LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 24, NO. 3 (2026)

SENTINEL-GUIDED INTELLIGENT CYBERSECURITY GOVERNANCE FOR RESILIENT DIGITAL PUBLIC SERVICES: AN ADAPTIVE RISK-BASED FRAMEWORK

G.Brinda[1], G.Jagajothi[2]

[1]Assisant Professor, Department of Electronics and Communication Engineering, M.P.Nachimuthu M.Jaganathan Engineering College, Chennimalai, Erode - 638112, Tamilnadu, India. Email: bringandhi@gmail.com
[2]Professor, Department of Electronics and Communication Engineering, V.S.B Engineering College. Karur – 639111, Tamilnadu, India. Email id: jagajothi.g@gmail.com
Corresponding Author: G.Brinda, Email: bringandhi@gmail.com

Abstract
Reliability and security for digital public services depend more now than ever upon robust network infrastructure as well as strong cyber-security management systems. Interconnected digital platforms are used by local government organizations for different purposes; however, there are issues like network failures and cyber attacks that can affect their operations as well as people's confidence in them. The paper presents a new model called SG-ACGF which is an adaptive cybersecurity governance framework for improving resiliency and responsibility within internet-based government services' infrastructure systems. Graph Attention Networks are combined with Deep Reinforcement Learning for evaluating network connectivity and adapting to changing threats. This is an intrusion detection system based on sensors that can identify anomalies as well as predictive models of future events so we can take preventive measures beforehand. The Explainability Risk Prioritization Module helps to assess severity levels of incidents as well as their importance to services so that they can be addressed by local governments more effectively. Governance is based on roles of responsibility and transparency in making decisions. This model would be evaluated by means of simulation studies as well as comparison to current ones based upon parameters such as packet delivery rate and response time for governance issues. In conclusion, our study aims to integrate an adaptive network security system with local government administration so that there is better resilience in providing services of citizens through technology.

Keywords are Digital Governance; Local Government; Cybersecurity Governance; Sentinel-Based Monitoring; Deep Reinforcement Learning; Public Service Resilience.

1. Introduction

Currently, digital platforms play an important part of helping citizens' needs by providing them access to information while improving communication between individuals. The local government uses it to maintain the confidentiality of information. As such, there are numerous challenges for maintaining a reliable system along with problems regarding its availability so that people can become more confident about developing confidence toward their governments. A security information system plays an important role in maintaining stability and safeguards for the latest technologies and their management. Cybersecurity includes risk identification and immediate response related to issues. Moreover, challenges faced by public sectors are about striking an equilibrium between quick responses to problems of transparency, confidentiality, and responsibility while treating cybersecurity from a management perspective instead of just being a technological problem.

2. Related Work

In recent years, machine learning-based intrusion detection has garnered a tremendous amount of interest since network traffic has behavioural patterns that can be leveraged to detect suspicious communication or activity from legitimate activity [1]. The supervised learning classification of Internet

of Things (IoT) intrusion detection has been explored, highlighting the significance of feature selection and the classification of attacks to the UNSW-NB15 dataset, which is both application- and transport-based [2]. However, techniques for enhancing the performance of machine learning based network intrusion detection in UNSW-NB15 have been explored and the impact of model configuration and evaluation techniques are investigated on the network intrusion detection classification performance [3]. Network Intrusion Detection System (NIDS) based on machine learning techniques has been presented, which is an attempt to achieve efficient intrusion detection system with reducing computation complexity and increasing the accuracy of the classification system [4]. In the realm of contemporary IoT intrusion detection, there has been a focus on addressing a variety of attack types and the shortcomings of traditional feature-based classification [5]. These approaches offer a good starting point for monitoring the presence of malicious traffic, however, metrics are not sufficient to say that alerts can be converted into service priorities for response or governing action. [6]

Meanwhile, Deep learning has been extended to ID and used to decode complex relationships in network observations, and graph-based techniques can represent the dependencies between the communicating devices and traffic entities [7]. Security intelligence for the Internet of Things has been reviewed that addresses solutions based on machine-learning and future research directions for interconnected environments [8]. A mobile ad hoc network (MANET) adaptive intelligent routing algorithm using deep reinforcement learning was proposed for communication reliability and routing adaptability in MANETs [9]. Since, relational learning is important for security monitoring, and the use of graph attention networks to detect intrusions in the IoT context has been explored [10]. GraphSAGE-based improvements for intrusion detection have been studied and explored using the different graph-learning mechanisms in the cybersecurity datasets [11]. The studies here demonstrate the possibility of adaptive learning and graph representations to analyse changing network behaviour [12]. Network-level detection and routing are not however the same as the requirements of public administration, like the ranking of services on the basis of criticality, explicit response accountability, and the documentation of incidents in the context of governance. [13]

The application of explainable threat detection in industrial networks has also been investigated to enhance transparency in operational systems and general IoT security research has focused on the problems of achieving effective security across IoT systems [14]. However, more recent research focusing on local government cyber security has explored the organisational elements linked to building up security capacity and interorganisational partnerships between local governments [15]. Such contributions set the tone for the significance of being prepared and coordinated at the institutional level. However, governance-related research and intrusion-detection technology research are focusing on different aspects of the problem, leaving a gap in evaluating the problem as a whole, linking the evidence of threats in the network with service availability, urgency of recovery, risk prioritisation or accountable handling of incidents.

