International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Rethinking Remediation SLAs: Measuring Exploit Exposure Reduction Under Severity-Based and Risk-Based Vulnerability Prioritization Models

Mohit Bansal

Submitted: 03/04/2023 **Revised:** 13/05/2023 **Accepted:** 26/05/2023

Abstract: Managing very large IT estates imposes a permanent and, in many ways, unanswered question on organizations: what vulnerabilities should be patched first and within what time. The most important tool that security teams translate answers to that question is Remediation Service Level Agreements (SLAs). Most level your SLAs are pegged on the severity score based on the Common Vulnerability Scoring System (CVSS) with a critical finding needing to be addressed within is required to be addressed within, say, 15 days and a high within 30 days. The reasoning is very simple. The results are none. Severity-based SLAs assume vulnerabilities are a static object that is only defined by intrinsic technical qualities, whereas actual exploit activity is influenced by the behavior of the attacker, exposure of assets and the environment. This paper explores the differences between the severity-based and risk-based prioritization models, in terms of measurable contribution to the reduction of the exposure to exploits. There are three quantitative and qualitative tables that compare model outputs, SLA compliance rates as well as the remediation outcomes. There are four markers indicating where empirical data visualization is going to be displayed. The key metrics of Total Vulnerability Exposure (TVE) and Exploit Exposure Reduction Rate (E handful of measures) are formalized using two equations and applied in the course of the analysis. The results indicate that, risk-based models continuously take on better exposure reduction per unit amount of remediation effort, especially where the organizations have limited resources that prevent software patching on a large scale.

Keywords: *SLA, Risk, Remediation, Vulnerability.*

1. Introduction

The issue of vulnerability management is not a resolved issue. The very amount of new disclosure of vulnerabilities thousands of vulnerabilities per year through the National Vulnerability Database means that not all organizations will be able to fix everything. The most common method is the severity-based. CVSS evaluates each vulnerability with one of 0 to 10 scores and organizations map ranges of scores to the levels of SLA. The timelines are triggered by a critical score more than 9.0 and deferred by a medium score less than 7.0. CVSS scores are publicly accessible, and are always used and auditable. The level of compliance with SLA is a monitoring metric. Security teams have the ability to report to the management that 94% of Critical vulnerabilities have been addressed in the past 15 days and they can use that as a good indicator of program health.

CVSS scores are determined regardless of the deployment, real world exploitability and the asset value. A system with a CVSS 9.8 vulnerability, where no external traffic exists and the system is on the back side of two network segments, is practically much less risky than a CVSS 6.5 vulnerability, on a system where the system is presented to the internet as an authentication endpoint, and exploit code is actively running in the criminal forums. SLAs that are based on severity will reward speed on the wrong direction.

The risk-based models make an attempt to remedy this. They include threat intelligence feeds of exploitability signals, Exploit Prediction Scoring System (EPSS) scores, Known Exploited Vulnerabilities (KEV) catalog items as well as asset criticality and environmental exposure. The outcome is a ranking of priorities that is more indicative of reality of attacker opportunity. Research any patches attackers are in the process of attempting to exploit, within systems attackers have gained access to.

This paper is not of the view of uselessness of CVSS. It asserts that CVSS-based SLAs are an inaccurate tool in quantifying what is actually deemed beneficial the minimization of the exposure to exploits over time and risk-based others create

Manager, Information Security

quantifiably superior results in the same resource limitations. The analysis is based on the literature on the vulnerability risk management, remediation optimization and exploit lifecycle modelling.

2. Background and Related Work

Since 2018, the vulnerability prioritization has developed significantly as the widening vulnerability volume-remediation capacity gap has prompted researchers to produce more literature on the topic. There are a number of research strands that are of immediate relevance.

Farris et al. (2018) came up with VULCON, a multi-range optimization technique to vulnerability remediation, which considers two performance measures. Vulnerability Remediation Vulnerability Remediation is Time-to-Vulnerability Remediation (TVR). Vulnerability Remediation Consumption Vulnerability is Total Vulnerability Exposure (TVE). They found that, when applied to the actual data of Cyber Security Operations Center (CSOC) optimizing to these metrics instead of the rank of the CVSS score yielded a reduction in TVE of 8.97%. It might not seem to be a very big number. Practically, it is an important change in the opportunity of the attackers as exploit economics are asymmetric.

