Assessing the Cyber Risk of Autonomous System Networks using Dynamic Bayesian Models

Elliot Chmil and Jacob Fulp
Department of Computer Science
Furman University
Greenville, SC, USA
chmiel5@furman.edu

Errin W. Fulp
Department of Computer Science
Wake Forest University
Winston-Salem, NC, USA
fulp@wfu.edu

***Abstract*—Accurately and continuously assessing cyber risk remains a core challenge in modern network defense. Conventional Cyber Threat Intelligence (CTI) products rely heavily on observed Indicators of Compromise (IoCs), which provide only partial or lagging visibility into malicious activity. As a result, defenders often lack persistent, data-driven measures of how threat exposure evolves across network entities such as Autonomous Systems (ASes). This paper introduces a probabilistic framework for estimating the security risk of ASes using time-varying latent risk scores derived from network threat intelligence. The method integrates heterogeneous threat signals, including IP reputation measures, observed threat events, and associated severity indicators, within a Bayesian Gaussian Random Walk model that infers a latent risk trajectory for each AS. Evaluation using both synthetic and real-world threat datasets demonstrates that the framework captures meaningful temporal trends, quantifies uncertainty during periods of sparse observation, and provides a continuous view of AS-level threat exposure.**

***Index Terms*—security, risk scoring, cyber threat intelligence, indicators of compromise, autonomous systems, security events**

I. Introduction

Cyber Threat Intelligence (CTI) often consists of Indicators of Compromise (IoCs), such as malicious IP addresses, CIDR ranges, domain names, and URLs, which enable defenders to identify, prioritize, and mitigate threats in real time. CTI is most commonly derived from observed malicious events, for example, data collected from firewalls, intrusion detection systems, or network honeypots, and IoCs are typically reported only after hostile activity has occurred [1]. However, it is unlikely that all malicious behavior is observed or reported, and relationships among IoCs (such as repeated attacks from the same network) may remain under-appreciated. Consequently, defenders may hold an incomplete view of the threat landscape and be unable to accurately assess the risk posed by emerging or recurring sources of malicious traffic.

Developing methods that can infer risk even in the presence of incomplete intelligence is essential to improving situational awareness and guiding adaptive defensive actions. Grouping IoCs by their operational context helps reveal broader patterns that isolated observations miss. Since IP-based IoCs can be mapped to an Autonomous System (AS) [2], a network operated by a single organization, aggregating CTI at the AS level exposes recurring behaviors, shared attack characteristics, and historical tendencies of the network as a whole. This makes the AS a natural unit for assessing cyber risk and motivates modeling approaches that infer network-level threat posture even when CTI is incomplete.

Traditional cybersecurity risk assessment focuses on the likelihood and impact of threats to systems, applications, or organizational infrastructures [3], with examples including CVSS, enterprise risk frameworks (NIST RMF, ISO 27005), threat modeling (STRIDE) [4], and network monitoring. These approaches, however, typically assess risk at the level of individual assets rather than entire networks or ASes [5], [6], [7]. Simple metrics, such as raw counts of malicious events or static blacklists, cannot distinguish between networks that occasionally generate high-severity threats and those that consistently produce lower-severity activity. Quality-of-Information (QoI)-based ASN reputation systems [8] improve upon raw counts by weighting observations based on source reliability, timeliness, and analyst validation. While effective at incorporating high-confidence, human-verified incidents, QoI remains retrospective, only previously documented activity contributes to AS risk, limiting responsiveness to emerging threats and short-term behavioral shifts.

This paper proposes a methodology for AS-level cyber risk estimation that integrates historical threat activity, threat diversity, and temporal dynamics. A dynamic Bayesian framework using Gaussian Random Walks is introduced to model latent risk evolution while combining multiple CTI signals. By discretizing continuous latent scores into risk levels, the framework provides actionable metrics for security operations. Risk is quantified based on intensity, severity, recency, longevity, and diversity of malicious activity from each AS. Experiments with simulated and actual CTI from multiple AS networks demonstrate that this approach produces risk scores that are both temporally and spatially informative, offering a nuanced and actionable view of network threats.

