

Algebraic topology methods in network security

D. Ganesh [‡]

*Department of Computer Science and Engineering
Vardhaman College of Engineering
Hyderabad 501218
Telangana
India*

M. Sudharsan [†]

*Presidency School of Information Science
Presidency University
Bengaluru 560119
Karnataka
India*

I. Katharaj [§]

*Department of Mechanical Engineering
Faculty of Engineering and Technology
JAIN (Deemed-to-be University)
Bangalore 562112
Karnataka
India*

Samir N. Ajani ^{*}

*School of Computer Science and Engineering
Ramdeobaba University (RBU)
Nagpur 440013
Maharashtra
India*

[‡] E-mail: dganesh@outlook.com

[†] E-mail: sudharsan.m@presidencyuniversity.in

[§] E-mail: kantharaj@jainuniversity.ac.in

^{*} E-mail: samir.ajani@gmail.com (Corresponding Author)

Saksham Sood [@]

Centre of Research Impact and Outcome (CRIO)
Chitkara University
Rajpura 140417
Punjab
India

Rupesh Mishra [#]

Department of Computer Science and Engineering
Rungta College of Engineering and Technology
Rungta International Skills University
Bhilai 490024
Chhattisgarh
India

Abstract

Attack surfaces are becoming dynamic, communications are increasingly encrypted and coordinated attacks are growing in complexity, creating more challenges for modern network security; traditional graph-based security solutions struggle to see the big picture and understand relationships among entities in the network. To secure the network an effective intrusion detection system, the authors propose a new model for Network Security-Persistent Homology-Based Network Security Model (PHNSM) which combines simplicial complex construction with homology computation, persistent homology, topological feature extraction, and graph neural networks-based threat classification. The proposed model is a multidimensional communication pattern, which is represented algebraically to be used for the effective identification of stealthy and distributed cyber-attacks in the global network, while maintaining the entire network function. Experimental evaluation is able to attain detection accuracy of 94.67%, precision of 94.18%, recall of 93.86%, F1-score of 94.02% and ROC-AUC of 96.08% and a 18.3% reduction in false-positive rate when compared to conventional graph-based security methods. The novelty is that the integration of persistent homology and deep topological learning for a cyber threat analysis with interpretation. The proposed framework provides mathematically sound, scalable and robust approach to anomaly detection, localization of attacks and large scale communication networks.

Subject Classification: 55M20.

Keywords: Algebraic topology, Persistent homology, Network security, Simplicial complexes, Graph neural networks, Intrusion detection.

[@] E-mail: saksham.sood.orp@chitkara.edu.in

[#] E-mail: drupesh.mishra@rungta.org

I. Introduction

Modern computer networks are increasingly complex, interdependent systems and the perimeter security approach is no longer effective to defend against the complex, distributed and stealthy nature of modern cyber-attacks. While graph-theoretic approaches have been the foundation of network security analytics for a long time they are mostly used to represent pairwise relations and rarely account for higher-order relations, which can occur when many nodes interact at the same time. Recent studies [1] have illustrated that the algebraic topology approach to these multi-node interactions provides a rich mathematical framework that can be applied to the study of changing geometry of the embedding spaces for neural networks, for example. Likewise, the topological descriptors have been used to identify structural patterns associated with the behavior of cyber attacks, which are not detectable from traditional metrics [2]. Algebraic graph structures have also been suggested for the security of distributed networks and cryptographic protocols, highlighting the wide range of applications of algebraic techniques in secure areas [3]. Advanced computing environments in large scale data analysis have been studied with the use of algebraic topology which has scalability benefits [4]. The use of the topological data analysis methods for anomaly detection tasks is further motivated by the fundamental studies that establish a connection between the two areas of computer science and topological data analysis [5].

Algebraic topology has also been combined with game-theoretic decision optimization for modelling dynamics across multilayer complex networks at the network level [6] and algebraic models have been applied to model redundant ring topologies in edge data centres [7]. However, similar advances in self-contained cyber-defense highlight the increasing demand for AI-based systems to be able to counter AI-based advanced persistent threats [8]. Algebraic structures have also been employed to describe malware insertion and detection procedures, providing a formal perspective towards adversarial procedures [9]. Last, but not least, quantitative approaches for quantifying network topology using mathematical models are introduced in global network security and confirm the relevance of structural invariants in global network security analytics [10].