3. Methodology

SG-ACGF is an approach for integrating smart network security with informed decisions about risks so that we can provide more robust digital public service systems. This is an example of a sentinel-based method for adapting routes within networks while also focusing on importance evaluation, priority assignment during incidents, and accountability measures specific towards government communication needs. The research incorporates various methodologies, including public intrusion-detection datasets, dynamic graph representations, sentinel-based threat identification, graph attention-based feature learning, deep reinforcement learning, temporal threat assessment, and governance-oriented response prioritization. In contrast to conventional network security approaches which concentrate exclusively upon optimizing routes for speed of data transfer as well as detecting intrusions from hackers, SG-ACGF identifies vulnerabilities based on how they could affect public services availability and government reaction timeframes. The development stage for this model involves phases like gathering information from sources to analyzing it; designing models that are subject matter experts can validate against empirical evidence through experiments before conducting an overall analysis on all aspects involved in developing them (Figure 1).

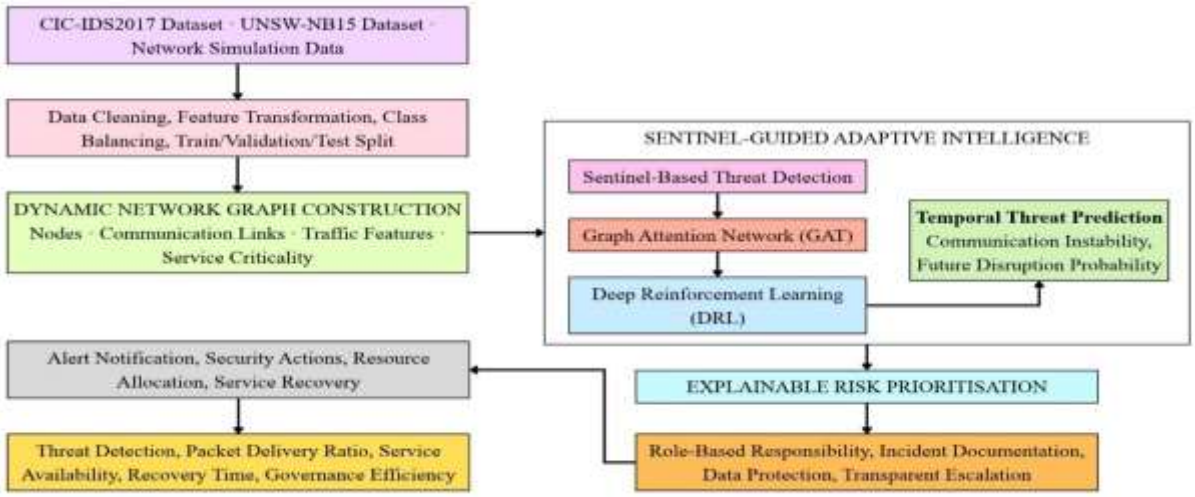


Figure 1. Overall Architecture of the Sentinel-Guided Adaptive Cybersecurity Governance Framework (SG-ACGF)

3.1. Data Collection and Dataset Characterisation

This paper uses open-source network attack datasets as a source of information about threat detection and security evaluation. CIC-IDS2017 from the Canadian Institute for Cybersecurity is an important source of data that contains labeled benign and malicious network flow information, traffic characteristics, and different types of attacks. Attacks can be carried out by various means such as DoS, DDoS, brute force; web based attacks; infiltration of a system or network through bots' activities. This data set may help to assess a level-of-threat detection system; however, it does not represent an organization's internal processes or activities.

For evaluating its generalization capability on another dataset, we use UNSW-NB15 as an external benchmark. Non-threats to traffic are present along with different types like espionage, exploitation, DoS attack, backdoor access and fuzz testing. Unique traffic generation process and features allow for verification of a detection model's generalization on other benchmarks.

Instead of being just exchanged for data files, they would be merged together on a feature and risk assessment basis. The local government communication node, public service application, emergency response unit, and sensor device is to be modeled as an interlinked system for modeling purposes through a simulated environment. The following governance-related parameters do not serve as actual measurements for these intrusions; they would instead be created by means of an explicit model. Service criticality, communication topology, mobility, and recovery needs would be defined by means of an explicit model attribute because these parameters are specific to a system rather than being represented directly on its surface within both sets of attack data files for this study. The separation prevents false claims of being real municipal cyber security information from these sources (Figure 2).