II. CTI, Threat Associations, and Risk

Cyber Threat Intelligence (CTI) provides reports of malicious activity through Indicators of Compromise (IoCs), such as IP addresses, CIDR ranges, domain names, URLs, and file hashes. Intelligence often includes contextual metadata characterizing the threat, such as MITRE ATT&CK tactics or

severity ratings [9]. For example, IP-based IoCs are frequently linked to Common Vulnerabilities and Exposures (CVEs), each with a corresponding CVSS score quantifying vulnerability severity [10].

While individual IoCs provide insight into specific threats, grouping them can reveal broader patterns. IP-based IoCs can be mapped to an Autonomous System (AS), an administrative domain identified by an Autonomous System Number (ASN) that manages a collection of IP addresses for a single organization, such as a service provider, enterprise, or university. Because ASes reflect operational control and routing-policy boundaries, they offer a natural level of abstraction for assessing network reputation and cyber risk. Aggregating threat information across all IoCs associated with an AS can reveal its overall security posture; for instance, repeated exploitation of high-severity CVEs from hosts within the same AS may indicate poor patch management or a persistently compromised environment.

Quantifying AS-level risk is challenging. IoCs are unevenly observed, the threat landscape evolves rapidly, and relationships among IoCs, contextual metadata, and ASes change over time. Simple metrics, such as event counts or average severity, cannot capture temporal dynamics, threat diversity, or contextual nuance, and may therefore yield incomplete or misleading assessments of AS-level risk. Thus, probabilistic modeling of cyber risk, explicitly incorporating uncertainty and temporal drift, can potentially provide defenders with more reliable and informative estimates.

III. A Bayesian Approach for AS Risk Assessment

The proposed approach estimates Autonomous System (AS) level cyber risk by integrating diverse threat intelligence signals into a dynamic probabilistic framework. This approach models the evolution of latent risk over time, enabling both temporal comparability (how risk changes day-to-day within the same AS) and spatial comparability (how risk differs across ASes at the same point in time). The latent risk of each AS is treated as a hidden variable evolving according to a Gaussian Random Walk (GRW) process.

A. Model Overview

Consider a Cyber Threat Intelligence (CTI) dataset of IP-based Indicator of Compromise (IoC) publications, which are updated at different times depending on provider cadence. Each publication may include contextual information, such as a severity score and a threat category. Each IP can also be mapped to its Autonomous System (AS), providing spatial association. Over time, the same IP may appear multiple times in CTI feeds with varying context as threats evolve. Therefore, it is important for the risk estimate to adjust dynamically to reflect the current environment.

The true risk state of an AS a at time t, denoted $r_{a,t}$, is not directly observable. Instead, it is reflected indirectly through observable threat intelligence signals, including the number of malicious IPs, diversity of threats, severity of attacks, and longevity of activity. These signals act as emissions from the latent risk process. The latent risk evolves over discrete time according to a GRW:

$$r_{a,t} = r_{a,t-1} + \epsilon_{a,t}, \qquad \epsilon_{a,t} \sim \mathcal{N}(0, \sigma^2_{\text{evol}}), \tag{1}$$

where $\epsilon_{a,t}$ represents stochastic fluctuations in the threat environment. The variance σ^2_{evol} controls the expected smoothness versus volatility of the latent trajectory, allowing gradual drift under normal conditions and occasional abrupt changes during sudden surges in malicious activity.

B. Per-IP Feature Computation

For an Autonomous System (AS) a, let I_a denote the set of IP addresses observed within that AS. For each IP $i \in I_a$, let P_i be the set of CTI publications associated with that IP. Each publication $p \in P_i$ includes a threat severity value $p(v)$ (e.g., CVSS), a timestamp $p(s)$, and optionally a reported threat identifier $p(c)$ (such as a MITRE tactic or CVE). Let t denote the current evaluation time, and define the elapsed time since publication as $\Delta t_p = t - p(s)$. Using this formulation, features that capture distinct temporal and behavioral aspects of malicious activity can be derived for each IP.

