HyperDGA: Comprehensive Learning of High-Order and Global Semantic Associations for DGA Botnet Detection

Yixin Chen †, Weizhe Chen †, Lihua Yin *, Nan Wei *, Hongyu Yang, Lei Xiao, Jiaxin Wu

Posted Date: 13 May 2026

doi: 10.20944/preprints202605.0792.v1

Keywords: hypergraph; DGA botnet detection; multi-head node attention; Mamba

 doi:10.20944/preprints202605.0792.v1

Article

HyperDGA: Comprehensive Learning of High-Order and Global Semantic Associations for DGA Botnet Detection

Yixin Chen [1,†], Weizhe Chen [1,2,†], Lihua Yin [1,*], Nan Wei [1,*], Hongyu Yang [1,2], Lei Xiao [2] and Jiaxin Wu [2,3]

[1] Cyberspace Institute of Advanced Technology,Guangzhou University, Guangzhou 510006, China
[2] Xiamen Software Supply Chain Security Public Technology Service Platform, Xiamen 361024, China
[3] China Telecom Corporation Limited Fuzhou Branch, Fuzhou 350000,China
* Correspondence: yinlh@gzhu.edu.cn (L.Y.); wei_nan@gzhu.edu.cn (N.W.)
† The authors contribute equally to this work.

Abstract

Domain Generation Algorithm(DGA) is widely used by botnets to evade detection by generating numerous pseudo-random domains to communicate with command and control servers. While existing Graph Neural Networks attempt to detect DGA botnets by exploiting the feature similarity of these domains to model semantic associations via similarity graphs, they are restricted to binary relationships, causing information decay during multi-hop propagation. To overcome this, we propose HyperDGA. Treating domains as nodes, HyperDGA utilizes K-Nearest Neighbors to construct hyperedges, explicitly capturing high-order group semantic correlations. Subsequently, a Local Topology Aggregation module employs multi-head node attention-based hypergraph convolution to dynamically assign distinct aggregation weights to intra-hyperedge nodes, extracting fine-grained structural features. To mitigate the limited receptive field of hypergraph convolutions, a Global Node Association module integrates the selective state space model, Mamba, to capture long-range dependencies across all nodes. Experiments on two public datasets demonstrate that HyperDGA outperforms all baselines and achieves over 99% accuracy, validating the efficacy of high-order semantic modeling for DGA botnet detection.

Keywords: hypergraph; DGA botnet detection; multi-head node attention; Mamba

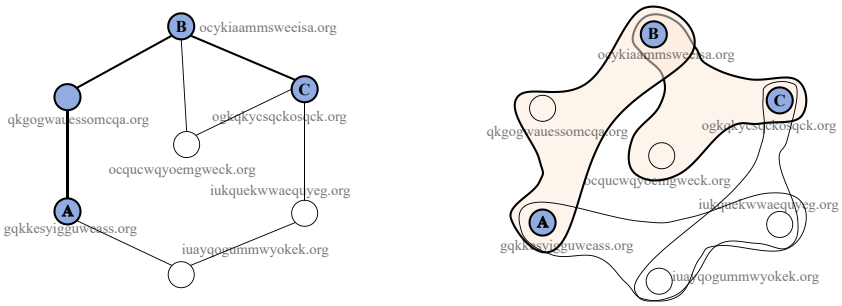
1. Introduction

Domain Generation Algorithm [1] represents a highly sophisticated evasion technique prevalent in IoT botnets [2] such as Conficker, Zeus, and Mirai. Infected hosts periodically invoke pseudo-random algorithms to dynamically generate thousands of candidate domains, sequentially attempting connections until successfully resolving a Command and Control(C&C) server. Early DGA botnet detection primarily leveraged machine or deep learning to model lexical features, including character entropy, n-grams, and domain length. Selvi et al [3]. utilized masked n-grams and Random Forest to maintain high performance with reduced features. These methods essentially treat domains as independent samples, neglecting the fact that DGA-generated botnets are structurally coupled via shared mechanisms and exhibit cohesive patterns within the feature space. Consequently, they fail to exploit intrinsic semantic associations, particularly when individual features are ambiguous but group-level patterns remain distinct. Graph Neural Network(GNN)-based methods are able to enhance inter-domain association modeling by constructing standard binary graphs. RMD-Graph [4] utilizes GNN to analyze domain-IP heterogeneous graphs derived from DNS logs, bridging the gap between individual features and group behaviors to enhance DGA botnet detection robustness. However, extracting domain query associations from DNS traffic is both computationally intensive and logistically