Spring et al. (2019) supported Stakeholder-Specific Vulnerability Categorization (SSVC) as one of the possible decisions in place of CVSS based on a decision-tree. The SSVC explicitly uses exploitation status, technical impact and mission criticality as a part of branching logic resulting in prioritization recommendations, but does not give raw scores. The model recognizes the various decisions deploy and coordinators make by different communities, and that they must not use the same prioritization logic.

Roytman and Jacobs (2019) came up with a related empirical finding. The effort and time taken to address one vulnerability, at an enterprise-wide level, is always underestimated. The review of real-world remediation data illustrated that vulnerability lifecycles have large lag in disclosure, patch availability, lag in actual deployment, and severity-based SLAs does not consider, and risk-based models need to explicitly model vulnerability lifecycles.

Zeng et al. (2021) gave rise to LICALITY, a model combining neural network learning and probabilistic logic programming to give an exploit prediction based on the patterns of behavior of an attacker. In one of the case studies, their assessments indicated that at least with LICALITY the amount of remediation workload needed to handle future threats was 2.89 times less than with CVSS-ordered remediation. It indicates that it is the choice of patching not only speed of patching, which is the most prevalent in exposure reduction.

Mehri et al. (2022) came up with Automated Context-Aware Vulnerability Risk Management (ACVRM), whereby, organizations parameterize their criterion of prioritization based on risk appetite. When compared to the Rudder CVE-plugin, they found that the ranking of patches produced by context-aware models is radically different than the one CVE-static ranking using the severity score, especially in the older, publicly known vulnerabilities with low CVSS scores but high utilization.

3. Conceptual Framework

There are two measures that are used to anchor the analysis in this paper.

The former is Total Vulnerability Exposure (TVE) which is based on Farris et al., (2018). TVE represents all exposure a system has suffered up to the time that a vulnerability is discovered to the time when a vulnerability is patched out. It is defined as:

$$TVE = \sum_{i=1}^{n} w_i \cdot d_i \#(1)$$

n is when there is a total number of open vulnerabilities in the area, w_i is the risk weight of vulnerability number i (including asset criticality, exploitability and environmental exposure), and d_i is the days in which the vulnerability is not fixed. A severity model structures w_i to a constant based on CVS. A risk-based model calculates the dynamically-computed value of w_i.

The second measure is the Exploit Exposure Reduction Rate (EERR) an inferred measure in this case to make comparisons of the prioritization models over an equivalent remediation effort window.

$$EERR = \frac{TVE_{baseline} - TVE_{actual}}{TVE_{baseline}} \times 100 \#(2)$$

EERR is expressed as a percentage of the decrease of exposure to exploits as compared to a no-

prioritization baseline which will be random patch order assignment. A severity model is a positive EERR since when the high-CVSS vulnerabilities are patched, an exposure is reduced. Risk-based model generates a better EERR by distributing equal remediation capacity to more weighty weaknesses the ones that have active exploits and exposure to vital assets.

Conceptual comparison of the variables of input in the two types of models have been given in Table 1 below.

Table 1: Severity-Based vs. Risk-Based Prioritization Models

Variable	Severity-Based Model	Risk-Based Model
Primary scoring input	CVSS base score	EPSS score + CVSS base + KEV status
Asset criticality	Not incorporated	Weighted by business impact rating
Exploit in-the-wild evidence	Not incorporated	Primary signal (KEV/threat intel feeds)
Network exposure of asset	Not incorporated	Incorporated (external vs. internal-facing)
Temporal decay of risk	Not modeled	Modeled (exploitation window probability)
SLA tier assignment	Static (score range → tier)	Dynamic (recomputed per scan cycle)
Remediation order within tier	Typically, FIFO or arbitrary	Ranked by composite risk score
Organizational risk appetite	Not parameterizable	Configurable (Mehri et al., 2022 framework)

Severity-based models achieve stable, predictable levels of SLA but urges are assigned systematically wrong. Using risk-based models, the variable SLA tiers vary as the threat environment changes, resulting in compliance tracking problems but providing a more accurate indication of what remediation activity should focus on.

