JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Survey on AI-Enhanced Network Vulnerability Assessment and the Hybrid AMC-MDP Framework

[1]M. Amareshwar Sai, [2]Kewal Ajaykumar Dharamshi, [3]K. Krithik Bose, [4]Vishal Rathod, [5]Padmavathi S
[1 2 3 4]Student, [5]Professor,
[1 2 3 4 5]Computer Science Engineering (Cyber Security),
[1 2 3 4 5]Dayananda Sagar College of Engineering, Bangalore, India

Abstract: Modern enterprises today have networks that are vulnerable to quickly changing threats, face attacks where thousands of discovered vulnerabilities interact with each other in a system, and thus need both probabilistic risk assessments, combined with adaptive planning for possible courses of action. Traditional vulnerability scoring models (such as CVSS and EPSS) provide potentially useful benchmarks, but fail to capture the evolution or propagation of attack scenarios across interconnected network assets.In this paper, we examine early work on network-level vulnerability assessment frameworks involving everything from deterministic scoring and adaptive learning to stochastic modeling, and we outline a broad pattern of failure and identify research gaps that tie together the two essential activities of risk measurement and decision making.To help resolve these gaps, we present a Hybrid Absorbing Markov Chain–Markov Decision Process (AMC–MDP) framework that estimates longterm compromise probabilities while learning optimal defensive actions over time.The model displays certain structural features based on graph paths and incorporates risk, cost, and reward into a unified action decision loop.The work synthesizes earlier literature that indicates this sort of hybridization can bridge the analytical transparencies needed for risk measures and adaptive controls, which can provide the basis for scaling up economically sustainable, interpretable cyber defense approaches.

IndexTerms **- Vulnerability Assessment, Attack Graphs, Markov Decision Process (MDP), Absorbing Markov Chain (AMC), Reinforcement Learning, Cybersecurity Automation.**

I. INTRODUCTION

Modern enterprise networks are experiencing an historic increase in both quantity and sophistication of cyberattacks. Studies report that new vulnerabilities appear every few minutes, creating an expanding attack surface that far exceeds the remediation capabilities of most organizations [1]. The 2025 Verizon-style breach analyzes consistently indicate that a significant fraction of breaches still arise from known unpatched vulnerabilities, while the average global breach cost remains high [2]. These numbers indicate a consistent disparity in the discovery of vulnerabilities, and how we prioritize them. Despite the maturity of scanning technologies, organizations still struggle to identify the most critical vulnerabilities within dynamic interconnected infrastructures [3], [4]. Earlier studies on biometric based cryptographic systems also show how strong authentication forms an important base for building reliable security models[5]

Conventional vulnerability management workflows— summarized as “scan → list → patch by severity”—rely primarily on the Common Vulnerability Scoring System (CVSS) [1]. Although CVSS provides a standardized method of determining the severity, it does not take into account the influence of topological placement, reachability or exploit chaining as a measure of risk in practice. The Exploit Prediction Scoring System (EPSS) extends CVSS by introducing short-term exploit probability [6], reflecting a shift towards data-driven prioritization. However, EPSS remains static and pointwise, evaluating each vulnerability independently without considering the movement of the attacker between the states of the network or the dependencies among the vulnerabilities [7]. Consequently, defenders still lack an analytical model that captures both exploit likelihood and attack progression within enterprise topology.

Research in network-level vulnerability assessment has progressed through three frameworks: static, adaptive, and stochastic frameworks. Static approaches focus on simplicity but lack temporal or contextual understanding. Adaptive or ML-driven models leverage threat intelligence and telemetry to dynamically estimate exploitability [8], [9], improving responsiveness but remaining primarily reactive and asset-specific. Lightweight encryption techniques like homomorphic encryption with elliptical curve methods have also been explored for protection mobile and fast charging networks[10]. In contrast, stochastic and graph-based frameworks— employing attack graphs, Bayesian networks, and Markov-based formulations— introduce probabilistic reasoning to represent sequential attacker behavior [3], [11], [12]. These models provide a deeper understanding of risk propagation, yet they often struggle with real-time adaptability and operational scalability.

This paper examines the evolution from static severity scoring to adaptive learning and stochastic decision frameworks, evaluates their contributions and shortcomings, and propose a hybrid AMC-MDP model ,drawing from related works [13], [14]. The remainder of the paper is organized as follows. Section II reviews prior work. Section III outlines gaps and requirements. Section IV presents the conceptual hybrid methodology. Section V discusses future directions, and Section VI concludes..