 doi:10.20944/preprints202605.0792.v1

challenging. Consequently, existing studies have validated the effectiveness of constructing topological graphs based on lexical features or latent semantic similarity. Jiao et al. [5] proposed DGGCN, which leverages string-based graphs and GCNs to capture latent semantic patterns missed by isolated models, thereby enhancing DGA botnet detection by modeling domain similarity. However, similarity-based binary graphs are restricted to pairwise associations. Decomposing cohesive DGA clusters into binary edges forces peripheral nodes to traverse multi-hop paths for core semantics, triggering information decay and oversmoothing that ultimately weakens group semantic representations.

Recently, hypergraph neural network(HGNN) [6] has emerged as a generalized graph topology that allows hyperedges to connect multiple nodes, demonstrating superior capability in modeling high-order group associations and complex semantic correlations. As shown in Figure 1, encapsulating semantically cohesive botnet clusters into hyperedges facilitates high-order group modeling via single-step interaction, bypassing multi-hop propagation limits. However, local hypergraph aggregation fails to capture the global dependencies across DGA families where global context association.



(a) domain graph (b) domain hypergraph

Figure 1. Comparison between domain graph and hypergraph. (a) In the traditional graph, A reaches B in two hops and C in three hops. (b) In the hypergraph, A reaches B in a single hop via a shared hyperedge and reaches C in only two hops. This illustrates how hypergraphs shorten propagation paths to capture high-order relationships more efficiently.

To address these limitations, we propose HyperDGA, a HGNN-based framework for DGA botnet detection. Specifically, HyperDGA utilizes K-Nearest Neighbors(KNN) algorithm to construct hyperedges from the k-nearest neighbors in domains' feature space, forming a DGA semantic hypergraph to model high-order group associations. A Local Topology Aggregation Module then employs multi-head attention-enhanced hypergraph convolutions to adaptively weight intra-edge nodes, extracting discriminative high-order topological representations. Furthermore, a Global Node Association Module leverages the Mamba architecture to model long-range dependencies across all nodes, effectively capturing cross-hyperedge and cross-cluster global associations. The main contributions of this paper are summarized as follows:

- We propose a feature similarity-based semantic hypergraph construction method that aggregates k related domains into a single hyperedge, overcoming traditional pairwise modeling limits to efficiently represent DGA botnet group semantic associations.
- We develop a synergistic modeling mechanism integrating local topology aggregation and global node association modules, which adaptively extracts high-order local topological features and captures grobal-node long-range dependencies, strengthening the feature representation of complex DGA associations.
- Extensive experiments on two real-world datasets demonstrate that HyperDGA achieves over 99% detection accuracy, validating its efficacy in high-order group modeling and node association analysis while significantly bolstering DGA botnet detection robustness.

 doi:10.20944/preprints202605.0792.v1

2. Related Work

2.1. Deep Learning-Based DGA Botnet Detection

Early DGA detection relied on manual feature engineering and traditional machine learning. Driven by deep learning advances, subsequent research adopted neural networks to extract character-level features. Tran et al. [7] and Chen et al. [8] improved detection by mitigating class imbalance and optimizing data distributions, respectively, while Tuan et al. [9] leveraged attention mechanisms to enhance feature modeling. Additionally, Sidi et al. [10] introduced Helix, which utilizes a CNN-LSTM autoencoder for effective representation learning and pattern identification. Liang et al. [11] developed HAGDetector, which leverages length-based segmentation and multi-feature fusion to improve detection capabilities across diverse DGA families. While deep learning boosts accuracy via character-level modeling, its focus on single domains neglects critical inter-domain associations. This failure to capture cross-sample correlations among co-occurring DGA botnet patterns limits the performance of conventional models.