The first feature, the recency-weighted severity, quantifies the immediate relevance and seriousness of the threat. It is defined as

$$f^r_{i,t} = \sum_{p \in P_i} \frac{p(v)}{1 + \Delta t_p}, \tag{2}$$

which assigns greater weight to recent and severe publications while gradually diminishing the influence of older ones. The second feature, longevity, captures the persistence of malicious behavior by measuring the duration over which the IP has remained active. It is expressed as

$$f^l_{i,t} = p_{\max}(s) - p_{\min}(s), \tag{3}$$

where $p_{\max}(s)$ and $p_{\min}(s)$ represent the most recent and earliest publication timestamps, respectively. Finally, diversity measures the functional breadth of threat behavior, let $p(c)$ be the threat category for publication p, then diversity can be measured by counting the number of distinct threat categories associated with the IP:

$$f^d_{i,t} = |\{p(c) : p \in P_i\}|. \tag{4}$$

Together, these features characterize how recently and severely an IP has been active, how long it has persisted, and how many different attack vectors or vulnerabilities it has exhibited.

C. Per-AS Score Aggregation

To obtain a single scalar score value per IP i that combines these aspects, the features are multiplied:

$$v_{i,t} = f^r_{i,t} \cdot f^l_{i,t} \cdot f^d_{i,t}. \tag{5}$$

Multiplication emphasizes joint expression of severity, persistence, and diversity, so high risk only emerges when multiple threat dimensions reinforce one another. Alternative combinations (e.g., additive or log-linear) can produce similar trends but are omitted due to space.

The per-AS score is then computed by summing across all IPs in the AS:

$$x_{a,t} = \sum_{i \in I_a} v_{i,t}. \tag{6}$$

This scalar captures the overall observed risk signal for the AS at time t, reflecting the combined recency, severity, persistence, and diversity of malicious activity across all its IPs.

D. Temporal Decay and Normalization

To reduce noise while remaining sensitive to recent activity, the per-AS score $x_{a,t}$ is smoothed using a decay-weighted temporal average:

$$\tilde{x}_{a,u} = \frac{\sum_{t=1}^{u} e^{-\lambda(u-t)} x_{a,t}}{\sum_{t=1}^{u} e^{-\lambda(u-t)}}, \tag{7}$$

where λ is the decay rate, t indexes past days, and u denotes the current evaluation day. For each day u, the smoothed score $\tilde{x}_{a,u}$ is computed as a weighted average over all previous daily scores $x_{a,1}, x_{a,2}, \ldots, x_{a,u}$, assigning exponentially higher weight to more recent activity. This approach preserves memory of past observations while emphasizing the most current signals. The resulting $\tilde{x}_{a,u}$ serves as the observation input to the Bayesian Gaussian Random Walk model. Thus, $\tilde{x}_{a,u}$ represents a temporally smoothed version of the raw score $x_{a,t}$ rather than the latent risk itself.

E. Observation Model and GRW Inference

Each AS-level observation $\tilde{x}_{a,u}$ is treated as a noisy measurement of the underlying latent risk $r_{a,u}$. Observations are assumed Gaussian:

$$\tilde{x}_{a,u} \mid r_{a,u} \sim \mathcal{N}(r_{a,u}, \sigma^2_{\text{obs}}), \tag{8}$$

where σ^2_{obs} captures variability in reporting, incomplete visibility, and noise in CTI sources. Again, "latent" refers to the unobserved true risk level, whereas "posterior" denotes the model's updated belief about that latent risk after incorporating the observation $\tilde{x}_{a,u}$.

Posterior inference is performed using a particle filter, a sequential Monte Carlo method that approximates the posterior distribution of a latent variable using discrete samples called particles [11]. At each time step, the model maintains n particles, each representing a plausible latent risk value $r_{a,u}$. By propagating and updating this ensemble over time, the filter captures uncertainty in the latent state and adjusts to sudden changes in the observed data.