Although these developments have taken place, the existing approaches lack the ability to combine both persistent homology and deep representation learning to perform classification of threats in an end-to-end manner and most current IDSs are limited to considering attack behavior at the level of a single host pair [11], [12]. In this paper, we fill this gap by introducing the Persistent Homology-Based Network Security Model (PHNSM) that consists of three key components: (i) a simplicial-complex-based network representation that extends to higher-order communication

structure, (ii) a persistent-homology guided topological feature extraction pipeline, and (iii) a graph neural network classification stage, all designed towards improving the detection accuracy, minimizing false positives, and providing interpretable, mathematically grounded intrusion detection for a large-scale and dynamic communication network.

II. Persistent Homology-Based Network Security Model (PHNSM)

The way PHNSM works is to convert raw network traffic into a simplicial complex containing information not only about node-to-node communications, but also the interactions between groups of multiple nodes. This complex is created using a filtration process in order to increase the similarity threshold, and the persistent homology captures the birth and death of topological features, like connected components, loops and voids.

A. Algebraic Topology Foundations

In algebraic topology, one approaches the study of networks by encoding more or less complex entities as vertices, edges, triangles and higher dimensional "similar" entities as a network of simplicial complexes. These simplices that are at the boundaries of the chain groups are mapped by boundary operators and the kernels and images of these operators provide homology groups of ranks, the Betti numbers, which quantify connected components, loops and voids, and thus uncover anomalies in the structure resulting from coordinated or distributed attacks.

Theorem: (Stability of Persistent Homology). *Let X be a point cloud. Let X be a point cloud. Suppose that $f, g: X \rightarrow \mathbb{R}$ are "tame" functions on a topological space X . Then their persistence diagram is encircled by the following pair of distances:*

$$dB(Dgm(f), Dgm(g)) \leq \|f - g\|_\infty$$

Proof (sketch): Let f and g be sublevel-set filtrations, respectively, on R , with $a \in R$. Every sublevel set of f is a subset of the corresponding δ -shifted sublevel set of g and conversely every sublevel set of g is a subset of the corresponding δ -shifted sublevel set of f , thus creating a δ -interleaving of the persistence modules $Hk(\{f \leq a\})$ and $Hk(\{g \leq a\})$. The algebraic stability lemma for interleaved persistence modules says that an ε -interleaving means that the induced persistence diagrams have a bottleneck distance of at most ε between each other. If we take $\varepsilon = \delta = \|f - g\|_\infty$ then we have the above bound.

$$K \subseteq 2^V \text{ such that } \sigma \in K, \tau \subseteq \sigma \Rightarrow \tau \in K \quad (1)$$

Describes what is meant by a simplicial complex: a family of faces of a simplicial complex is downward closed: it means a collection of subsets of vertices, where each subset represents a face of a lower dimension, is a face of a lower dimension.

$$C_k(K) = \bigoplus \mathbb{Z} \cdot \sigma, \sigma \in K, \dim \sigma = k \quad (2)$$

Represents the k -chain group as a free abelian group on all k dimensional simplices of the complex.

$$\partial_k[v_0, \dots, v_k] = \Sigma(-1)^i [v_1, \dots, \hat{v}_i, \dots, v_k] \quad (3)$$

The eq. (3) defines the Boundary operator that will remove a vertex one after another and downwards, alternating signs.

$$\partial_k \partial k + 1 = 0 \quad (4)$$

States the "twin boundary" property, that is, that twice the boundary is zero.

$$Z_k = \ker(\partial_k) \quad (5)$$

It describes k -cycles as chains with vanishing boundary, in terms of closed topological loops or surfaces.

$$B_k = \text{im}(\partial_{k+1}) \quad (6)$$

Defines k -boundaries as chains which are the boundaries of higher dimensional simplices in the complex.

$$H_k = \frac{Z_k}{B_k} \quad (7)$$

Computes the k -th homology group, as cycles with respect to the boundaries, which is equivalent to genuine topology holes.