2.2. Graph Learning-Based DGA Botnet Detection

To model DGA botnet associations, graph-based methods utilize GNNs to extract structural representations and uncover latent patterns, effectively enhancing detection performance. Jiao et al. [5] proposed DGGCN, which utilizes domain segmentation and string-based DomainGraphs to mine inter-domain associations via GCNs, enabling cross-semantic DGA Botnet detection. Zhang et al. [4] proposed RMD-Graph, which integrates domain-IP heterogeneous graphs with denoising and contrastive learning to bolster robustness in complex environments. Similarly, Wang et al. [12] developed HANDOM, utilizing heterogeneous attention networks to model intricate associations among domains, IPs, and clients. Although these methods partially model DGA botnet associations, most remain restricted to pairwise relationships, failing to capture high-order group-level semantic structures. Furthermore, their reliance on local neighborhood propagation hinders the capture of cross-cluster global associations, thereby limiting the representation capacity for complex DGA botnets.

3. Preliminary

3.1. Hypergraph Convolutional Networks

HGNNs generalize traditional GNNs by extending binary relations to high-order modeling. While standard graphs are restricted to pairwise edges, hypergraphs utilize hyperedges to link multiple nodes, capturing complex group interactions. This allows HGNNs to leverage high-order structural information for a more effective representation of intricate dependencies. To enable effective message passing on hypergraphs, Feng et al. proposed the Hypergraph Convolutional Networks(HGCN) [6] model, utilizing a "node-hyperedge-node" propagation mechanism for high-order information aggregation.

Given a hypergraph $G = (V, E)$, where V and E denote the sets of vertices and hyperedges respectively, let H be the incidence matrix representing their relationships. The hypergraph convolution operation is defined as:

$$X^{(l+1)} = \sigma(D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2} X^{(l)} \Theta^{(l)}) \tag{1}$$

Here, D_v and D_e denote the degree matrices of nodes and hyperedges, W is the hyperedge weight matrix, and $X^{(l)}$ represents the l-th layer node features. While this node-hyperedge-node cycle propagates high-order information, uniform node weighting fails to capture the varying impacts of different associations. We introduce multi-head node attention-based HGCN for local topological aggregation. By assigning adaptive weights during node-hyperedge-node propagation, the module achieves weighted feature aggregation and bolsters high-order topological modeling.

 doi:10.20944/preprints202605.0792.v1

3.2. Mamba

State Space Models(SSMs) [13] are a class of methods for recursive sequence modeling via hidden states, fundamentally expressed as:

$$h_t = Ah_{t-1} + Bx_t, y_t = Ch_t \tag{2}$$

where h_t, x_t, and y_t represent the hidden state, input, and output, respectively. This recursive structure enables sequence modeling and possesses the capability to capture long-range dependencies. Gu et al. proposed the Mamba model [14], which introduces a selective SSM. By maintaining the structural stability of the state transition matrix and employing input-dependent dynamic parameters for input and output projections, the state update process is defined as:

$$h(t_{k+1}) = Ah(t_k) + B_t x(t_k), \quad y(t_k) = C_t h(t_k)$$

We leverage Mamba to implement a global node association module, which performs global modeling on DGA nodes, achieving comprehensive global relationship learning.

4. Methodology

As shown in Figure 2, HyperDGA first constructs the semantic hypergraph to explicitly encode high-order group associations among domains. Building on this structure, the Local Topology Aggregation Module extracts fine-grained structural information from neighborhood topologies, while the Global Node Association Module leverages Mamba to capture long-range dependencies across all nodes, effectively modeling global relationships and overcoming the limited receptive field of local aggregation.