II. LITERATURE SURVEY

A. Static and Deterministic Vulnerability Scoring

Novice vulnerability assessment frameworks were mainly static or deterministic scoring systems such as those implemented in Nessus and OpenVAS, backed by CVSS at its core [1]. CVSS provided standardized means of calculating exploitability and impact but it focused on inherent attributes of each vulnerability. Organizations usually prioritized vulnerabilities through organised CVSS scores, that produced long lists of remediations and were a major cause of inefficient resource utilization . To tackle this inefficiency, EPSS framework was developed that estimated the probability of a vulnerability being exploited within a 30-day time limit. [6]. EPSS marked a shift toward predictive prioritization, but both CVSS and EPSS treat vulnerabilities independently and do not represent them as a part of a chronological sequence or as a dependence across systems [2]. Without modeling multistep attack path, these static approaches cannot capture the structured nature of enterprise risk [4]. Limitations: (1) No dynamic or sequential attack behavior. (2) Neglect of topology, reachability, and attacker pathways. (3) No modeling of longterm cost–benefit trade-offs.

B. Adaptive and Machine Learning–Based Approaches

The limitations of static systems lead to the adaptation of ML-based frameworks that leverage statistical learning and pattern recognition for real-time detection and response. Turukmane and Devendiran introduced M-MultiSVM with adaptive resampling and dimensionality reduction to handle class imbalance, achieving high accuracy [8]. Dugar-LSTM employed chaotic optimization with LSTMs for intrusion detection [15]. Robustness work examined adversarial perturbations in deep IDS and proposed defenses [16]. Data augmentation in SDN with GANs improved generalization [17]. Probabilistic learners and hybrid HMM/GMM pipelines were also explored for real-time detection. Collectively, these methods showcased significant improvement in adaptivity and accuracy but highly relied on supervised data and controlled experimental setups.[18], [19]. Limitations: (1) Optimization for local detection metrics rather than long-term network resilience. (2) Topology treated as features, not as the governing substrate of attack progression. (3) Heavy dependence on labeled data and batch retraining, limiting adaptivity [9], [20].

C. Stochastic and Graph-Based Framework

The shift from adaptive to stochastic modeling represented a change from frameworks based on reactive decisions to the ones based on probabilistic reasoning and decision optimization . Stochastic frameworks treat the enterprise network as a dynamic system in which both attackers and defenders operate under uncertainty. Essential work synthesized by comprehensive surveys on attack graphs revealed the concept of logical paths an attacker can follow through interconnected vulnerabilities, providing fundamental support to understand accessibility, lateral movement, and cumulative attack probability [21]. Further research connected probabilities and transition costs with graph edges, enabling quantitative evaluation of multi-step attacks paths and the spatiotemporal transformation of compromise [11], [12]. This probabilistic methodology allowed researchers to calculate not only whether an attack is possible but also its likelihood, expected duration, and most probable path to success.

In this stochastic framework, several mathematical models capture system interactions and support defense optimization. Absorbing Markov Chains (AMCs) interpret attack graphs as state-transition systems in which absorbing states represent attacker goals and transient states represent the connections between vulnerable and normal states in a network. By calculating the fundamental matrix $N = (I - Q) - 1$, defenders can estimate expected time-to-compromise, state visit counts, and steady-state risk probabilities [22], [13]. AMC-based analysis produces interpretable and verifiable quantification of longterm exposure; however, it is intrinsically passive and does not explicitly advise defensive actions.

To incorporate decision-making, a framework known as Markov Decision Process (MDP) was developed in which states represent network posture, actions represent steps taken to transit between states , transitions evolve stochastically, and rewards justify operational objectives. Luo et al. proposed MDP-AD for real-time adaptive responses to variable and unknown attacks [23]. Liu et al. modelled defender action for partial observability by a POMDP and by deep Q-learning [24]. Optimizing proactive defence schemes such as Moving Target Defense (MTD) has utilized factored and receding-horizon MDPs to further mitigate scalability and anticipatory control challenges [14], [25]. Complementary probabilistic learners such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) model hidden phases of attack and traffic blends for purposes of anomaly detection and prediction but lack the sequential control aspect that characterizes the MDPs [19]. Collectively, the stochastic frameworks—AMCs, HMMs, GMMs, MDPs/POMDPs—lead to core principle of quantifying and reasoning of risk over time