$$\beta_k = \text{rank}(H_k) \quad (8)$$

Gives the meaning of the Betti number: the rank of the homology group (number of independent k -dimensional holes).

$$K_{\varepsilon_1} \subseteq K_{\varepsilon_2} \subseteq \dots \subseteq K_{\varepsilon_n} \quad (9)$$

Generates a hierarchy of filters of increasing distance range and follows the topologies.

$$\text{Wp}(\text{Dgm}(f), \text{Dgm}(g)) = \left(\inf_{\gamma} \sum \|x - \gamma(x)\| p^t \right)^{\frac{1}{p}} \quad (10)$$

Introduces the Wasserstein distance as a bijective distance to compare persistence diagrams that is a generalization of the bottleneck distance.

B. Network Representation and Persistent Homology

Hosts, servers and communication endpoints are represented in the network as vertices and the weights of the edges are based on similarity measures, which are calculated based on the volume of traffic, protocol behavior and temporal correlation. A Vietoris–Rips construction links vertices to within distance threshold ε , and higher-order simplices will form when groups of three or more vertices have the same distance threshold. As ε grows, the complex changes in nature, passing through a series of filtrations, and persistent homology captures the birth and death of the features of the complex, thus encoding them as persistence diagrams, which

summarize the evolving connectivity, cycles and higher-dimensional voids of the network.

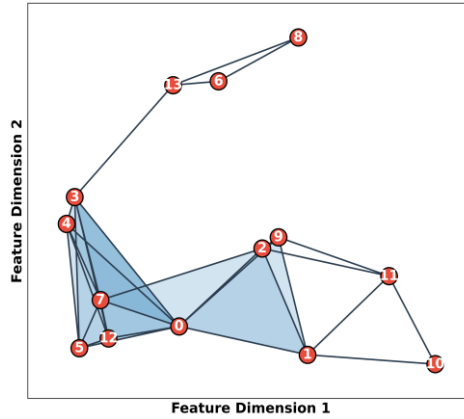


Figure 1

Simplicial complex representation of a communication network constructed from node feature similarity

The simplicial complex representation of a communication network is shown in Figure 1, where the nodes are represented by red vertices and the 2-simplices (by blue triangles) are formed by nodes triplets. Edges show direct pairwise communication, and shaded triangles represent higher order group interactions which are not expressible by traditional graph models. Overlapping clusters of triangles represent areas of dense communication, both possibly legitimate multiparty services and co-ordinated attacks, and highlight the need for distinguishing between legitimate density and coordinated attack structure by using persistent homology.

$$d(u, v) = \|X_u - X_v\| \quad (11)$$

It calculates pairwise Euclidean distance between node feature vectors to get simplicial connectivity thresholds.

$$VR_\varepsilon(V) = \{\sigma \subseteq V : d(u, v) \leq \varepsilon, \forall u, v \in \sigma\} \quad (12)$$

Constructs the Vietoris–Rips complex that is created by all subsets of nodes at distance within threshold ε from each other.

$$PD = \{(b_i, d_i)\} \quad (13)$$

Represents persistence diagram as a multiset of birth-death coordinate pairs of the topological features.

$$\text{pers}_i = d_i - b_i \quad (14)$$

Computes persistence lifetime difference between death and birth thresholds for feature.

$$L(t) = \max(0, \min(t - b_i, d_i - t)) \tag{15}$$

Combines the persistence landscape function that maps diagrams into stable functional representations suitable for learning, which are vectorizable.

C. Threat Detection Framework

The threat detection framework starts from raw network traffic ingestion to the construction of simplicial complexes, which is done based on the similarity between nodes. Then, the changes in the evolution of these features are followed through a whole series of increasing thresholds, with persistence diagrams the result. These diagrams are vectorized into fixed length (nonscale) topological feature vectors that contain Betti numbers, persistence lifetimes and landscape functions. The output of these features are then passed on to a graph neural network (GNN) classifier, which learns to classify benign traffic from malicious traffic for the final binary threat classification and anomaly confidence score for downstream response and mitigation.