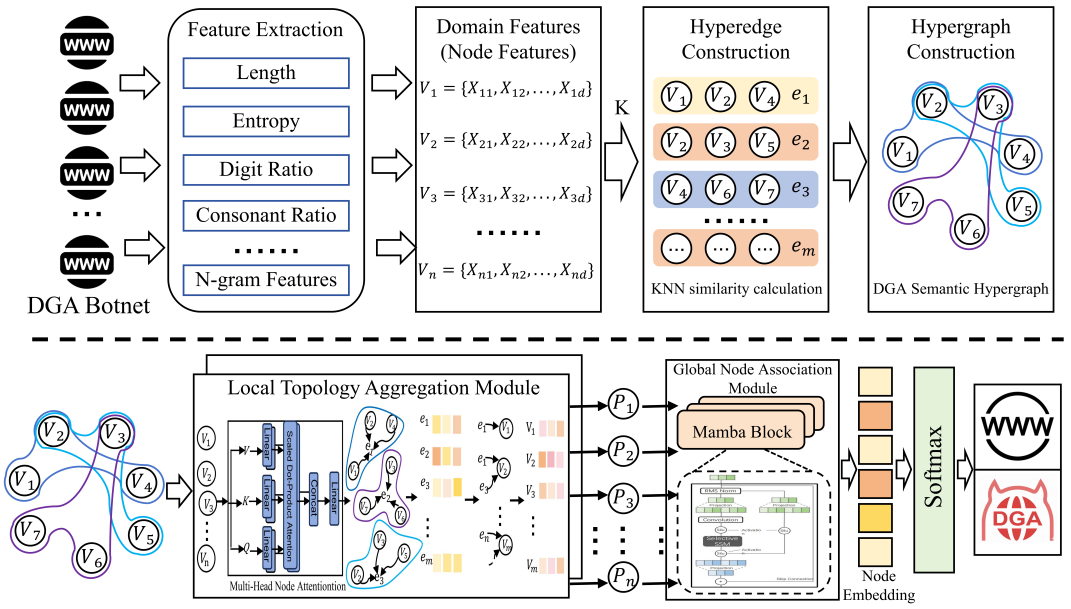


Figure 2. The Overview Architecture of HyperDGA.

4.1. Semantic Hypergraph Construction

Let $D = \{(d_i, y_i)\}_{i=1}^{N}$ be the DGA Botnet set, where $y_i \in \{0, 1\}$ indicates benign domain 0 or malicious DGA botnet 1. Each DGA botnet is represented by an F-dimensional feature vector, forming the feature matrix $X \in \mathbb{R}^{N \times F}$. By mapping each feature x_i to a node $v_i \in V$, we identify the k-nearest neighbors $\mathcal{N}_k(v_i)$ for each node in the Euclidean space:

$$\mathcal{N}_k(v_i) = \underset{ki \neq i}{\text{top-k}}\{v_{ki} : \|\mathbf{x}_i - \mathbf{x}_{ki}\|_2\} \tag{3}$$

 doi:10.20944/preprints202605.0792.v1

The hyperedge e_j is defined as:

$$e_j = \{v_i\} \cup \{v_{ki} | v_{ki} \in \mathcal{N}_k(i)\} \tag{4}$$

Iterating over all nodes yields $M = N$ hyperedges, forming the set $E = \{e_1, e_2, \ldots, e_M\}$. The resulting incidence matrix $H \in \{0,1\}^{N \times M}$ is defined as:

$$H_{i,j} = \begin{cases} 1, & v_i \in e_j \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The DGA semantic hypergraph is defined as $G = (V, E, H)$. In G, the degree of node v_i is defined by the number of its incident hyperedges: $d(v_i) = \sum_{j=1}^{M} H_{i,j}$. Consequently, the node degree matrix is $D_v = \text{diag}(d(v_1), \ldots, d(v_N))$, and the hyperedge degree matrix $D_e = \text{diag}(d(e_1), \ldots, d(e_N))$ is constant, with:

$$d(e_j) = \sum_{i=1}^{N} H_{i,j} = k + 1, \quad \forall j \tag{6}$$

This construction organizes feature-similar domains into high-order structures, capturing group semantic associations beyond pairwise relations. For instance, domains from the same DGA family that cluster by features like entropy, length, and character ratios are mapped into high-order units via KNN hyperedges.