To empirically evaluate decision-theoretic approaches, Luo et al. [23] compared MDP-AD against SVM, CNN, RNN, Transformer-IDS, and GNN across four simulated attack scenarios using KDD-Cup99 and UNSW-NB15. Experiments on a GPU-accelerated Ubuntu server reported accuracy, precision, recall, and F1-score. The MDP-based model achieved the highest average accuracy of about 94% in low-intensity settings and maintained > 91% accuracy under unseen attacks, outperforming the neural baselines by roughly 5–8%. These results demonstrate the adaptability and stability of reinforcement-learning–driven defense across varying attack intensities, quantifying the advantages of real-world decisiontheoretic policies over temporal dependencies and sequential decision dynamics that static or purely ML-based alternatives lack. Limitations: (1) Many stochastic approaches face scalability challenges, as attack graphs and MDP state spaces grow rapidly with network size. (2) Reward and transition functions in MDPs may overlook structural dependencies or economic costs of defensive actions , influencing policies that lack

foresight. (3) Most methods either quantify risk (AMC) or select actions (MDP) but rarely integrate both efficiently. These limitations motivate hybrid methodologies that combine efficient probabilistic quantification with interpretable and adaptive decision control [23], [13], [24], [14]..

TABLE I: Summary of Reviewed Literature

No.	Authors & Year	Title / Source	Core Contribution	Key Limitation / Gap
[1]	Jiang et al., 2025	A Survey on Vulnerability Prioritization (survey context aligned with [5])	Taxonomy of vulnerability scoring/prioritization and metrics	Lacks integration of topology and probabilistic attack progression
[2]	Barchuk & Volkov, 2024 [4]	Limitations of Modern Vulnerability Scanners and CVE Systems	Empirical critique of scanner outputs and CVE workflows	No exploit-likelihood modeling; low structural context
[3]	Shimizu & Hashimoto, 2025 [5]	Vulnerability Management Chaining (arXiv)	Chained scoring for prioritization	Static chaining; no stochastic/decision layer
[4]	Luo et al., 2025 [2]	MDP-AD (Alexandria Eng. J.)	Adaptive MDP for evolving/unknown attacks	High compute; simplified topology; no exploit priors
[5]	Krishnan, 2024 [6]	Generative AI for Vulnerability Management	Blueprint for AI automation in VM	No probabilistic risk integration
[6]	Akkemgari & Binny, 2025 [7]	Intelligent Security Automation	Pipeline from scan to remediation	Reactive; no long-horizon optimization
[7]	Yang & Yang, 2021 [3]	Markov Model Attack Graph	Optimal protection via Markov modeling	Static topology; limited adaptivity
[8]	Koscinski et al., 2025 [8]	Conflicting Scores, Confusing Signals (CCS)	Empirical score divergence study	No unified predictive/decision model
[9]	Jacobs et al., 2020 [9]	Exploit Prediction for Remediation	Short-term exploit likelihood (EPSS motivation)	Short horizon; no topology or control
[10]	Liu et al., 2022 [25]	AMC for Situation Assessment	AMC-based security quantification	Passive quantification; no action selection

TABLE II: Performance Comparison of Detection Methods in Four Scenarios (adapted from [2])

Method	Scenario A		Scenario B		Scenario C		Scenario D	
	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)	Acc (%)	F1 (%)
MDP-AD [2]	94.3	94.0	93.8	93.4	92.5	91.6	91.6	90.1
SVM	89.2	89.3	85.7	85.1	85.0	84.6	84.5	84.0
CNN	92.1	93.9	87.3	86.8	86.5	85.9	85.7	85.1
RNN	91.5	93.1	86.0	86.0	85.2	85.1	84.8	84.5
Transformer-IDS	91.36	93.06	85.98	85.91	85.06	84.95	84.61	84.31
GNN	91.49	92.90	85.88	85.86	85.07	84.95	84.71	84.46

III. IDENTIFIED RESEARCH CHALLENGES AND FRAMEWORK DESIGN OBJECTIVES

Although the evolution into adaptive and stochastic based defense frameworks , current research remains divided between risk quantification and decision-making. Markov Decision Process (MDP) and reinforcement learning (RL) based systems achieve real time responsiveness and circumstantial control but they still lack network topological awareness and long-term economic trade-offs. [23], [24]. On the contrary , probabilistic quantification methodologies such as Absorbing Markov Chains (AMCs) provide audit-able estimates of compromise probabilities but cannot inherently determine optimal defense actions [13], [22]. This gap between quantifying risk and deciding optimal actions defines the theoretical shortfall that a next-generation network-level vulnerability assessment framework must address.