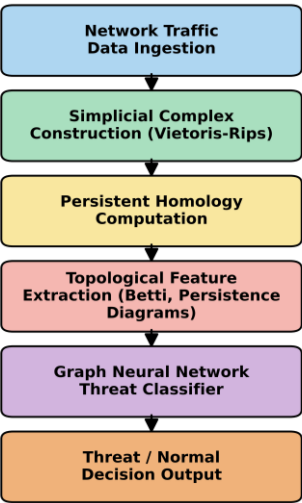


Figure 2
Six-stage working pipeline of the PHNSM threat detection framework

The pipeline of the PHNSM is shown in Figure 2, starting with the ingestion of data into the pipeline, followed by the construction of simplicial complexes, computation of persistent homology, extraction of topological features, classification using GNN, and output of the threat decision.

III. Experimental Methodology

A. Dataset and Experimental Setup

Experiments were performed on the benchmark intrusion detection datasets such as CICIDS2017 and NSL-KDD that contained labeled benign and malicious traffic of various attacks including DDoS, port scanning, botnet, and attacks. The network flows were preprocessed to feature vectors, normalized and were used to build Vietoris–Rips simplicial complexes at different filtration levels. The data set was divided into 70% training, 15% validation and 15% testing data sets. A Python-based topological data analysis libraries framework with a graph neural network framework was used for the experiments and trained on GPU-accelerated hardware.

B. Performance Evaluation Metrics

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (16)$$

Computes the total accuracy (rate) of classification of all the traffic instances for both the benign and malicious sets.

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (17)$$

Quantifies the proportion of predicted threats that are genuinely malicious, penalizing false positive classifications.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (18)$$

Calculates the percentage of the threats classified as malicious that were actually malicious, which means that false positives are penalized.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (19)$$

Capture the ratio of real threats that were correctly identified by the model indicating the sensitivity of the model for real threats.

$$\text{ROC} - \text{AUC} = \int \text{TPR} \, d(\text{FPR}) \quad (20)$$

Represents the area under the receiver operating characteristic curve, summarizing discrimination across all thresholds.

C. Comparative Methods

The four representative baselines used for benchmarking are within the standards set by PHNSM. Traditional flow based, statistical detection, non-structural is termed Support Vector Machine (SVM) classifiers. Random Forest models can deal with non-linear feature interactions but are not able to deal with the structure of the network. We also include a standard Graph Neural Network without topological features, to isolate the contribution of persistent homology and to only use raw adjacency-based message passing. Last, an Autoencoder-based intrusion detection system is an unsupervised reconstruction-error anomaly detection, which is frequently used in

pipelines in industry. All these baselines cover the different paradigms of statistical, ensemble, graph based and unsupervised.

IV. Results and Discussion

The proposed PHNSM was tested on the baseline intrusion detection techniques on a variety of classification scores and attack categories. Overall, PHNSM was shown to be the most successful method when compared with all of the other methods, with the best accuracy, precision, recall, and ROC-AUC scores and a significant reduction of the false-positive rate. There was little degradation in the detection reliability when performance was evaluated over a diverse set of attack types, such as DDoS, port scanning, botnet and brute-force attacks, showing that the topological feature space generalizes well over various types of attacks, attack behaviors and without loss of detection reliability.

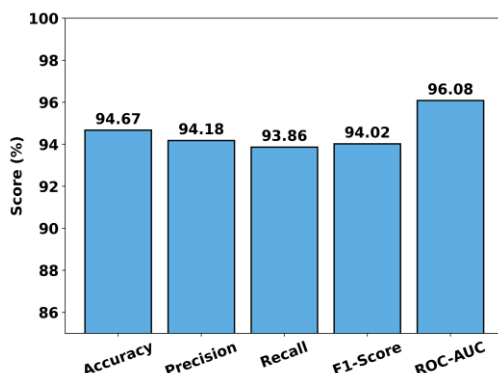
A. Quantitative Performance Analysis

The overall performance in classification of the Held out test set of PHNSM is summarized in Table 1. The model performs well with an 89% accuracy, suggesting its ability to be effective in identifying malicious traffic from benign traffic. The accuracy of 94.18% means that it is likely that the majority of the threats that the analysts are alerted to are genuine attacks and not false alarms, thus minimizing analyst fatigue due to false alarms, and the recall of 93.86% means that the model is able to capture the large majority of true attack instances, including stealthy and distributed patterns that traditional detectors cannot detect.