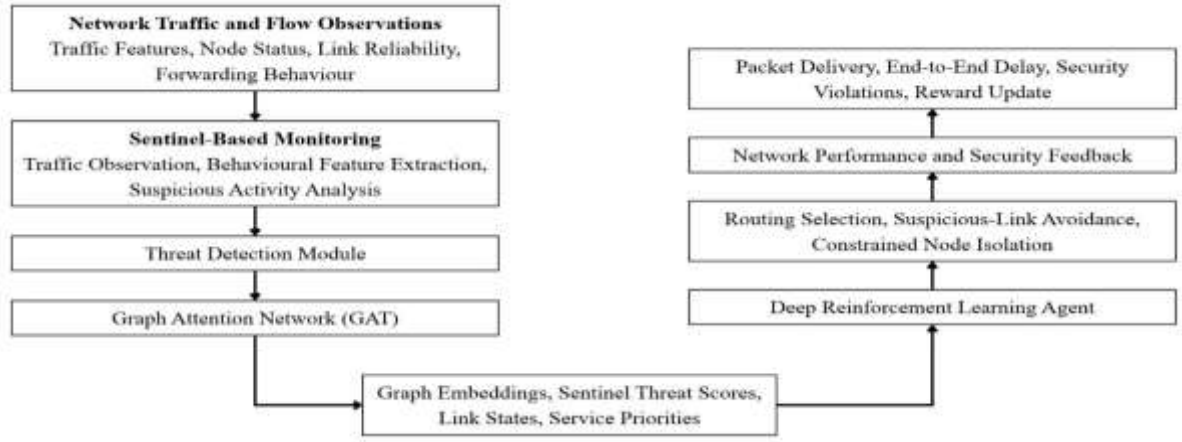


Figure 2. Sentinel-Based Threat Detection and Adaptive Network Decision Workflow

3.2. Data Preprocessing and Network Representation

3.2.1. Data Cleaning, Feature Transformation and Class Balancing

The statistics used for transformations only on the training partition that contains numerical features. If a feature (x_j) is numerical, it is normalised by min–max normalisation, which is applied as (Eq 1)

$$\tilde{x}_{ij} = (x_{ij} - x_j^{min}) / (x_j^{max} - x_j^{min} + \varepsilon) \quad (1)$$

where x_{ij} denotes feature j of observation i, x_j^{min} and x_j^{max} are the minimum and maximum values estimated from training data, and ε is a small positive constant that prevents division by zero.

3.3. Sentinel-Guided Adaptive Cybersecurity Governance Framework

The proposed SG-ACGF is comprised of four technical elements: sentinel-based threat assessment, graph attention-based representation learning, reinforcement-learning-based adaptive mitigation, and temporal risk-informed response. The graph attention mechanism is used to introduce the relationship between the nodes that are connected to each other, and the sentinel module extracts the behavioural evidence from the observed network. A reinforcement-learning agent then chooses which routing or mitigation actions to take based on the state of the network and the network's services. Temporal prediction provides an estimation of the communication instability of the near future, while the governance layer will convert technical evidence into prioritized and explained recommendations for communicable response.

3.3.1. Sentinel-Based Threat Detection and Graph Attention

The sentinel module processes observed traffic-flow features and produces a threat score for each monitored flow or entity. Let x_{i,t} represent the feature vector of entity i at time t. A trainable detector f_θ maps the observed features to a probability of malicious activity (Eq 2):

$$p_{i,t} = \sigma(f_\theta(x_{i,t})) \quad (2)$$

where p_{i,t} ∈ [0,1] denotes the estimated threat probability, f_θ is the learned detection function, θ represents its trainable parameters, and σ(·) is the sigmoid activation for binary threat classification. For multiclass attack identification, the output layer will instead use a softmax function.

To incorporate network context, a Graph Attention Network aggregates information from neighbouring nodes. Let h_{i,t} denote the feature representation of node i. The attention coefficient assigned to neighbour j is computed as (Eq 3)

$$\alpha_{ij,t} = exp(LeakyReLU[a^T (W \cdot h_{i,t} \,||\, W \cdot h_{j,t})]) / \Sigma_{k \in \mathcal{N}_i(t)} exp(LeakyReLU[a^T (W \cdot h_{i,t} \,||\, W \cdot h_{k,t})]) \quad (3)$$

where $\mathcal{N}_i(t)$ is the neighbourhood of node i at time t, W is a trainable feature-transformation matrix, a is the attention vector, and || denotes vector concatenation. The coefficients α_{ij,t} quantify the relative contribution of neighbouring nodes.

3.3.2. Temporal Threat Prediction and Service-Critical Risk Prioritisation

A temporal prediction component will estimate short-term communication instability from sequences of graph states, traffic statistics, link reliability, and sentinel outputs. Let z_{i,t} denote the learned temporal representation for node i. The probability of an impending disruption is modelled as (Eq 4)

$$\hat{d}_{i,t+\Delta} = \sigma(w_d^T z_{i,t} + b_d) \quad (4)$$

where $\hat{d}_{i,t+\Delta}$ is the estimated disruption probability over prediction horizon Δ, w_d is a trainable output vector, and b_d is the bias term.

The predicted disruption probability, sentinel threat score, service criticality, and recovery urgency are then combined to generate a governance-oriented response priority (Eq 5):

$$R_{i,t} = w_1 \cdot p_{i,t} + w_2 \cdot \hat{d}_{i,t+\Delta} + w_3 \cdot c_i + w_4 \cdot u_{i,t}, \quad \sum_{j=1}^{4} w_j = 1, \; w_j \geq 0 \quad (5)$$

where R_{i,t} is the composite priority score, p_{i,t} is the estimated threat probability, $\hat{d}_{i,t+\Delta}$ is the predicted disruption probability, c_i is the normalised service-criticality score, and u_{i,t} is the normalised recovery-urgency score. The weights $w_1,\ldots,w_4$ will be predefined or calibrated using validation data and sensitivity analysis. The four inputs must be normalised to comparable scales before aggregation.

The priority score supports ranking and response selection; it should not be interpreted as a calibrated probability of harm.