4.2. Local Topology Aggregation Module

Following hypergraph construction, domain nodes are integrated into multiple hyperedges rather than being isolated. The Local Topology Aggregation Module(LTAM) then employs node attention-based hypergraph convolution to aggregate information from high-order neighborhoods, utilizing adaptive weights to reflect each sample's semantic contribution to the group.

In the LTAM, node features $V \in \mathbb{R}^{N \times F}$ undergo L layers of node attention-based hypergraph convolution. At layer l, the input node representation is denoted as $Z^{(l)} \in \mathbb{R}^{N \times d_l}$, where $z_v^{(l)}$ is the current representation of domain node v. For each hyperedge e, the module first generates a hyperedge context vector based on the representations of all constituent nodes:

$$\mathbf{c}_e^{(l)} = \frac{1}{d(e)} \sum_{v \in e} z_v^{(l)} \tag{7}$$

In the r-th attention head, the module integrates the node representation with its hyperedge context to compute the contribution score $s_{ve}^{(l,r)}$ of node v to hyperedge e:

$$s_{ve}^{(l,r)} = \text{LeakyReLU}(\mathbf{a}_r^T [\Theta_r^{(l)} \mathbf{z}_v^{(l)} \parallel \Phi_r^{(l)} \mathbf{c}_e^{(l)}]) \tag{8}$$

where $\parallel$ denotes concatenation, and $\mathbf{a}_r^T$, $\Theta_r^{(l)}$, and $\Phi_r^{(l)}$ are learnable parameters. The attention coefficient $\alpha_{ve}^{(l,r)}$ is then derived by normalizing these scores within the hyperedge:

$$\alpha_{ve}^{(l,r)} = \frac{\exp(s_{ve}^{(l,r)})}{\sum_{u \in e} \exp(s_{ue}^{(l,r)})} \tag{9}$$

The attention coefficient $\alpha_{ve}^{(l,r)}$ signifies the importance of node v within hyperedge e. Instead of simple averaging, this mechanism adaptively aggregates discriminative neighbors to form high-order messages:

$$m_e^{(l,r)} = \sum_{v \in e} \alpha_{ve}^{(l,r)} \Theta_r^{(l)} \mathbf{z}_v^{(l)} \tag{10}$$

 doi:10.20944/preprints202605.0792.v1

Subsequently, each node receives messages from its incident hyperedges and updates its representation as follows:

$$\mathbf{z}_v^{(l+1,r)} = \sigma\left(\frac{1}{d(v)}\sum_{e\in E(v)} w(e)\frac{1}{d(e)}\mathbf{m}_e^{(l,r)}\right), \quad E(v) = \{e \mid v \in e\} \tag{11}$$

where $w(e)$ is the hyperedge weight matrix. To enhance representation capacity, the LTAM employs multi-head attention and concatenates the results from R heads. The final output for node v at layer $l+1$ is given by:

$$\mathbf{z}_v^{(l+1)} = \|_{r=1}^{R}\mathbf{z}_v^{(l+1,r)} \tag{12}$$

Different attention heads model neighborhood relationships across distinct feature subspaces. The concatenated result undergoes Batch Normalization, ReLU activation, and Dropout to enhance training stability and mitigate overfitting, yielding node representation Z embedded with high-order local topological information.

4.3. Global Node Association Module

While the LTAM captures local high-order associations, its information propagation remains constrained by hyperedge structures despite stacking layers to expand the receptive field. To address this, HyperDGA incorporates the Global Node Association Module(GNAM), enabling comprehensive interaction across all nodes' topological representations to capture global dependencies.