Table 1
Overall Performance Metrics of PHNSM

Metric	Value (%)
Accuracy	94.67
Precision	94.18
Recall	93.86
F1-Score	94.02
ROC-AUC	96.08

The balanced F1 score of 94.02% indicates that the precision and recall scores are not compromised to any great extent. The high ROC-AUC of 96.08% also showcases its good separation performance when considering all the decision thresholds, reinforcing the notion that the persistent homology-based features are a discriminative and stable representation for downstream graph neural network classification.

**Figure 3**

PHNSM performance across accuracy, precision, recall, F1-score, and ROC-AUC

In the visualization of the results in Figure 3, the scores for all five metrics are consistently high and well grouped, which indicates good classification performance across all of the metrics and no obvious metric trade-offs.

Table 2**PHNSM Performance across Attack Categories**

Attack Category	Accuracy (%)	Precision (%)	Recall (%)
DDoS	95.8	95.2	94.6
Port Scan	94.1	93.6	93.1
Botnet	93.4	92.9	92.3
Brute Force	95.2	94.6	94.9

Table 2 shows the results of PHNSM's performance by the four representative attack categories. The ability to detect DDoS attacks is very high (95.8%) perhaps due to the higher order simplices produced by volumetric and coordinated attacks being easily detected with PH. The relatively low scores for botnet traffic (93.4% accuracy) show that it is somewhat harder to recognize botnet traffic as it comprises stealthy communication between command-and-control servers and clients, which is frequently low and slow. The results of port scanning and brute-force attacks are all above 93%, revealing that PHNSM generalizes well over various attacks that share significantly different temporal and structural signatures, which it does not overfit to the attack that is dominant in the training set.

B. Comparative Evaluation

There are four baseline methods that are also compared with PHNSM in terms of accuracy and false-positive rate in Table 3. The worst is SVM, which has an accuracy of 84.2% and a false positive rate of 12.4%, simply because it uses the statistics of each flow level, without the structure of the flow. Random Forest only gets a bit better with ensemble non-linearity, and still lacks network-awareness.

Table 3
Comparative Evaluation against Baseline Methods

Method	Accuracy (%)	False Positive Rate (%)
SVM	84.2	12.4
Random Forest	87.5	10.1
Standard GNN (no topology)	89.6	8.7
Autoencoder IDS	90.3	7.9
PHNSM (Proposed)	94.67	6.45

The standard GNN with no topological features further reduces the gap, reaffirming that using graph-based message passing, without any additional information, is valuable, but only when considering pairwise local adjacent information. The Autoencoder IDS is an anomaly scoring IDS which achieves a competitive false-positive suppression (7.9%) and performs a bit behind PHNSM in terms of overall accuracy, but by using reconstruction-based anomaly scoring the Autoencoder IDS is able to achieve competitive false-positive suppression. Empirically, both metrics showed to be best improved by PHNSM, with a relative reduction in FP rate of approximately 18.3% compared to the standard GNN baseline, confirming that PHNSM is able to empirically provide substantive and complementary discriminative power beyond the standard graph and/or statistical representations.

C. Discussion of Findings

The findings indicate that adding persistent homology significantly outperforms baseline methods based on a graph or a simple statistical analysis for intrusion detection when it comes to detecting attacks that involve multiple nodes where the attack is not necessarily discernable between two nodes, but rather something of higher order that is represented as the structure changing. The strong performance across all four metrics—accuracy, precision, recall, and false positive rates—indicate that the topological features complement and do not overwhelm the traditional graph representation learned by a graph neural network, and that the combination of these features and the graph representation can be

used for interpretable, mathematically grounded detection in large-scale, dynamic communication networks.