At each time step, the latent risk associated with the j-th particle (out of n total particles) for AS a, denoted $r^{(j)}_{a,u}$, is first propagated according to the GRW prior:

$$r^{(j)}_{a,u} = r^{(j)}_{a,u-1} + \epsilon^{(j)}_{a,u}, \qquad \epsilon^{(j)}_{a,u} \sim \mathcal{N}(0, \sigma^2_{\text{evol}}), \tag{9}$$

which allows the latent risk to evolve gradually while retaining stochastic variability. Each particle is then assigned a weight based on the Gaussian likelihood of the current observation $\tilde{x}_{a,u}$ given that particle:

$$w^{(j)} \propto \exp\left[-\frac{(\tilde{x}_{a,u} - r^{(j)}_{a,u})^2}{2\sigma^2_{\text{obs}}} \right]. \tag{10}$$

This weighting corresponds to the probability density of observing $\tilde{x}_{a,u}$ if the true latent risk were $r^{(j)}_{a,u}$, assuming Gaussian measurement noise with variance σ^2_{obs}. Particles that are closer to the observed value receive higher weight, reflecting higher likelihood under the model, whereas particles further from the observation receive lower weight.

After weighting, particles are resampled so that high-weight particles are more likely to be duplicated and low-weight particles are more likely to be discarded. This ensures that the particle population continues to represent the posterior distribution accurately, focusing computational effort on likely latent states. When observations are stable, particles remain near previous values, producing smooth evolution; when sudden spikes occur, particles near the new level are favored, producing rapid adjustments. The posterior mean of the resampled particles gives the final AS-level risk estimate, providing a single, time-varying measure of AS cyber risk.

F. Risk Score Parameters

The proposed risk scoring framework depends on three primary parameters λ, σ^2_{obs}, and σ^2_{evol}, each shaping the behavior of the posterior estimate. The decay factor λ determines how quickly historical events lose influence over time: smaller values retain older activity, whereas larger values emphasize recent behavior. The observation variance σ^2_{obs} controls the model's trust in newly observed features; small values make the filter highly responsive to spikes, while larger values favor smoother estimates. The evolution variance σ^2_{evol} governs expected day-to-day variability of latent risk, with small values implying slow drift and larger values allowing rapid shifts.

Although expert knowledge can guide these settings, the parameters may also be estimated dynamically from historical data to maintain strict causality. A sliding-window procedure examines short-term and long-term variability in past observations to infer appropriate values: σ^2_{obs} can be estimated from recent residual variance, σ^2_{evol} from fluctuations in successive differences, and λ from the characteristic temporal scale observed in the window. This enables the model to adjust to evolving threat behavior without relying on future information.

IV. Experimental Evaluation

In this section, the performance of the proposed Bayesian AS risk assessment method is evaluated against two static causal baselines, *static* and *decayed-static*. For each baseline, a daily IP-level score is computed from the earliest last-seen date to the analysis date. These scores use three components: a recency weight $w_d = 1/(1+\text{days since last seen})$, a longevity factor $1 + 0.05 \times$ mean days seen, and a diversity factor $1 + 0.1 \times \log(1 + \text{unique categories})$. The *static* method sums daily scores cumulatively, while *decayed-static* applies an additional exponential decay with a 10-day half-life.

Bayesian estimation models latent AS risk as a Gaussian random walk with 10,000 particles. The parameters λ, σ^2_{obs}, and σ^2_{evol} control the decay of historical influence, the trust placed in daily observations, and the day-to-day volatility of the latent process. These parameters are inferred dynamically

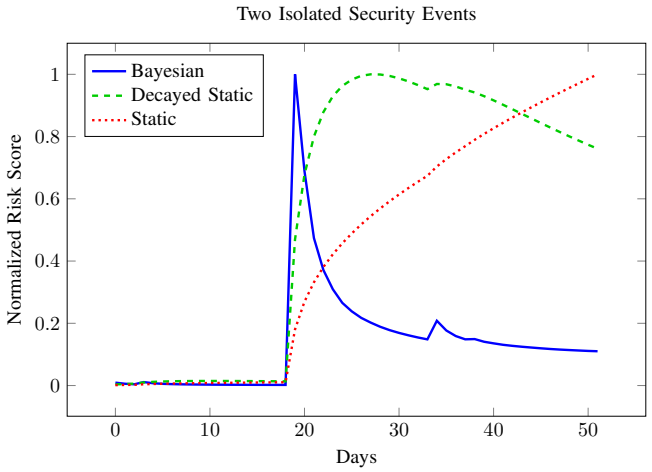

Fig. 1. Normalized risk trajectories for a single AS in response to two isolated security events. Bayesian risk highlights event impact while maintaining a stable baseline.