The node embeddings obtained from LTAM are treated as a representation sequence across all domain samples. Let the output Z be:

$$Z = [\mathbf{z}_1; \mathbf{z}_2; \ldots; \mathbf{z}_N] \in \mathbb{R}^{N\times d} \tag{13}$$

Pre-normalization and linear projection are then applied to the node representations to obtain U:

$$U = \mathbf{W}_{in}\mathrm{LN}(Z) \in \mathbb{R}^{N\times d_m} \tag{14}$$

Given the node representation sequence $U = (u_1, u_2, \ldots, u_N)$ of length N, Mamba models global associations based on the SSM. The core recurrence process is defined as:

$$s_t = \bar{A}_t s_{t-1} + \bar{B}_t u_t,\ o_t = C_t s_t + D u_t \tag{15}$$

where s_t denotes the state vector and o_t is the output at the t-th node position. The parameters $\bar{A}_t$, $\bar{B}_t$, and C_t are input-dependent and vary dynamically.This selective mechanism enables the model to adaptively retain, update, or attenuate global information based on current node representations, capturing long-range dependencies across the entire set of nodes. The global output is then linearly projected back to the embedding dimension d, utilizing a residual connection to preserve local structural information learned by the LTAM:

$$Z^* = Z + \mathrm{Dropout}(W_{out}\mathrm{Mamba}(U)) \tag{16}$$

Z^* serves as the final node representation, fusing local high-order topology with global node associations. Finally, an MLP classification head and softmax function yield the predicted label probability $\hat{y}_i$, with the entire framework trained end-to-end via cross entropy loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)) \tag{17}$$

 doi:10.20944/preprints202605.0792.v1

5. Experiments

5.1. Datasets and Experimental Settings

We evaluate our model on two public datasets. The open source DGA dataset [15] introduced by T. Wang et al. is denoted as DGA and consists of DGA botnet and benign domains. After preprocessing and deduplication, the dataset comprises 91599 samples, including 42076 DGA botnet and 49523 benign domains. We also employ the "Benign and Malicious Domains based on DNS Logs" dataset [16] from Mendeley Data (hereafter referred to as benign_malicious). This dataset, built from real DNS logs, contains 90000 samples, including 45000 DGA botnets. Table 1 details the dataset specifications along with the node and hyperedge counts for the constructed hypergraphs.

To evaluate HyperDGA, we adopt Accuracy, Precision, Recall, and F1-score as evaluation metrics. Accuracy measures overall classification correctness, while Precision and Recall reflect identification precision and detection capability for malicious samples, respectively. F1-score provides a comprehensive assessment of model performance. HyperDGA is trained end-to-end, with key hyperparameters summarized in Table 2.

Table 1. Dataset Description.

Dataset Name	Domains	Features	Nodes	Hyperedges
DGA	91599	7	91599	91599
benign_malicious	90000	280	90000	90000

Table 2. Hyperparameter Configurations.

Hyperparameter	Value
Number of neighbors k	3
Distance metric	Euclidean distance
Number of HGCN layers	2
Number of attention heads	4
Learning rate	0.001
Training epochs	100

5.2. Performance Comparison

To ensure objectivity and comprehensiveness, we select nine representative baseline models. Random Forest(RF) [3] represents traditional machine learning for lexical statistical analysis. The deep learning methods including MLP [17], CNN [18], CNN-GRU [19], BotDetector(CNN-LSTM) [20], and Bert [21] covers various representation learning methods from statistical features to contextual dependencies. GCN [22] and GAT [23] model binary relationships via similarity graphs. And HGNN [6] is utilized to compare performance in high-order relationship modeling.

As shown in Table 3, individual-based models such as RF, MLP, and CNN-GRU achieve F1-scores around 94% on the DGA dataset, trailing HyperDGA by around 5.4% due to a lack of inter-sample association modeling. Bert attains the best baseline performance through intra-domain context modeling, yet its F1-score remains 2.6% lower than HyperDGA due to the absence of cross-domain association modeling. Although graph learning methods introduce node relationships, their binary structures fail to capture high-order semantics. Similarly, while HGNN expresses high-order relationships, its local aggregation mechanism lacks global long-range dependency modeling capabilities, strictly limiting its performance gains. In contrast, HyperDGA effectively overcomes the limitations of isolated features, binary topologies, and local aggregation, achieving optimal detection performance.

 doi:10.20944/preprints202605.0792.v1

Table 3. Performance Comparison on DGA Dataset.