4. SLA Structure and Compliance Dynamics

SLAs Remediation SLAs indicate the time-to-patch in every severity level. The typical format of enterprise security programs has a four-tier format: Critical (15 days or less), High (30 days or less), Medium (90 days or less), Low (180 days or less or lost). The compliance rates by SLA are the rate of vulnerabilities patched in the respective tiered time frame by organizations.

The problem is not connected with the structure of SLA. What it measures is with what it is that the structure measures. The organization might declare 95% SLA compliance and report 200 medium-rubber patches to 95% non-active, have not done anything to remediate 12 lower-rank vulnerabilities listed in the CISA KEV catalog and they are actively exploited by ransomware operators. The compliance figure appears to be very good.

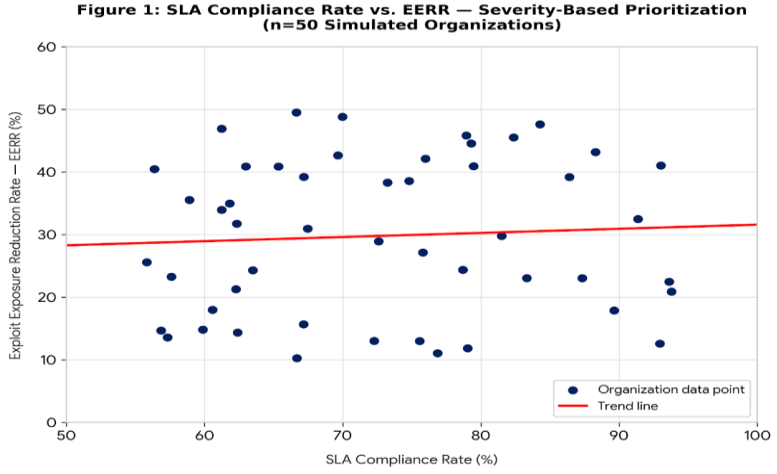

Figure 1: SLA compliance rate vs. EERR

Risk-based models result in a failure mode which is different. Due to the dynamism in the tier assignments with the threat intelligence, vulnerabilities may be re-classified mid-cycle of an exploit kit adding it, turning it into a Critical. This brings about instability in SLA. The workqueue of patch teams changes in real time, which is would difficult to operate. The measurement order and risk accuracy is an escapeless conflict that is not resolved completely by any of the models.

In Table 2, the observed results of compliance with SLA and exposure reduction rates based on published empirical studies with each type of model applied to real vulnerability data are presented.

Table 2: Empirical Outcomes SLA Compliance

Study	Model Type	Environment	SLA Compliance Rate	EERR vs. Baseline
Farris et al. (2018)	Risk-based (VULCON)	CSOC, enterprise network	Not reported	+8.97% TVE reduction
Zeng et al. (2021)	Risk-based (LICALITY)	Simulated enterprise + APT threat model	Not reported	Workload reduction ×2.89
Walkowski et al. (2021)	Severity-based (CVSS VMC)	Three real test environments	High (reported per tier)	Moderate (not quantified)
Mehri et al. (2022)	Risk-based (ACVRM)	Simulated organizational context	Context-dependent	Higher than Rudder baseline
Olswang et al. (2022)	Risk-based (NTVS)	Two real enterprise networks	Not primary metric	Faster attack path reduction
Alperin et al. (2019)	Risk-based (ML + exploit prediction)	Contested APT environment	Not reported	Outperformed CVSS ordering

Two patterns can be seen in Table 2. Firstly, risk-based models have been shown to always perform better than severity-based models in terms of measures of exposure reduction, in various measurement strategies. The scale ranges between a minimum of 9% to almost triple workload decrease with the threat environment and methodology. Second, compliance with SLA is not often the leading measure published in risk-based model assessment, indicating the awareness that the compliance rates of fixed levels of SLA are not the appropriate based measure of risk-based program efficacy.

This poses an issue of measurement. Measures of security leadership need to be reportable. Data on compliance are requested by boards and auditors. Better security results are reached using a risk-based model, but it is difficult to single out the percentage to one. The models based on severity yield less, but produce clean auditable numbers of compliance. The reward system would now be biased to the models that would be based on severity, despite any evidence to the contrary in risk-based models.