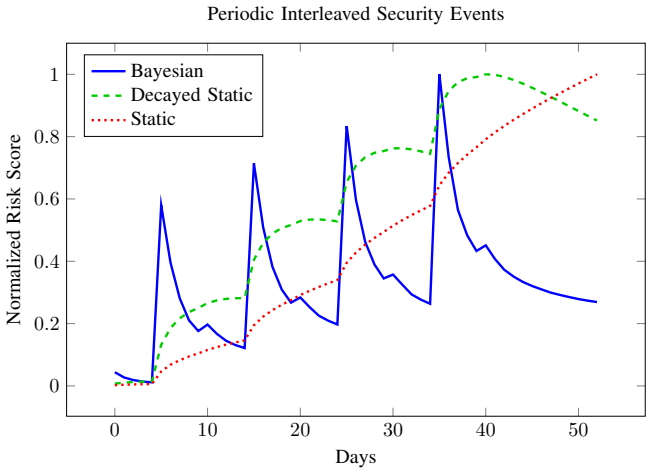

Fig. 2. Normalized risk trajectories for a single AS under alternating high- and moderate-severity events. Bayesian risk differentiates spike magnitude and maintains baseline stability.

from the previous 30 days as described in the previous section. Particle weights are updated according to the likelihood of observed scores, producing a posterior distribution that captures uncertainty while remaining causal and temporally adaptive.

Performance of the risk assessment approaches is assessed using the normalized area under the curve (AUC) of each method's cumulative risk trajectory. A larger AUC indicates greater total allocated risk, though this may reflect accumulation rather than meaningful discrimination. Conversely, smaller AUC values may indicate more selective or more responsive behavior.

A. Experimental Results

In the first experiment, an AS with 256 IPs experiences multiple event spikes (IoC publications consisting of an IP address, CVSS value, and CVE). Spikes are either large (40 unique events, CVSS 10) or small (5 unique events, CVSS 5). A large spike occurs at day 19, followed by a small spike at day 34, with low-level background activity throughout the 60-day period. As shown in Figure 1, the Bayesian estimator localizes the large spike precisely, with normalized risk peaking on day 19, whereas the static and decayed-static methods peak later (days 51 and 27). These delays reflect the cumulative and smoothing behavior of the baselines, which integrate both spikes and background noise. The Bayesian method concentrates risk sharply around the large spike, yielding an AUC of 3.78 in the spike window, slightly above static (3.65) and well below the broader accumulation of decayed-static (7.78). Pre-spike variability further distinguishes the methods: Bayesian and static maintain low variance (std $\approx$ 0.0027 and 0.0024), while decayed-static exhibits nearly double (std $\approx$ 0.0043). All methods detect the smaller spike at day 34, but the Bayesian estimate reflects it more distinctly relative to noise, whereas decayed-static spreads its influence and static accumulates it gradually. Overall, Experiment 1 shows that the Bayesian estimator isolates both major and minor spikes from background noise, providing temporally precise risk estimates, while the static and decayed-static methods trade temporal accuracy for smoothness and long-term accumulation.