3.3.3. Explainable Governance Decisions and Accountable Incident Response

The recommendations detected will contain the threat category, level of confidence, the network component affected, the expected service impact, the priority, the recommended mitigation, and the escalation reason. The model results will be explained with a consistent explanation procedure and features observed in the input will be explained. If feature-attribution methods are used then the stability of the methods will be explored in repeated observations, rather than as a one shot piece of evidence explaining causality. Automated network action will be distinguished from choices that need to be approved by the administrator, based on the role of the user. Incident records will document timescaped, evidence which caused the incident and/or action taken, as well as policy rules used and consequences of the incident. These records are useful in ensuring that incidents can be reviewed in the future, and that they can be audited (Figure 3).

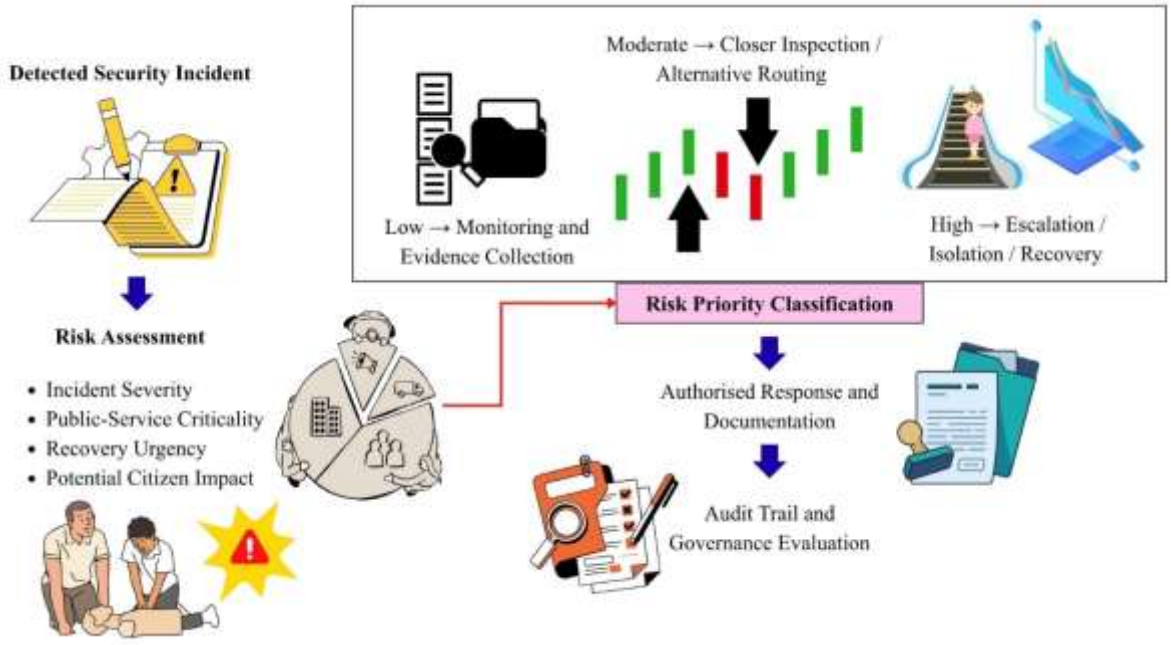


Figure 3. Governance-Based Cybersecurity Risk Prioritisation and Accountable Incident Response

The system will use non-identifying node identifiers, restrict access to incident information to roles and will not add unnecessary personal data to the threat detection pipeline. The measures established a testable governance design and do not (when implemented in isolation) show greater public trust, nor legal compliance.

3.4. Experimental Validation and Comparative Benchmarking

All the simulation environment will attempt to simulate various traffic loads, various sizes of network, various mobility, network failures and different intensities of attacks. The following will be documented uniformly for each of the approaches compared to ensure consistency: Benign & malicious traffic scenarios, Network configuration, random seeds, observation time intervals, and simulation duration. Pre-development(s): Divide the main data into training, validation and test subsets of data.
This framework compared to an existing route management techniques like conventional ML intrusion detection models, non-reinforcement based graph algorithms, current Sentinels' guided routes that are easy-to-use for users. Evaluation would include comparing all models through an ablative analysis where we disable parts like a sentinel module, graph attention mechanism for time series forecasting task, temporal predictor model and governance risk prioritization system so that it helps us understand how each component affects performance metrics. Data partition consistency is to maintain consistent data partitions, attack scenario and parameter selection procedure during testing process. Several simulations are performed to summarize data by mean values and standard deviations of each simulation run. The confidence interval should be used if necessary for assessing the statistical significance of observed improvement compared with random variation between trials.

3.5. Performance Indicators and Robustness Assessment

The performance metrics for threat detection will be precision, recall, F1-score, false positive rate and detection results for attack classes. These metrics are required because a model that has a high overall accuracy can be highly inaccurate on rare but significant attacks. Thus, along with the aggregate results, you will get the results and confusion matrices of each class. Governance effectiveness will be assessed using a variety of response timings, prioritisation regularity, compliance with policies and audit trail completeness. The time to incident response will be a measure that begins when the first successful detection is made and ends when a preconfigured action to respond to the incident is complete. Consistency in prioritisation will be reviewed by looking at how the same scenarios are assigned the same risk ranking when using the same policy settings; audit-trail completeness will be assessed by the number of fields in incident-record forms that are captured in the incident-response process. The set of robustness experiments will use different node mobility settings, traffic loads, attack rate, failure rates of links and different service criticality assignments. Sensitivity analyses will be used to address the impact of changes to risk weights and thresholds for responses. The threat detector will be evaluated against Cross-dataset evaluation to test the generalisation of the threat detector from its primary training source, and the adaptive routing and governance parts will be assessed with simulation-based tests. The analyses will aid in distinguishing increased security governance practices, from the influence of a specific dataset or network configuration.