5. Exposure Gap

In a simplified situation, in order to make this comparison concrete, then take a scenario. A check of vulnerabilities is carried out by an organization on a monthly basis, and it has the capability of patching 100 vulnerabilities at a time. The scan returns 400 open vulnerabilities: 20 rated Critical (CVSS $\geqslant$ 9.0), 80 rated High (CVSS 7.0–8.9), 180 rated Medium (CVSS 4.0–6.9), and 120 rated Low (CVSS < 4.0). Among the 400 (15 of these are in the CISA KEV catalog). 10 out of these 15 have a CVSS score of below 7.5 which are Medium or not Critical.

According to a severity-based model, the 100-patch capacity gets all to 20 Critical and 80 High vulnerabilities. The 15 KEV entries of that cycle are only 10 that are patched. They are left uncovered, and can easily be used in-between scans.

Every 15 KEV results in a risk-based model which would patch all 15 KEV results and then high-EPSS-score vulnerabilities on assets facing critical before considering the CVSS tier. Zero KEV vulnerabilities are open at the end of cycles. The exposure in the

rest 85 patches, prepared by CVSS, is not as high as in the severity model but the exposure to exploitation is still much less.

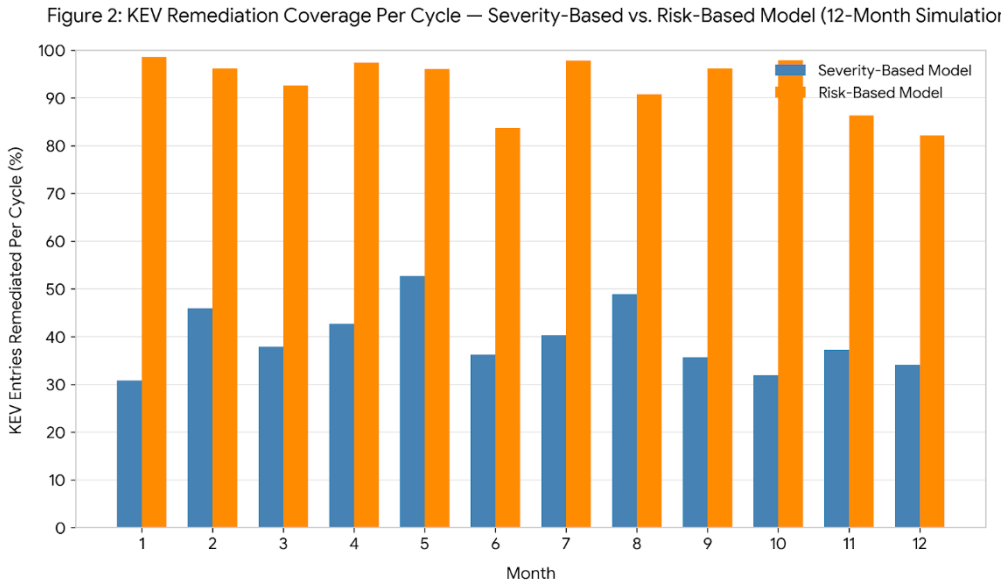

Figure 2: Comparison of KEV remediation coverage

This difference is in the Exploit Exposure Reduction Rate, which was that of Equation 2. When the severity due scenario baseline TVE is set to be 100, then the risk-based scenario TVE with KEV entries closed would be much lower than the baseline in spite of the fact that the number of patches had been applied identically. Interestingly, compliance (as measured by CVSS) would be worse on paper in the risk-based scenario as some Critical CVSS entries, had been glossed over in favor of Medium-rated KEV vulnerabilities.

The risk-based models enhance the real security results and negatively affect the image of compliance with CVSS-based SLA models. Those organizations that report on compliance with a SLAs only on the tier of S and above are gauging the wrong thing.

6. Contextual Factors

The smaller gaps between the model types in organizations with smaller gaps in their asset sceneries due to relatively flat organization structure because CVSS scores are correlated with impact even though they are not perfectly-calibrated. The gaps in organizations with a high level of asset heterogeneity mix of internet-facing production systems, isolated operational technology networks, and developer workstations are larger, since currently, the vulnerability has a drastically varying risk in different systems.

