features in a Contested Environment." Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, 2019.

[44] Allodi, L., and F. Massacci (editors). "A Preliminary Analysis of Vulnerability Scores for Attacks in Wild: The EKITS and SYM Datasets." Proceedings of the 2012 ACM Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, 2012.

[45] Tatarinova, Y., and O. Sinelnikova. "Extended Vulnerability Feature Extraction Based on Public Resources." *Theoretical and Applied Cybersecurity*, vol. 1, no. 1, 2019.

[46] Notess, G. R. *The Wayback Machine: The Web's Archive*. Vol. 26, no. 2, pp. 59–61, 2002.

[47] Horawalavithana, S., A. Bhattacharjee, R. Liu, N. O. Choudhury, L. Hall, and A. Lamnitchi (editors). "Mentions of Security Vulnerabilities on Reddit, Twitter and GitHub." Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 2019.

[48] Jacobs, J., S. Romanosky, I. Adjerid, and W. Baker. "Improving Vulnerability Remediation Through Better Exploit Prediction." The 2019 Workshop on the Economics of Information Security, 2019.

[49] Singhal, A., and X. Ou. "Security Risk Analysis of Enterprise Networks Using Probabilistic Attack Graphs." *Network Security Metrics*, Springer, pp. 53–73, 2017.

[50] Ou, X., W. F. Boyer, and M. A. McQueen (editors). "A Scalable Approach to Attack Graph Generation." Proceedings of the 13th ACM Conference on Computer and Communications Security, 2006.

[51] Ou, X., S. Govindavajhala, and A. W. Appel (editors). "MulVal: A Logic-Based Network Security Analyzer." Proceedings of the USENIX Security Symposium, 2005.

[52] Homer, J., X. Ou, and D. Schmidt. "A Sound and Practical Approach to Quantifying Security Risk in Enterprise Networks." Kansas State University Technical Report, 2009-3, http://people.cs.ksu.edu/~xou/publications/tr_homer_0809.pdf, 2009.

[53] Gallon, L., and J. J. Bascou (editors). "Using CVSS in Attack Graphs." Proceedings of the 2011 Sixth International Conference on Availability, Reliability and Security, 2011.

[54] Noel, S., E. Harley, K. H. Tam, and G. Gyor. "Big-Data Architecture for Cyber Attack Graphs." MITRE Case Number 14-3549, https://csis.gmu.edu/noel/pubs/2015_IEEE_HST.pdf, 2014.

[55] Noel, S., E. Harley, K. H. Tam, and G. Gyor (editors). "Big-Data Architecture for Cyber Attack Graphs Representing Security Relationships in NoSQL Graph Databases." Proceedings of the IEEE Symposium on Technologies for Homeland Security (HST), 2015.

[56] Obes, J. L., C. Sarraute, and G. Richarte. "Attack Planning in the Real World." arXiv preprint arXiv:13064044, 2013.

[57] Allodi, L., and S. Etalle (editors). "Towards Realistic Threat Modeling: Attack Commodification, Irrelevant Vulnerabilities, and Unrealistic Assumptions." Proceedings of the 2017 Workshop on Automated Decision Making for Active Cyber Defense, 2017.

BIOGRAPHIES

CORREN MCCOY serves as adjunct faculty to several universities, where she instructs courses in the information systems, computer science, and cybersecurity curriculums. She has earned multiple industry certifications and has been a frequent presenter at technology conferences across the country. Dr. McCoy holds a bachelor's degree in computer science from Pennsylvania State University, a master's degree in computer science from ODU, an M.S. in management from Regent University, and a Ph.D. in computer science from ODU, where her research focused on data-driven approaches for assessing cyber vulnerabilities.

ROSS GORE is a research associate professor at the Virginia Modeling, Analysis and Simulation Center (VMASC) at ODU. His current work focuses on data science and predictive analytics. Dr. Gore holds a bachelor's degree in computer science from the University of Richmond and a master's degree and Ph.D. in computer science from the University of Virginia.

MICHAEL L. NELSON is the deputy director of the School of Data Science at ODU, where his research interests include web science, web archiving, digital libraries, and information retrieval. Prior to his current position and teaching at ODU, he was an electronics engineer at NASA's Langley Research Center. He also received a joint appointment with VMASC. Dr. Nelson holds a B.S. in computer science from Virginia Tech and M.S. and Ph.D. degrees in computer science from Old Dominion University.

MICHELE C. WEIGLE is a professor of computer science at ODU, where her research interests include web science, social media, web archiving, and information visualization. She currently serves on the editorial boards of the *Journal of the Association for Information Science and Technology* and the *International Journal on Digital Libraries*. She has published over 115 articles in peer-reviewed conferences and journals and served as PI or Co-PI on external research grants totaling $6M from a wide range of funders, including the National Science Foundation, the National Endowment for the Humanities, the Institute of Museum and Library Services, and the Andrew W. Mellon Foundation. Dr. Weigle holds a B.S. in computer science from Northeast Louisiana University (now University of Louisiana at Monroe) and an M.S. and Ph.D. in computer science from the University of North Carolina.