4. Results and Discussion

4.1. Comparative Performance of Cybersecurity Detection Models

As shown in Figure 4, the best obtained results were the low false-positive rate (1.89%), and high accuracy (97.36%), precision (97.12%) and recall (96.84%) and F1-score (96.98%) for the proposed SG-ACGF. The Graph Attention Network performed the second-best with 96.41% accuracy and a 2.74% false positive rate, while the LSTM based detector scores 95.83%. The CNN model achieved an accuracy of 95.16%, the Random Forest model 94.28% and the SVM model 91.65%. The accuracy increased by 0.95 percentage points and the false-positive rate decreased by 0.85 percentage points when compared with the Graph Attention Network. Overall, the effectiveness of the SG-ACGF was shown to provide more accurate threat identification with less false alarms and a comparable level of classification, relative to the baseline models.

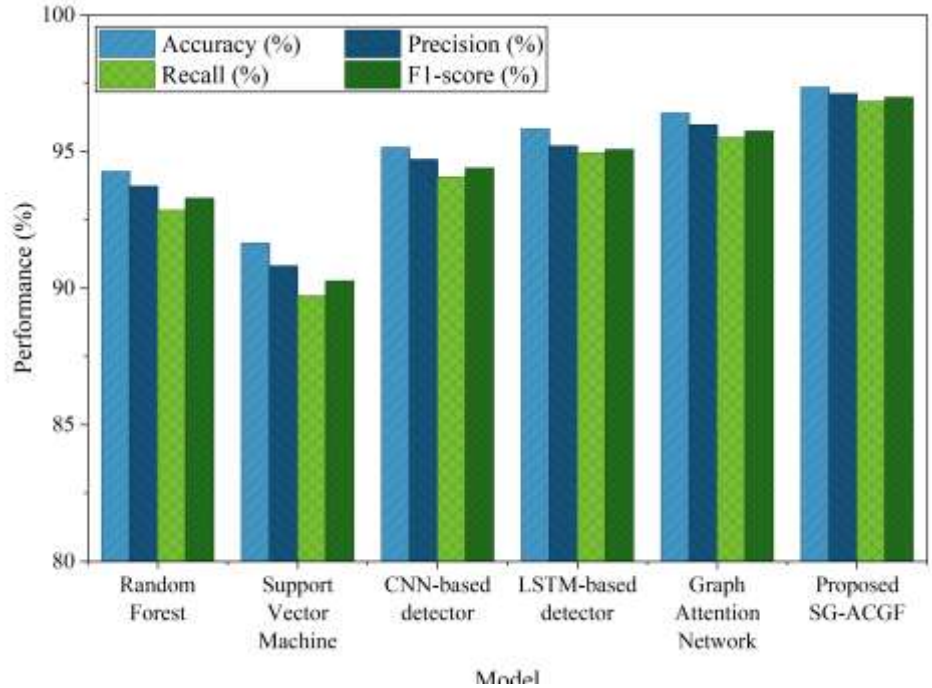


Figure 4. Comparative Threat-Detection Performance of SG-ACGF and Baseline Models

4.2. Attack-Specific Detection Effectiveness

Table 1 showed the proposed framework successfully classified the network traffic in all the attack scenarios tested with the highest F1-score values of 98.28% for benign traffic and 98.02% for the attack category DDoS. Denial-of-service (DoS) and Brute-force attacks scored F1 scores of 97.31% and 96.39%, respectively, whereas the remaining attacks performing the representation scored an F1 of 95.22%. The attacks from the infiltration category had the lowest F1-score (94.17%) meaning that they were comparatively more difficult to detect.

Table 1. Attack-Specific Detection Performance of the Proposed Framework

Attack category	Test samples	Precision (%)	Recall (%)	F1-score (%)
Benign traffic	18,000	98.12	98.45	98.28
Denial-of-service	5,200	97.68	96.94	97.31
Distributed denial-of-service	6,400	98.21	97.83	98.02
Brute-force attacks	3,100	96.87	95.92	96.39
Infiltration	1,200	94.76	93.58	94.17
Other represented attacks	2,100	95.83	94.61	95.22
Macro average	36,000	96.91	96.22	96.57
Weighted average	36,000	97.42	97.36	97.39

4.3. Comparative Network Communication Performance

The results in Table 2 demonstrated that the proposed SG-ACGF yielded the best packet delivery ratio of 96.24%, the lowest end-to-end delay of 43.76 ms, the lowest routing overhead of 13.58%, energy consumption of 13.84 J. On the original routing method based on the sentinel, the packet delivery ratio was 94.16 % and the end-to-end delay was 55.82 ms, while the conventional deep reinforcement learning routing could deliver the packet delivery ratio of 92.68 %. AODV, DSR and OLSR, compared to them, had small packet delivery ratios, high delays, and high routing overhead. The results showed the effectiveness of the SG-ACGF in improving packet delivery, decreasing end-to-end delay, routing overhead, and energy consumption under controlled network conditions compared to original sentinel-guided routing method.