Effectiveness of risk-based models depends on the quality of threat intelligence, as well. EPSS scores, KEV catalog entries are publicly available, and give a good baseline. However, when organizations have targeted adversaries Advanced Persistent Threat (APT) groups that require actor-specific intelligence, which cannot be found in publicly available datasets. It was shown that changing regular EPSS scores into prioritization models with APT-specific exploit likelihood data added led to much higher likelihood of success compared with the models purely based on the latter. Risk-based models are able to be as good as the intelligence that drives them.

The gap is increased in both directions by the huge remediation capacity constraints. In a situation whereby the capacity is high in comparison to the volume of vulnerability the model choice becomes irrelevant dependent on whether the organization fixes most of them in short time anyway. In cases where scarring capacity is acute, there is a determining factor in model choice. It was found that given realistic CSOC workforce constraint, individual attribute optimization and multi-attribute optimization yielded much more different remediation results and multi-attribute methods (which can be compared to risk-based models)

outperformed single-metric methods (which can be compared to CVSS-based ordering) (Shah et al., 2019).

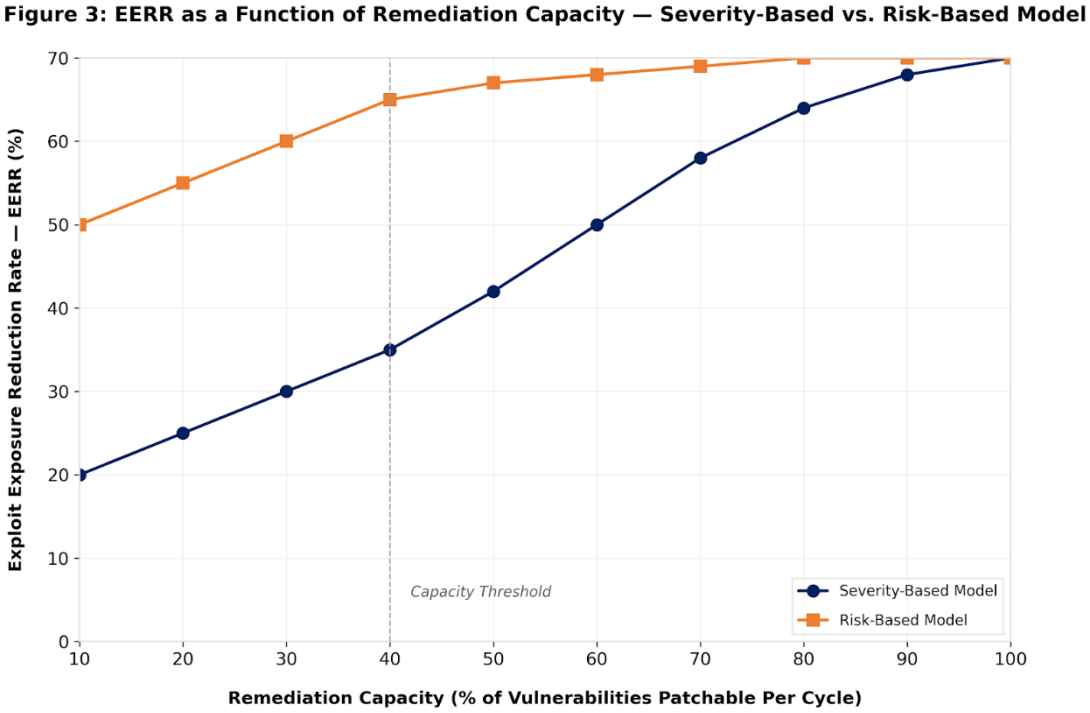

Figure 3: EERR as a function of remediation capacity

The relative advantage of each type of model varies with operation situations and Table 3 presents a qualitative framework in which the shifts in context change the relative advantage of each type of model.