The second experiment introduces alternating periodic spikes for the same 256-IP AS (Figure 2). Small spikes occur at days 0, 10, 20, ... and large spikes at days 5, 15, 25, ..., continuing until day 40. The Bayesian estimator tracks spike timing closely, with normalized risk peaking on day 35 (the final large spike), while static and decayed-static peak later (days 52 and 40). AUC ratios highlight differences in sensitivity to spike magnitude: the Bayesian method yields comparable AUCs for large and small spikes (0.115 vs. 0.104, ratio 1.11), indicating balanced responsiveness across event sizes. Static and decayed-static methods disproportionately emphasize large spikes, with higher ratios (1.37 and 1.33). Variability trends mirror these behaviors: Bayesian risk exhibits the highest variance (std $\approx$ 0.203), static the lowest (std $\approx$ 0.074), and decayed-static falls between them (std $\approx$ 0.150). Short-term pre-spike rolling variability shows the same pattern (0.084, 0.011, and 0.022). Overall, Experiment 2 demonstrates that the Bayesian estimator provides high temporal resolution and balanced sensitivity to event magnitude, whereas static and decayed-static methods delay peak detection and disproportionately weight larger events.

The third experiment evaluates how AS size influences Bayesian risk when all ASes experience the same event at the same time. Four ASNs with 32, 64, 128, and 256 IPs each receive ten CVSS 10 events on day 1. As shown in Figure 3, peak normalized risk is inversely proportional to AS size: the 32-IP AS reaches the highest peak (approx. 1.0), while the 256-IP AS peaks near 0.125. Since identical events represent a larger fraction of a smaller address space, smaller ASes naturally appear riskier. Correspondingly, AUC values decrease with AS size, ranging from 2.85 for the smallest to 0.36 for the largest. Experiment 3 thus confirms that the Bayesian method reflects expected scaling properties: identical

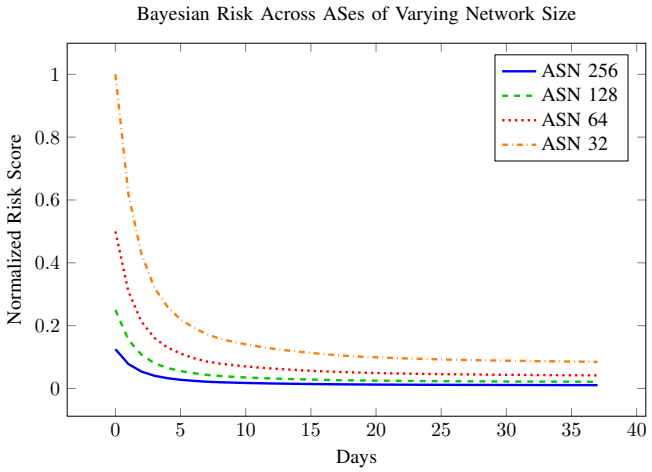

Fig. 3. Normalized risk values for multiple ASes with different network sizes, demonstrating proportional risk scaling and comparability.

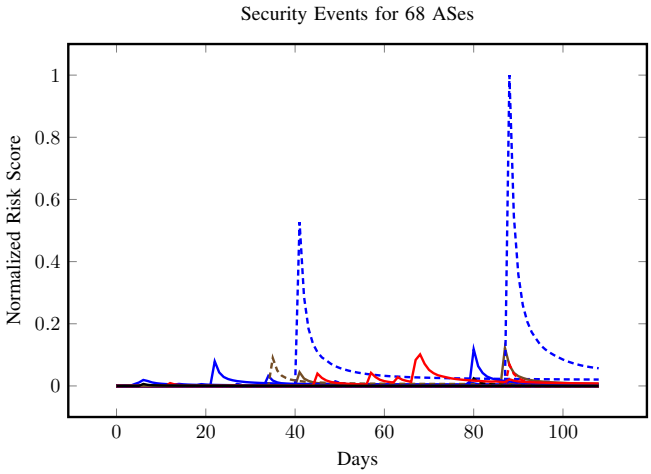

Fig. 4. Normalized risk trajectories for actual ASes. Bayesian risk maintains fairness and responsiveness across all ASNs.

events have proportionally greater impact in smaller networks.

The final experiment analyzes 68 ASes with real security-event data from Greynoise (June until September 2025). AS sizes range from 256 to over 15 million IPs (mean approx. 1.33 million). Daily event volumes per AS range from 1 to 11,324 (mean 448), with CVSS values between 2.7 and 10 (mean 8.66). Event activity is extremely sparse at the IP level: each IP is associated with only 0.008 events on average over the full period. Figure 4 illustrates the diversity of this dataset and its suitability for evaluating responsiveness and robustness.