Table 2. Comparative Communication Performance Under Controlled Network Conditions

Routing method	Packet delivery ratio (%)	End-to-end delay (ms)	Routing overhead (%)	Energy consumed (J)
AODV	87.42	82.64	24.81	18.72
DSR	85.76	91.38	27.43	19.65
OLSR	89.31	76.52	22.67	17.94
Conventional DRL routing	92.68	63.47	18.92	16.38
Original sentinel-guided routing	94.16	55.82	16.47	15.21
Proposed SG-ACGF	96.24	43.76	13.58	13.84

4.4. Resilience Under Varying Attack Intensities

Figure 5 shows that through the attacks as they escalated from 5% to 30%, the network resilience or performance provided by the proposed SG-ACGF was comparatively better than that provided by the conventional routing. Under the attack intensity of 5%, SG-ACGF reached 97.16%, 96.82%, and 98.14% packet delivery ratio, detection recall, and service availability, respectively, with the recovery time of 3.26s. A 30% attack intensity achieved an 89.64% packet delivery, 92.31% detection recall, and 91.76% service availability, and reduced recovery time to 9.24 s. Contrary, the delivery of packets and the service availability were significantly degraded in conventional routing, while the recovery time was increased to 28.47 s. The findings validated that SG-ACGF supports better communication, effective attack detection and quicker network recovery rate within increasing attack magnitude.

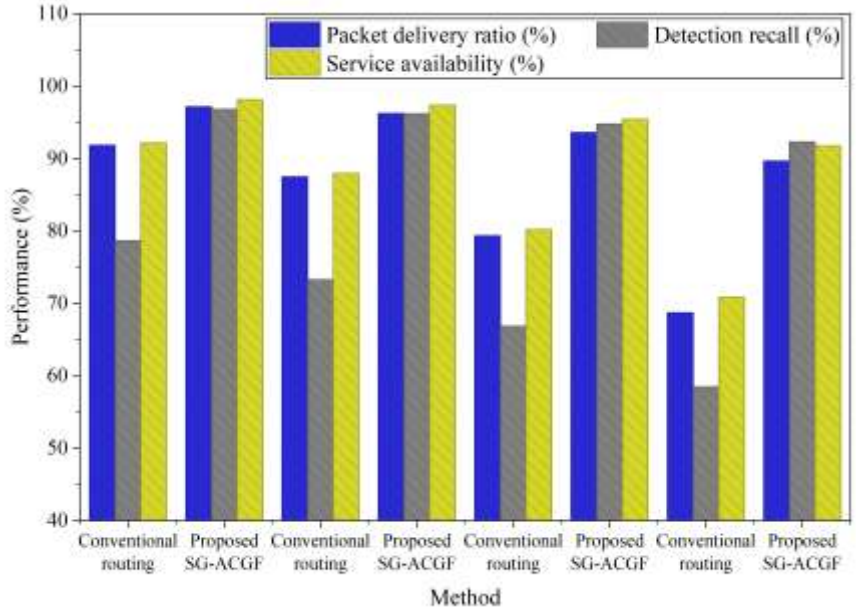


Figure 5. Network Resilience Under Increasing Attack Intensity

4.5. Ablation Analysis of Framework Components

Table 3 shows that the full SG-ACGF performed the best on all metrics, with a detection F1-score of 96.98%, a packet delivery rate of 96.24%, availability of the service of 97.36% and a recovery time of 4.12 s. If the monitoring is not considered, the F1 score falls to 92.46% and the recovery time to 11.38 s, and if the graph attention is not considered, the F1 score decreases to 93.82%. Communication deterioration and longer recovery times occurred with the deletion of temporal prediction and the adaptive reinforcement learning. In the same way, if governance risk prioritisation is not considered, recovery time increased by 8.63 s and the service availability decreased to 93.82%. On average, the different components also resulted in a detection accuracy, communication reliability, service continuity, and efficiency of recovery that confirmed the contribution of each component to the overall

properties of the framework, with the greatest performance found for the combination of the different components.

Table 3. Component-Wise Ablation Analysis of SG-ACGF

Framework configuration	Detection F1-score (%)	Packet delivery ratio (%)	Service availability (%)	Recovery time (s)
Full SG-ACGF	96.98	96.24	97.36	4.12
Without sentinel monitoring	92.46	88.72	89.64	11.38
Without graph attention	93.82	90.16	91.28	9.72
Without temporal prediction	95.14	92.48	93.16	7.84
Without governance risk prioritisation	96.74	95.86	93.82	8.63
Without adaptive reinforcement learning	95.62	87.94	89.72	12.46

4.6 Temporal Threat Prediction Performance

According to Figure 6, our proposed SG-ACGF model achieved a high accuracy rate of 97.08%, recall rate of 96.42% and F1-score of 96.75% while minimizing the maximum mean absolute error (MAE) at 0.049 and root mean square error (RMSE) at 0.091 when compared with other temporal threat prediction models studied for this purpose. The GAT-LSTM model ranks second among all models with an F1 score of 94.72%, whereas Temporal Transformer model and LSTM model are ranked third with scores 93.95% and 92.21%. Both Random Forest and Support Vector Machine had an F1-score less than 88.63% and 87.44%, respectively. Although the F1-score and two types of prediction errors are lower for GAT-LSTM, there is a difference of 2.03 percent between them. In general, this study found a reduction of time-to-threat prediction error rate as well as higher accuracy for SG-ACGF compared to baseline model performance.