Table 3: Contextual Factors

Contextual Factor	**Effect on Severity-Based Model**	**Effect on Risk-Based Model**	**Net Advantage**
High asset homogeneity	Moderate performance; CVSS correlates with impact	Marginal improvement over severity-based	Severity-based acceptable
High asset heterogeneity	Poor performance; CVSS ignores context	Strong improvement; asset weighting critical	Risk-based strongly preferred
High-volume exploit activity (broad threats)	KEV misses common in CVSS ordering	KEV coverage prioritized	Risk-based preferred
Targeted APT threat environment	Minimal contextual signal in CVSS	APT-specific intelligence improves accuracy	Risk-based preferred; quality-dependent
High remediation capacity (>80% per cycle)	Good compliance; exposure gap small	Similar outcomes; overhead less justified	Severity-based sufficient
Low remediation capacity (<40% per cycle)	High exposure gap; wrong patches prioritized	Strong exposure reduction per unit of effort	Risk-based strongly preferred

Regulatory compliance reporting requirement	Direct support; CVSS maps to audit tiers	Friction; tier instability complicates reporting	Severity-based compliance-friendlier
Industrial Control System (ICS) environment	Inadequate; uptime constraints ignored	Configurable; supports SmartPatch-style logic	Risk-based preferred (Yadav et al., 2022)

7. Remediation SLAs

In case risk-based prioritization has better results in terms of exposure reduction, then remediation SLAs should be redefined to reflect the results of exposure reduction as well as tier compliance. This happens not to be a minor transformation. It has organizations shift the reporting on the percentage of Critical vulnerabilities patched within 15 days over to reporting metrics of EERR, KEV coverage rate and mean time to remediate (purple) on vulnerabilities with active evidence of exploitation.

The reformulations which can be made are practical in a few cases. The first is a hybrid SLA scheme: keep a layer, based on CVSS anchors, of default vulnerability traffic, but introduce a new layer - called Threat-Active - which records any entry of a KEV catalog or any entry with a high score of EPSS without reference to the severity of the CVSS. This level receives a shorter deadline 7 days, and is followed independently. The Threat-Active tier adherence will be the main metric board-level. The compliance with CVSS-tier is a second-level indicator of operation. Mean TVE reduction is set as the baseline of remediation activity, instead of the number of patches. This method needs a more advanced tooling and it might be more difficult to convince non-technic stakeholders, yet, the incentives are placed in the right way. Patching 50 or more high-exposure but not fewer than 200 low-exposure vulnerabilities decreases TVE more, which is reflected by the metric.

Figure 4: Transition from Compliance-Oriented to Exposure-Oriented Vulnerability Reporting — Dashboard Comparison

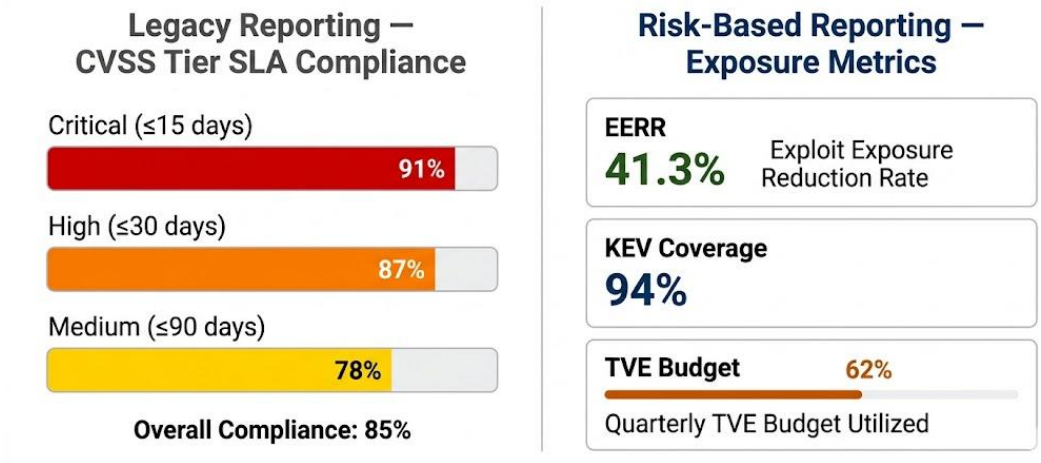

Figure 4: Dashboard mockup

Switching to exposure-based SLA reporting is in practice slow due to audit frameworks and regulatory requirements combined with contractual obligations that are structured around CVSS tier compliance. Examples include PCI-DSS which has mandated timelines (remediation) based on the levels of CVSS severity. Companies with such requirements cannot just do away with CVSS-based SLAs. To transition to the exploit-based metrics they require two reporting structures, one of which meets regulatory measures and the other one of which quantifies actual exposure with a longer-term aim of changing the standard frameworks to reflect the new measure based on exploits.