A common limitation in many adaptive systems is structural blindness. Countless MDP or DRL-based frameworks define states using packet-level or flow-based measurable features, neglecting the graph structure of the network that governs attack propagation paths. [28], [29], [8]. Research in attack graphs and Bayesian network quantification demonstrates that topology—particularly centrality measures such as degree, betweenness, and eigenvector importance—determines lateral movement and systemic risk [3], [27], [11], [21]. However, in most adaptive systems, topology remains outlying rather than intrinsic to decision modeling. For network defense to be effective, topology must influence both transition probabilities and policy selection, enabling prioritization of actions on structurally critical assets.

Another recurring challenge is reward myopia, where reinforcement-based systems aim to maximize short-term cri- teria such as accuracy or false-alarm reduction while ignoring long-horizon defender utility. Many MDP and DRL frameworks define rewards only in terms of detection correctness [17], [18], [24], often overlooking the economic cost of remediation and the collective impact of risk reduction in critical nodes [28], [14]. The theory of reinforcement learning postulates that reward design directly governs policy optimality [30], [31]; suboptimal reward shaping yields locally effective but globally inefficient defenses. An efficient reward function should combine three aspects: (1) expected reduction in risk, (2) structural importance of the defended node, and (3) economic cost of the chosen action. Balancing these aspects allows policies to align with real-world return-on-investment (ROI) objectives.

Scalability and tractability also remain major limitations. With the growth in size in enterprise networks, state-action matrices and attack graphs exponentially increase and result in the "state explosion" issue in MDPs and AMCs [21], [24]. Recent advances also consider factored and hierarchical MDPs for scalable decomposition in policies and modular decision learning in POMDPs [14], [25]. AMC-based quantification can also use sparse-matrix approximations or probabilistic sampling in order to stay efficient [13]. Inclusion of AMCgenerated abstractions in MDP state modeling can reduce dimensionality and maintain structural semantics and therefore guarantee interpretability and computational tractability.

Another crucial prerequisite is managing uncertainty and partial observability. Network states and attack progression are rarely fully visible to real-world defenders. Although POMDP and belief-based RL frameworks make an effort to tackle this uncertainty, they are still constrained by computational overhead and static observation models [24]. Robustness under incomplete information can be improved by combining these with probabilistic inference mechanisms like Bayesian updating, Hidden Markov Models (HMMs), or Gaussian Mixture Models (GMMs) [19], [32]. Defenders can predict attacker behaviour, infer hidden states, and dynamically modify mitigation strategies thanks to this hybridisation.

Deployment is made more difficult by adaptation and nonstationarity. Static transition probabilities soon become outdated, and attack tactics are constantly changing. The need for continuous policy learning is emphasised by dynamic reinforcement learning and stochastic optimisation frameworks [28], [16], [33], [14]. Adaptability is naturally supported by a hybrid AMC–MDP framework: MDP layers use temporaldifference learning to improve decision policies, while AMC layers update risk probabilities as vulnerabilities change. Realtime responsiveness and analytical stability are both preserved by this interaction, which is essential for enterprise-scale defence.

From this analysis, several key requirements emerge for a next-generation framework:

1) Dual quantification and control: Combine AMC-based risk estimation with MDP-driven action optimization
. 2) Structural and topological awareness: Ensure network connectivity and node centrality influence transitions, rewards, and state representation.
3) Reward alignment with long-horizon ROI: Integrate cost, criticality, and long-term risk reduction.
4) Scalability and abstraction: Employ hierarchical or factored MDPs and graph compression to handle largescale networks.
5) Uncertainty modeling: Consider the use of probabilistic estimators or POMDP variants to improve the accuracy of the decision under incomplete observability.
6) Continuous adaptation: Enable dynamic risk reconfiguration and policy learning to address evolving vulnerabilities.