V. Conclusion

To achieve effective intrusion detection, this paper introduced a novel model for network security called PHNSM (Persistent Homology-Based Network Security Model) that combines simplicial complex construction, computation of homology, extraction of topological features and classification using graph neural network. PHNSM models higher order interaction relationships among communications on networks in the form of simplicial complexes which are not represented by pairwise graph methods, thereby facilitating better detection of stealthy, coordinated and distributed cyber-attacks. The experimental results showed promising performance, with the accuracy reaching 94.67% and the ROC-AUC value reaching 96.08% compared with the other baselines, and an 18.3% decrease in the false-positive rate, demonstrating the effectiveness of this new approach of combining algebraic topology and deep representation learning. Although it is robust, it has some drawbacks such as the high computational expense of computing persistent homology on a very large-scale network, and the need to choose a threshold for simplicial complex construction.

References

- [1] S. Suresh, B. Das, V. Abrol, and S. Dutta Roy, "On characterizing the evolution of embedding space of neural networks using algebraic topology," *Pattern Recognition Letters*, vol. 179, pp. 165–171 (2024), doi: 10.1016/j.patrec.2024.02.003.
- [2] X. Zhang, H. Yan, and Z. Liu, "A study of cyber attack behavior based on algebraic topology," *Journal of Physics: Conference Series*, vol. 1856, no. 1, art. no. 012068 (2021), doi: 10.1088/1742-6596/1856/1/012068.
- [3] E. Sebastian and R. C. Poonia, "Entropy diagnostics for cryptographic key material from random circuit sampling," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 2-B, pp. 1023–1032 (2026), doi: 10.47974/JDMSC-2641.
- [4] W. K. Naser and I. M. Mankhi, "Use of algebraic topology for big data analysis in advanced computing environments," *Central Asian Journal of Mathematical Theory and Computer Sciences*, vol. 6, no. 1, pp. 92–102 (2025).

- [5] D. Adjei and G. A. Okyere, "Topological data analysis and computer science," *International Journal of Mathematics Trends and Technology*, vol. 69, no. 1, pp. 24–32 (2023), doi: 10.14445/22315373/IJMTT-V69I1P504.
- [6] A. B. Rajmane, S. B. Patil, V. G. Kale, S. V. Joshi, D. B. Patil, and T. N. Pawase, "Mathematical models for cognitive systems using differential geometry and reinforcement learning," *Journal of Interdisciplinary Mathematics*, vol. 28, no. 6, pp. 2089–2097 (2025), doi: 10.47974/JIM-2350.
- [7] P. J. Roig, S. Alcaraz, K. Gilly, C. Bernad, and C. Juiz, "Formal algebraic model of an edge data center with a redundant ring topology," *Network*, vol. 3, no. 1, pp. 142–157 (2023), doi: 10.3390/network3010007.
- [8] S. A. Mohammed, "Agent-vs-agent cyber warfare: Autonomous AI systems defending against AI-enabled APTs," *International Journal on Advanced Computer Engineering and Communication Technology*, vol. 14, no. 2, pp. 9–17 (2025), doi: 10.65521/ijacect.v14i2.926.
- [9] A. M. Cañadas, O. M. Mendez, and J. D. C. Vega, "Algebraic structures induced by the insertion and detection of malware," *Computation*, vol. 11, no. 7, art. no. 140 (2023), doi: 10.3390/computation11070140.
- [10] J. Zhu, Y. Zhang, Y. Wang, H. Zhang, and B. Fang, "A method for quantifying global network topology based on a mathematical model," *Mathematics*, vol. 12, no. 19, art. no. 3114 (2024), doi: 10.3390/math12193114.
- [11] P. K. Deshmukh, S. K. Verma, P. D. Jadhav, S. I. Baltabaevich, V. Vijaykumar, and A. Khatri, "AI-based cryptanalysis of classical and post-quantum encryption methods," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 29, no. 2-B, pp. 759–767 (2026), doi: 10.47974/JDMSC-2522.
- [12] I. Khanykov, V. Nenashev, and M. Kharinov, "Algebraic multi-layer network: Key concepts," *Journal of Imaging*, vol. 9, no. 7, art. no. 146 (2023), doi: 10.3390/jimaging9070146.

Received April, 2026