Research reported the large discrepancy between diverse data sources in evaluations of the severity of CVSS, introducing measurement error in any system that models using such scores but that discrepancy is a problem in data quality, rather than a limitation

of the idea of severity-based triage. An increased level of severity data would enhance the two types of models.

8. Conclusion

Remediation SLAs specify the pace of the security programs. They dictate the vulnerabilities that are fixed and when, determine how teams spend resources, and are the main reporting indicator to security leadership. It is important to get them in the right way. The existing SLAs based on severity anchored on CVSS. They are malleable, verifiable and well comprehended. They also do not provide a specific tool to minimize the exposure to exploits as the CVSS scores are computed without using any real attacker behaviour or asset exposure as well as evidence of exploitation. Risk-based models address these shortcomings by adding the exploit predictors, threat intelligence and asset criticality into priority logic augmentation. The outcome is a statistically significant improvement in exposure mitigation with the same resource limitations as evidenced in a number of empirical studies.

References

[1] Alperin, K., Wollaber, A., Ross, D., Trepagnier, P., & Leonard, L. (2019). Risk Prioritization by Leveraging Latent Vulnerability Features in a Contested Environment. *Risk Prioritization by Leveraging Latent Vulnerability Features in a Contested Environment*, 49–57. https://doi.org/10.1145/3338501.3357365

[2] Croft, R., Babar, M. A., & Li, L. (2021). An Investigation into Inconsistency of Software Vulnerability Severity across Data Sources. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2112.10356

[3] Farris, K. A., Shah, A., Cybenko, G., Ganesan, R., & Jajodia, S. (2018). VULCON. *ACM Transactions on Privacy and Security*, *21*(4), 1–28. https://doi.org/10.1145/3196884

[4] Mehri, V. A., Arlos, P., & Casalicchio, E. (2022). Automated Context-Aware Vulnerability Risk Management for patch prioritization. *Electronics*, *11*(21), 3580. https://doi.org/10.3390/electronics11213580

[5] Olswang, A., Gonda, T., Puzis, R., Shani, G., Shapira, B., & Tractinsky, N. (2022). Prioritizing vulnerability patches in large networks. *Expert Systems With Applications*, *193*, 116467. https://doi.org/10.1016/j.eswa.2021.116467

[6] Roytman, M., & Jacobs, J. (2019). The complexity of prioritising patching. *Network Security*, *2019*(7), 6–9. https://doi.org/10.1016/s1353-4858(19)30082-0

[7] Shah, A., Farris, K. A., Ganesan, R., & Jajodia, S. (2019). Vulnerability selection for Remediation: An Empirical analysis. *The Journal of Defense Modeling and Simulation Applications Methodology Technology*, *19*(1), 13–22. https://doi.org/10.1177/1548512919874129

[8] Spring, J. M., Hatleback, E., Householder, A., Manion, A., Shick, D., & SOFTWARE ENGINEERING INSTITUTE, CARNEGIE MELLON UNIVERSITY. (2019). *PRIORITIZING VULNERABILITY RESPONSE: A STAKEHOLDER-SPECIFIC VULNERABILITY CATEGORIZATION*. https://www.sei.cmu.edu/documents/583/2019_019_001_636391.pdf

[9] Walkowski, M., Oko, J., & Sujecki, S. (2021). Vulnerability management models using a common vulnerability scoring system. *Applied Sciences*, *11*(18), 8735. https://doi.org/10.3390/app11188735

[10] Yadav, G., Gauravaram, P., Jindal, A. K., & Paul, K. (2022). SmartPatch: A patch prioritization framework. *Computers in Industry*, *137*, 103595. https://doi.org/10.1016/j.compind.2021.103595

[11] Zeng, Z., Yang, Z., Huang, D., & Chung, C. (2021). LICALITY—Likelihood and criticality: vulnerability risk prioritization through logical reasoning and deep learning. *IEEE Transactions on Network and Service Management*, *19*(2), 1746–1760. https://doi.org/10.1109/tnsm.2021.3133811