A hybrid Absorbing Markov Chain–Markov Decision Process (AMC–MDP) model satisfies these objectives by combining the quantitative capability of AMC-based network modeling with the decision optimization capabilities of MDPs [23], [13], [14].The AMC component mathematically evaluates the likelihood of system compromise and the expected time until absorption, while the MDP part identifies strategies that reduce the long-term expected risk. This combination transforms vulnerability assessment from a reactive and analytic task into a proactive one, where its decision driven mechanisms ensure a sustainable and a topologically aware resilient network..

IV. METHODOLOGY

The proposed framework combines the strength of Absorbing Markov Chains to quantify the long-term probabilities with the adaptive, decision-optimizing nature of Markov Decision Processes (MDP). This integration allows the system to evaluate both the probability of network compromise and the expected duration before a successful attack, while simultaneously learning the most effective defense strategies. To ensure structural awareness of the graph, node central metrics such as be-tweeness and eigenvectors are incorporated, to ensure that the policies derived align with the network's topology and accurately represent the behavior of risk propagation.

A. Formal Model Elements

We define an MDP $M = (S, A, P, R, \gamma)$, where S is the set of states, A the set of actions, $P(s'|s, a)$ the transition probability, $R(s, a)$ the reward, and $\gamma \in [0, 1]$ the discount factor [30], [31]. Each state $s \in S$ encodes both the current network security posture (e.g., vulnerability status or telemetry) and structural attributes (e.g., node centrality). Actions $a \in A$ represent defender interventions such as patching, isolating, blocking, or re-segmenting nodes. The optimal policy $\pi*(s)$ satisfies the Bellman optimality criterion:

$$\pi*(s) = \arg\max_a Q*(s, a), \quad (1)$$

where $Q*(s, a)$ denotes the optimal state–action value. The AMC layer models the probabilistic propagation of an attacker through the network. From an attack graph, an AMC is constructed with a transient block Q and an absorbing block R. The fundamental matrix is:

$$N = (I - Q)^{-1}, \quad (2)$$

which yields expected visits to transient states before absorption and the expected time-to-compromise. Absorption probabilities from $N \times R$ quantify the long-term likelihood of attackers reaching critical assets [13], [22]. These AMCderived measures are used to parameterize both the reward and transition functions of the MDP, bridging quantification with decision control..

B. Reward and Transition Coupling

To jointly model structural importance and operational cost, the reward function is formulated as:

$$R(s, a) = Rbase(s, a)+\omega\ C(v)-Cop(a)-\lambda RiskAMC(s), \quad (3)$$

where C(v) denotes the centrality of node v, Cop(a) represents the action cost, and RiskAMC(s) is the AMC-quantified compromise probability. Coefficients ω and λ balance the influence of structural significance and risk reduction [23], [14]. Transitions $P(s' |s, a)$, estimated from telemetry or simulation, are adjusted by AMC-derived risk and node centrality, emphasizing realistic attacker preferences toward high-impact nodes [21], [24].

C. Learning Process

Policy learning is achieved using Q-learning or Deep QNetworks (DQN) when state dimensionality is large. The temporal-difference update rule is:

$$Q(st, at) \leftarrow Q(st, at)+\alpha\ rt+1+\gamma \max a' Q(st+1, a')-Q(st, at) , \quad (4)$$

where α is the learning rate and rt+1 is the reward observed after executing at. This iterative process converges toward a policy maximizing the long-term expected return while adapting to AMC-updated risk estimates [28], [24].

D. AMC Quantification Layer

Within the AMC layer, each transient state corresponds to an intermediate network configuration or a partial system compromise, while absorbing states denote full system compromise or successful defense. The AMC is constructed from the attack graph by dividing it into transient block Q and absorbing block R. The fundamental matrix derived from Equation (2),

$$N = (I - Q) -1 , \quad (5)$$

provides interpretable measures of network risk dynamics. N reflects the expected number of transitions through transient states before reaching an absorbing state. This formulation supports the computation of several essential security indicators, including: : • Expected path length: the average number of steps before an attacker reaches an absorbing (compromised) state. • Expected absorption time: the anticipated duration or persistence of an attack campaign before compromise. • State visitation frequency: the chance of each node being visited in an attack chain, reflecting its systemic importance. These metrics are continuously recalculated whenever the network topology changes (e.g., via segmentation, patching, or isolation), ensuring that the AMC layer remains synchronized with evolving network states and reflects real-time security posture [22], [13], [19].