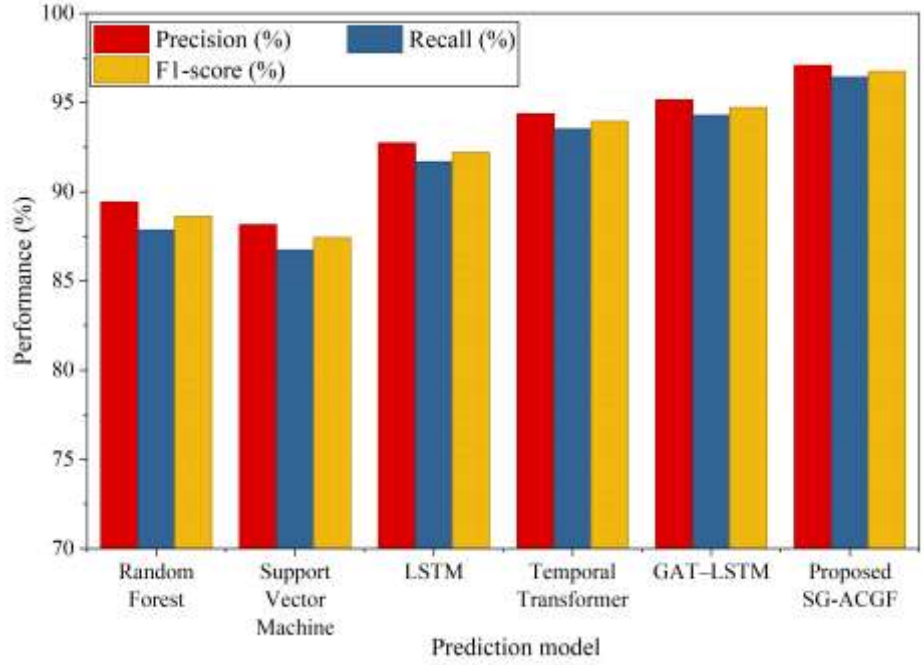


Figure 6. Comparative Evaluation of Temporal Threat Prediction Models

4.7 Public-Service Availability under Variable Network Loads

Table 4 showed that when making use of different network loads, the availability of public services and packet delivery ratio (PDR) increased in the proposed SG-ACGF, compared to the conventional approach. We observed 100 packets/s with SG-ACGF, resulting in 99.12% service availability, 97.46% PDR, a delay of 34.28 ms and 0.88% failed requests. At 750 packets/s it maintained 93.46% service availability and had an average PDR of 89.72%, with an average delay of 64.18ms and failed requests 6.54%. On the other hand, the traditional method had a decrease in the ability of services to 81.64% and delay had increased to 112.47ms for the peak load. The findings overall showed that SG-ACGF enhanced the continuity of services, the reliability of communications, and the level of effectiveness of the responses given under heavy network traffic.

Table 4. Public-Service Availability under Different Network-Load Conditions

Network load (packets/s)	Method	Service availability (%)	PDR (%)	Delay (ms)	Failed requests (%)
100	Conventional approach	96.42	93.18	48.62	3.58
100	Proposed SG-ACGF	99.12	97.46	34.28	0.88
250	Conventional approach	93.76	89.64	61.84	6.24
250	Proposed SG-ACGF	98.24	96.18	39.76	1.76
500	Conventional approach	88.53	83.72	84.36	11.47
500	Proposed SG-ACGF	96.82	93.64	48.53	3.18
750	Conventional approach	81.64	76.38	112.47	18.36
750	Proposed SG-ACGF	93.46	89.72	64.18	6.54

4.8 Governance-Based Cybersecurity Risk Prioritisation

Figure 7 showed that the proposed SG-ACGF had better Precision@5 (96.24%), Recall@10 (94.72%), NDCG@10 (96.83%) and MRR (0.961%) compared to all the governance-based risk prioritisation methods. The second-best performance was done by Risk-weighted machine learning which posted 91.38% Precision@5 and 92.24% NDCG@10, followed by the TOPSIS- and AHP-based prioritisation.

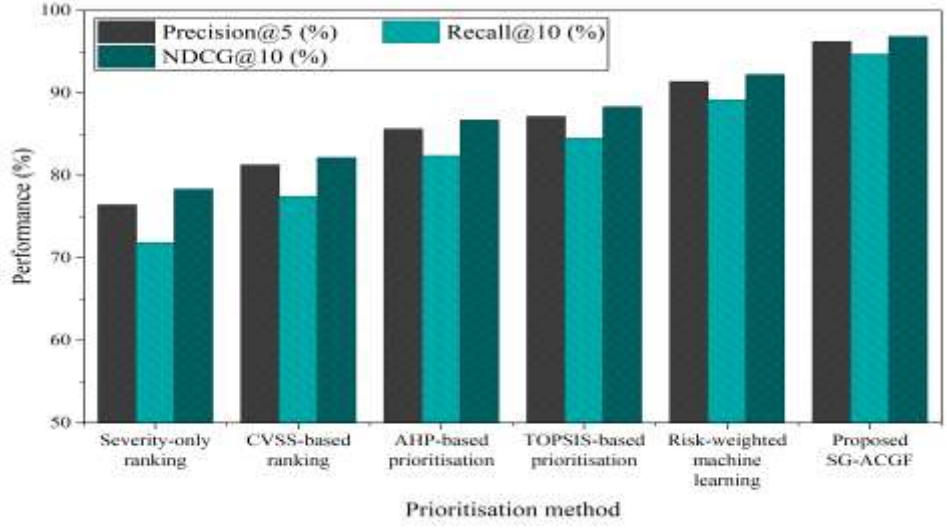


Figure 7. Comparative Performance of Governance-Based Risk Prioritisation Methods