Across these ASes, Bayesian risk consistently outperforms decayed-static risk in ranking ASes by event activity and identifying those contributing to the top 20% of daily events. Rank correlations show this improvement: Bayesian risk achieves slightly higher Spearman (rho $\approx$ 0.135) and Kendall (tau $\approx$ 0.111) values than decayed-static (rho $\approx$ 0.128, tau $\approx$ 0.112). Discriminative performance also differs: Bayesian risk attains a ROC AUC of 0.883 versus 0.838 for decayed-static, indicating better separation of high- and low-activity ASes. These results confirm that the Bayesian method provides more accurate and timely AS-level risk signals on real-world CTI data.

V. Conclusions and Future Work

This paper introduced a dynamic Bayesian framework for assessing Autonomous System (AS) level cyber risk using time-varying threat intelligence signals. By modeling latent risk as a Gaussian Random Walk, the approach captures smooth yet adaptive temporal evolution, distinguishing between transient fluctuations and sustained changes in threat behavior. Unlike static or exponentially decayed risk models, the Bayesian formulation integrates uncertainty directly into inference, ensuring that risk scores remain robust under sparse or noisy observations.

Future work will explore several promising directions. First, incorporating network topology and inter-AS dependencies could enable the model to capture threat propagation across related networks. Second, further experiments are needed to better understand the limitations of the linear assumptions underlying the current model and to evaluate the potential of hierarchical Bayesian approaches to address these limitations. Third, modeling feature-wise variances or using a multivariate emission model may improve performance, since different features can operate at different scales. Finally, empirical validation using operational incident data or known compromise timelines would demonstrate the model's practical value for network defense and policy management.

References

[1] X. Bouwman, H. Griffioen, J. Egbers, C. Doerr, B. Klievink, and M. Van Eeten, "A different cup of {TI}? the added value of commercial threat intelligence," in *29th USENIX security symposium (USENIX security 20)*, 2020, pp. 433–450.

[2] D. Magoni and J. J. Pansiot, "Analysis of the autonomous system network topology," *ACM SIGCOMM Computer Communication Review*, vol. 31, no. 3, pp. 26–37, 2001.

[3] P. Cheimonidis and K. Rantos, "Dynamic risk assessment in cybersecurity: A systematic literature review," *Future Internet*, vol. 15, no. 10, p. 324, 2023.

[4] N. Shevchenko, T. A. Chick, P. O'riordan, T. P. Scanlon, and C. Woody, "Threat modeling: a summary of available methods," Tech. Rep., 2018.

[5] J. Karlin, S. Forrest, and J. Rexford, "Autonomous security for autonomous systems," *Computer Networks*, vol. 52, no. 15, pp. 2908–2923, 2008.

[6] A. Y. Nur and M. E. Tozal, "Identifying critical autonomous systems in the internet," *The Journal of Supercomputing*, vol. 74, no. 10, pp. 4965–4985, 2018.

[7] J. Chang, K. K. Venkatasubramanian, A. G. West, S. Kannan, B. T. Loo, O. Sokolsky, and I. Lee, "As-trust: A trust quantification scheme for autonomous systems in bgp," in *International Conference on Trust and Trustworthy Computing*. Springer, 2011, pp. 262–276.

[8] S. Hutchinson and R. F. Erbacher, "Asn reputation system model," in *Cyber Sensing 2015*, vol. 9458. SPIE, 2015, pp. 61–68.

[9] A. Kuppa, L. Aouad, and N.-A. Le-Khac, "Linking cve's to mitre att&ck techniques," in *Proceedings of the 16th International Conference on Availability, Reliability and Security*, 2021, pp. 1–12.

[10] K. Scarfone and P. Mell, "An analysis of cvss version 2 vulnerability scoring," in *2009 3rd International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2009, pp. 516–525.

[11] A. Doucet, N. de Freitas, and N. J. Gordon, *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer, 2001.