E. Integration and Topology Awareness

The hybrid AMC–MDP framework functions as a selfadaptive system in which the probabilistic output of the AMC continuously dictates the policy-learning process of MDP. After each training cycle, the updated AMC-derived risk values are inserted into the reward R(s, a) and transition probability $P(s' |s, a)$ components, ensuring that policy optimization mirrors the current attack surface and evolving threat conditions. Topological awareness is maintained by calculating graphbased centrality metrics—such as between ness, eigenvector, and closeness centrality—which are part of both state features and reward formulations [14], [28]. Whenever the network undergoes changes, the AMC matrices and centrality values are recalculated, allowing the model to adaptively reshape its perception of attack pathways and defense leverage points [4].

F. Addressing the Identified Gaps

This unified AMC–MDP framework systematically addressees the key challenges identified in prior research: • Quantification and Control: The AMC component generates interpretable probabilistic estimates of compromise likelihood and time-to-absorption, while the MDP layer interprets these insights to develops them in actionable and defense policies [13], [14]. • Structural Awareness: Network topology directly influences both state representation and policy optimization, ensuring that nodes that are critical to attack path lateral movement are prioritized for defense. [27], [4]. • Long-Horizon Optimization: Reinforcement learning is capable of maximizing security returns over long time horizons , rather than solely focusing on short term detection metrics. [9], [20]. • Scalability: Through hierarchical abstraction and graph compression, the framework maintains computational feasibility even with large and complex network environments. [14], [24]. • Adaptability: With temporal-difference learning, the system keeps improving its decision based on new data. This means even if the network environment changes or new attack patterns appear , the model can adapt quickly without needing to be re-trained. [16], [17]. • Interpretability: The probabilistic foundation of AMC ensures that all model outputs and policy decisions remain auditable and explainable to security analysts and system administrators [22], [13].

G. Scope and Future Evaluation Plan

This work presents a conceptual framework and does not include experimental validation; implementation and validation will be pursued in subsequent research. Planned experiments will use simulated enterprise networks with diverse topologies—random, scale-free, and small-world—and benchmarked or synthetically generated attack graphs [27], [4]. Comparative baselines will include:

1) CVSS-only severity ranking,
2) AMC-only probabilistic risk assessment,
3) MDP-based defense without structural modulation, and

4) heuristic patching of top-k central nodes. Evaluation criteria will include metrics such as: (a) AMC absorption probability and expected time-to-compromise, (b) cumulative cost-adjusted risk, (c) average remediation time, (d) policy stability and convergence, and (e) computational efficiency. Stress-testing will analyze performance under progressive vulnerability landscapes (non-stationarity), partial observability via POMDP extensions, and adversarial adaptation modeled through game-theoretic scenarios. Furthermore scalability assessments will examine the benefits of graph embeddings and factored MDP decompositions for largescale deployments [14], [24]. Collectively, these experimental studies aim to validate the framework's real world applicability, scalability and efficiency within complex enterprise environments. V. DISC

V. RESEARCH OPPORTUNITIES AND FUTURE WORK

The proposed Hybrid AMC–MDP framework offers a unified approach that integrates probabilistic quantification, structural reasoning, and long-horizon decision optimization. While the model conceptually addresses prior gaps in vulnerability assessment, scalable deployment and real-world adaptation introduce new frontiers for research. The following directions extend beyond the limitations already addressed within the hybrid framework.

1) Scalable architectures and state abstraction: Although the hybrid model mitigates complexity by grouping states using AMC, large enterprise networks still demand additional abstraction mechanisms. A possible improvement is to investigate hierarchical and factored reinforcement learning architectures, where decisions are broken in manageable layers or components using graph-based embeddings, and modular subnet decomposition to support expansive, diverse infrastructures [14], [24]. Recent works highlights how adaptive routing in IOT systems can simultaneously imporve security and reduce energy consumption[34]. The integration of graph neural encoders with reinforcement learning can enable compact structural representation of the network while still maintaining the relation between the nodes [28], [9].

2) Advanced uncertainty handling: In real world security operations , defenders often operate under incomplete or delayed information. Extending the framework toward partially observable MDPs (POMDPs) and Bayesian inference allows maintenance of decision accuracy under incomplete information [24], [19]. Hidden Markov Models (HMMs) can also help better estimate hidden attack states, providing greater resilience to stealthy or ambiguous attack behaviors in dynamic environments.