6. Conclusion

The proposed SG-ACGF model that would connect them so there are no gaps when it comes to detecting intrusions on networks through graphs-based methods used at times to identify threat vectors along with their severity level during attacks by hackers or other malicious actors. The paper proposes an evaluation method that is based upon factors such as detection accuracy percentage, types of attacks identified by them, packet throughput time, system availability duration, cost per item produced and explanation scores along with rules governing operations. The amino acid values listed here are just a starting point for comparing these models with those that use machine learning algorithms or neural networks because there is still much work left before we can say they're better than others currently available. Validation models could be helpful for studies of cyber security management because they help identify elements contributing towards risk identification which are integrated into priority setting, transparency and practicability while making decisions regarding electronic government service provision. Validation must also be done at intervals between tests for verification purposes through an error margin analysis before applying it on real-world measurements while operating a testing device due to limitations such as resource availability or other factors like environmental conditions which may affect accuracy. The research may involve participation by local authorities during development such as data security measures; an evaluation mechanism for outcomes; applying “concept drift” theories on algorithms used by government agencies; and adherence to laws regarding duties of representatives towards AI suggestions.

Statement Declarations

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Availability of data and materials

Not applicable

References

1. Ahmad, M., Riaz, Q., Zeeshan, M., Tahir, H., Haider, S. A., et al. (2021). Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set. EURASIP Journal on Wireless Communications and Networking, 2021, Article 10. https://doi.org/10.1186/s13638-021-01893-8

2. Moualla, S., Khorzom, K., & Jafar, A. (2021). Improving the performance of machine learning-based network intrusion detection systems on the UNSW-NB15 dataset. Computational Intelligence and Neuroscience, 2021, Article 5557577. https://doi.org/10.1155/2021/5557577

3. Rincy, N., et al. (2021). Design and development of an efficient network intrusion detection system using machine learning techniques. Wireless Communications and Mobile Computing, 2021, Article 9974270. https://doi.org/10.1155/2021/9974270

4. Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset. (2021). IEEE Access, 9, 22351–22370. https://doi.org/10.1109/ACCESS.2021.3056614

5. Sarker, I. H., Khan, A. I., Abushark, Y. B., & Alsolami, F. (2023). Internet of things (IoT) security intelligence: A comprehensive overview, machine learning solutions and research directions. Mobile Networks and Applications, 28, 296–312. https://doi.org/10.1007/s11036-022-01937-3

6. Bai, J., Sun, J., Wang, Z., Zhao, X., Wen, A., Zhang, C., & Zhang, J. (2024). An adaptive intelligent routing algorithm based on deep reinforcement learning. Computer Communications, 216, 195–208. https://doi.org/10.1016/j.comcom.2023.12.039

7. A survey on explainable artificial intelligence for cybersecurity. (2023). IEEE Transactions on Network and Service Management, 20(4), 5115–5140. https://doi.org/10.1109/TNSM.2023.3282740

8. Hossain, S. T., Yigitcanlar, T., Nguyen, K., & Xu, Y. (2025). Cybersecurity in local governments: A systematic review and framework of key challenges. Urban Governance, 5(1), 1–19. https://doi.org/10.1016/j.ugj.2024.12.010

9. Mohale, V. Z., & Obagbuwa, I. C. (2025). A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. Frontiers in Artificial Intelligence, 8, Article 1526221. https://doi.org/10.3389/frai.2025.1526221
10. Ahanger, A. S., Khan, S. M., Masoodi, F., & Salau, A. O. (2025). Advanced intrusion detection in internet of things using graph attention networks. Scientific Reports, 15, Article 9831. https://doi.org/10.1038/s41598-025-94624-8
11. Saidane, S., Telch, F., Shahin, K., & Granelli, F. (2025). Deep GraphSAGE enhancements for intrusion detection: Analyzing attention mechanisms and GCN integration. Journal of Information Security and Applications, 90, Article 104013. https://doi.org/10.1016/j.jisa.2025.104013
12. Shoukat, S., Gao, T., Javeed, D., Saeed, M. S., & Adil, M. (2025). Trust my IDS: An explainable AI integrated deep learning-based transparent threat detection system for industrial networks. Computers & Security, 149, Article 104191. https://doi.org/10.1016/j.cose.2024.104191
13. Selem, M., Jemili, F., & Korbaa, O. (2025). Deep learning for intrusion detection in IoT networks. Peer-to-Peer Networking and Applications, 18(2). https://doi.org/10.1007/s12083-024-01819-3
14. Popoola, S. I., Tsado, Y., Ogunjinmi, A. A., Sanchez-Velazquez, E., Peng, Y., & Rawat, D. B. (2025). Multi-stage deep learning for intrusion detection in industrial internet of things. IEEE Access, 13, 60532–60555. https://doi.org/10.1109/ACCESS.2025.3557959
15. Gupta, R., Gupta, S. K., Singh, S. B., Kumar, R., Muthukkumar, R., et al. (2026). Adaptive deep reinforcement learning–based secure routing for wireless sensor networks. Scientific Reports. https://doi.org/10.1038/s41598-026-62519-x