Scheme	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9431	0.9422	0.9439	0.9428
MLP	0.9456	0.9448	0.9462	0.9454
CNN	0.9407	0.9398	0.9411	0.9404
CNN-GRU	0.9434	0.9425	0.9442	0.9432
Bert	0.9735	0.9735	0.9735	0.9735
BotDetector	0.9028	0.9024	0.9050	0.9026
GCN	0.9443	0.9433	0.9453	0.9440
GAT	0.9437	0.9445	0.9437	0.9438
HGNN	0.9434	0.9424	0.9443	0.9431
HyperDGA	**0.9992**	**0.9991**	**0.9992**	**0.9992**

As shown in Table 4, HyperDGA maintains the highest detection performance on the benign_malicious dataset, achieving 99.67% across all metrics and outperforming all baselines. While baselines like GCN, GAT, and MLP achieve F1-scores over 99%, which indicates the highly discriminative nature of the dataset features, HyperDGA still achieves optimal performance. This proves its ability to further mine latent high-order group associations and long-range dependencies even with strong initial features. Conversely, the sharp performance drops in BotDetector to 75.99% and HGNN to 92.01% demonstrate that relying solely on intra-sample feature modeling or local hypergraph aggregation is inadequate for capturing complex feature distributions and associations.

Overall experimental results from both datasets demonstrate that by combining high-order relationship and global node association modeling, the proposed method maintains strong discriminative capability, superior adaptability, and stable detection performance across different datasets.

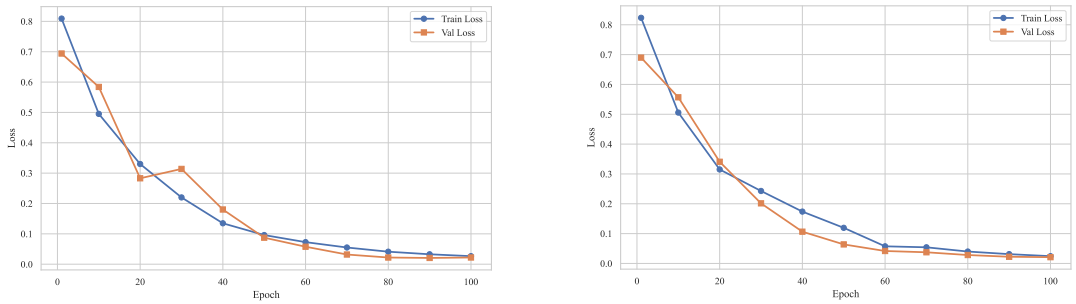
Table 4. Performance Comparison on benign_malicious Dataset.

Scheme	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9878	0.9879	0.9878	0.9878
MLP	0.9913	0.9914	0.9913	0.9913
CNN	0.9625	0.9625	0.9625	0.9625
CNN-GRU	0.8936	0.8969	0.8936	0.8933
Bert	0.9732	0.9732	0.9732	0.9732
BotDetector	0.7630	0.7866	0.7676	0.7599
GCN	0.9933	0.9933	0.9933	0.9933
GAT	0.9937	0.9937	0.9937	0.9937
HGNN	0.9205	0.9289	0.9205	0.9201
HyperDGA	**0.9967**	**0.9967**	**0.9967**	**0.9967**

5.3. Convergence Analysis

As shown in Figure 3, both training and validation losses decrease steadily and eventually converge across both datasets. The rapid initial loss reduction indicates that the model swiftly captures core discriminative information, while the subsequent stable convergence demonstrates the effective fusion of local high-order topology and global association features. Furthermore, the close alignment between training and validation loss curves confirms the absence of significant overfitting. Specifically, the DGA dataset stabilizes after approximately 60 epochs, whereas the benign_malicious dataset exhibits a smoother convergence process. Overall, HyperDGA demonstrates excellent convergence performance and training stability on both datasets.

 doi:10.20944/preprints202605.0792.v1



(a) DGA Dataset (b) benign_malicious Dataset

Figure 3. Training and validation losses on the DGA and benign_malicious datasets.