3) Real-world data integration and empirical calibration: For the hybrid model to work in real world enterprise environments, its parameters need to reflect real world telemetry and threat data. Integrating live exploit prediction and threat intelligence feeds—such as the Exploit Prediction Scoring System (EPSS)—can provide continuously updated priors for AMC edge likelihoods and MDP reward terms [6]. This data coupling ensures that learned defensive policies evolve alongside actual exploitation patterns, maintaining alignment between modeled risk and empirical threat trends. These directions collectively make the hybrid AMC–MDP framework more scalable, uncertainty-aware, and empirically grounded deployment within enterprise security operations, marking the next stage in transitioning from conceptual design to operational realization

VI. CONCLUSION

The evolution of network vulnerability assessment reveals a shift from static scoring to adaptive and stochastic decision frameworks. Static systems focused on severity scores to individual vulnerabilities, without considering how attack progresses step-by-step within a network; learning-based detectors react locally; probabilistic attack-graph models quantify longterm compromise but do not choose defenses [3], [23], [11], [13]. The proposed Hybrid AMC–MDP framework unifies these by combining AMC's interpretable quantification with MDP's long-horizon optimization, enriched by topology-aware state and reward design. Although this study is conceptual, it provides a rigorous foundation for empirical validation and a roadmap toward adaptive, interpretable, and economically optimized defense systems.

VII. ACKNOWLEDGMENT

The authors would also like to say their sincere thanks to Padmavathi S. at the Department of Computer Science and Engineering (Cybersecurity) at Dayananda Sagar College of Engineering, who has guided, advised, and supported them throughout the research period with the invaluable guidance, insightful advice, and untenting support.

REFERENCES

[1] B. Barchuk and K. Volkov, "Limitations of modern vulnerability scanners and CVE systems," World Journal of Advanced Engineering Technology and Sciences, vol. 12, pp. 973–989, 2024.

[2] V. Koscinski, D. Cromar, and A. Goucher, "Conflicting scores, confusing signals: An empirical study of vulnerability scoring systems," in Proc. 32nd ACM Conf. on Computer and Communications Security (CCS), 2025, pp. 1–15.

[3] M. Frigault and L. Wang, "Measuring network security using Bayesian networks," in Proc. IEEE 33rd Conf. on Local Computer Networks (LCN), 2008, pp. 901–908.

[4] N. Shimizu and M. Hashimoto, "Vulnerability management chaining: An integrated framework for efficient cybersecurity risk prioritization," arXiv preprint, arXiv:2506.01220, 2025.

[5] M. Tajuddin and C. Nandini, "Secured crypto biometric system using retina," International Advanced Research Journal in Science, Engineering and Technology (IARJSET), vol. 2, no. 1, pp. 28–33, 2015.

[6] J. Jacobs, A. Edwards, and M. Zmuda, "Improving vulnerability remediation through better exploit prediction," Journal of Cybersecurity, vol. 6, 2020.

[7] S. K. R. Akkemgari and B. Binny, "Intelligent security automation: Standard continuous process from scan to remediation for vulnerability management," in Proc. 6th Int. Conf. on ICICV, 2025, pp. 711–716.

[8] A. V. Turukmane and R. Devendiran, "M-MultiSVM: An efficient feature-selection-assisted network intrusion detection system using machine learning," Computers & Security, vol. 137, 103587, 2024.

[9] T. Yi, M. Xu, and K. Xu, "Review on the application of deep learning in network attack detection," Journal of Network and Computer Applications, vol. 212, 103580, 2023.

[10] D. V. S. Deepthi, "Attribute-policy homomorphic encryption with elliptic curve cryptography for MANET security," International Journal of Computer Science and Engineering, vol. 10, no. 5, pp. 112–118, 2022.

[11] D. Telsoc, "A threat computation model using a Markov chain and CVSS and its application to cloud security," Journal of Telecommunications and the Digital Economy, vol. 7, no. 1, 2017.

[12] T. Xu, F. Luo, and J. Li, "Markov chain-based modeling of attack propagation in enterprise networks," Information Sciences, vol. 647, 118021, 2024.

[13] Z. Liu, J. Chen, and Y. Zhao, "Network security situation assessment method based on absorbing Markov chain," in Proc. Intl. Conf. on Information Technology and Applications, 2022.

[14] A. Bose, R. Paruchuri, and S. Kumar, "Factored MDP approach to moving target defense," arXiv preprint, arXiv:2408.08934, 2024.

[15] R. Devendiran and A. V. Turukmane, "Dugar-LSTM: Deep learning-based network intrusion detection using chaotic optimization," Expert Systems with Applications, vol. 245, 123027, 2024.

[16] K. Roshan, S. Jain, and M. Maheshwari, "Untargeted white-box adversarial attack with heuristic defence in real-time deep-learning-based intrusion detection systems," Computer Communications, vol. 218, pp. 97–113, 2024.

[17] M. Maddu and Y. N. Rao, "Network intrusion detection and mitigation in SDN using deep learning models," International Journal of Information Security, vol. 23, pp. 849–862, 2024.

[18] H. M. Saleh, S. E. Ismail, and Y. A. Ramadan, "Stochastic gradient descent intrusion detection for wireless sensor network attack detection systems using machine learning," IEEE Access, vol. 12, pp. 3825–3836, 2024.

[19] W. Wang, P. Zhao, and L. Wu, "Multi-stage network attack detection algorithm based on Gaussian mixture hidden Markov model and transfer learning," IEEE Trans. Automation Science and Engineering, 2024.

[20] A. Rahman, S. Hussain, and L. Karim, "Reinforcement learning for cyber operations: Applications of artificial intelligence for penetration testing," Preprint, 2025.

[21] J. Zeng, X. Liu, and R. Yang, "Survey of attack graph analysis methods," Security and Communication Networks, Article ID 2031063, 2019.

[22] S. Abraham and S. Nair, "Cybersecurity analytics: A stochastic model for security quantification," Journal of Communications, vol. 9, no. 12, pp. 947–955, 2014.

[23] F. Luo, T. Xu, J. Li, and F. Xu, "MDP-AD: A Markov decision process-based adaptive framework for real-time detection of evolving and unknown network attacks," Alexandria Engineering Journal, vol. 126, pp. 480–490, 2025.

[24] X. Liu, H. Zhang, and K. Xu, "Network defense decision-making based on a stochastic POMDP and DQN," Computers & Security, vol. 108, 102301, 2021.

[25] Y. Qian, X. Fu, and Q. Zhu, "Receding-horizon MDP for moving target defense," arXiv preprint, arXiv:2002.05146, 2020.

[26] V. V. Krishnan, "Generative AI for vulnerability management: A blueprint," Journal of Artificial Intelligence & Cloud Computing, vol. 3, pp. 1–4, 2024.

[27] J. Yang and Y. Yang, "Optimal security protection strategy selection based on Markov model attack graph," J. Phys.: Conf. Ser., vol. 2132, 012020, 2021.

[28] M. Sewak, P. Sethi, and M. Singh, "Deep reinforcement learning in advanced cybersecurity threat detection and protection," Information Systems Frontiers, vol. 25, pp. 589–611, 2023.

[29] H. Benaddi, S. El Fkihi, and A. Bourjij, "Robust enhancement of intrusion detection systems using deep reinforcement learning and stochastic game," IEEE Trans. Veh. Technol., vol. 71, pp. 11089–11102, 2022.

[30] S. E. Hashemi-Petroodi, H. Rezaei, and P. Ghasemi, "Markov decision process for multi-manned mixed-model assembly lines with walking workers," International Journal of Production Economics, vol. 255, 108661, 2023.

[31] S. Ramesh, R. Sinha, and M. Reddy, "Comparative analysis of Q-learning, SARSA, and deep Q-network for microgrid energy management," Scientific Reports, vol. 15, 694, 2025.

[32] K. Hu, L. Lin, and J. Liu, "A review of research on reinforcement learning algorithms for multi-agents," Neurocomputing, 128068, 2024.

[33] J. Yuan, T. Zhang, and W. Li, "Deep reinforcement learning-based energy consumption optimization for peer-to-peer communication in wireless sensor networks," Sensors, vol. 24, 1632, 2024.

[34] R. Kiran P., S. Khaiyum, C. M., and P. B., "Improvising safety and energy efficiency of IoT based networks data routing," International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), vol. 11, no. 10, pp. 831–837, Oct. 2023. doi:10.17762/ijritcc.v11i10.8599