5.4. Parameters Sensitivity Analysis

Figure 4 illustrates the parameter sensitivity experiments analyzing the impact of the KNN neighbor count K on semantic hypergraph construction. On both the DGA and benign_malicious datasets, the model achieves peak accuracy at K=3, indicating that a compact neighborhood scale optimally constructs effective semantic hyperedges. As K increases, overall performance declines. This is particularly evident on the DGA dataset where accuracy drops significantly at K=6, likely because larger neighborhoods introduce weakly correlated nodes into a single hyperedge and generate noise during local aggregation. Since K=3 yields optimal results across accuracy, precision, recall, and F1-metrics by effectively capturing group semantic associations among feature-similar domains, it is selected as the final hypergraph construction parameter.

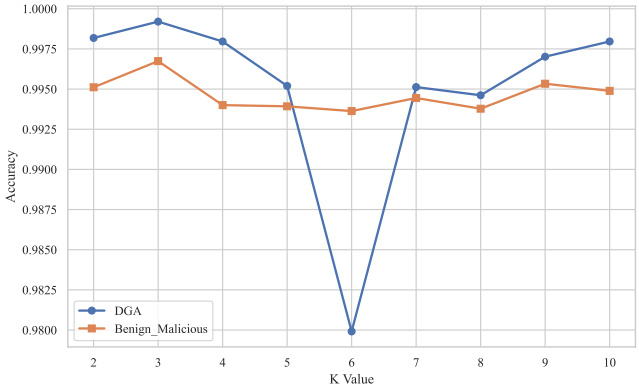


Figure 4. Impact of the number of K on Semantic Hypergraph Construction.

5.5. Ablation Study

To verify the effectiveness and synergistic modeling capabilities of the core components in the HyperDGA model, we designed six groups of ablation experiments across two datasets to investigate the impact of semantic hypergraph construction, the Local Topology Aggregation Module(LTAM), and the Global Node Association Module (GNAM) on model performance. Specifically, Group 1 retains only the semantic hypergraph construction module and employs standard hypergraph convolution without multi-head attention mechanism for feature learning to validate the baseline performance of the hypergraph structure. Group 2 and Group 5 set the parameter K to 1, which degrades the hypergraph into a traditional binary graph. Furthermore, Group 3 removes all graph structures and applies GNAM directly to the original features to assess the independent effectiveness of global association learning.

The experimental results in Table 5 and Table 6 demonstrate that the complete Group 6 achieves optimal performance across all metrics on both datasets with accuracies reaching 99.92% and 99.67% respectively, which comprehensively verifies the effectiveness of the synergistic module design. For Group 3, which exclusively utilizes the GNAM module, HyperDGA attains accuracies of 97.15% and

 doi:10.20944/preprints202605.0792.v1

98.79% on the two datasets. Analyzing this alongside Group 4 and Group 6 confirms the effectiveness of the Mamba model in mitigating local aggregation deficiencies and establishing global node associations. Furthermore, it indicates that relying solely on global association modeling without extracting fine-grained high-order neighborhood features remains insufficient. A comparison between Group 4 and Group 6 reveals that depending exclusively on the LTAM module creates a distinct performance bottleneck due to a restricted receptive field, resulting in an accuracy decrease of approximately 6.1% on the DGA dataset. A comparison between Group 5 and Group 6 reveals that high-order hyperedges capture the group semantic consistency of DGA botnets more effectively than traditional binary relationships. Overall, the semantic hypergraph in HyperDGA represents high-order group relationships among domains and successfully integrates local high-order group representations with global long-range associations to achieve high precision DGA botnet detection.

Table 5. Ablation Study on DGA Dataset.

HyperGraph	LTAM	GNAM	Accuracy	Precision	Recall	F1
1	0	0	0.9371	0.9362	0.9381	0.9369
0	1	0	0.9402	0.9393	0.9409	0.9399
0	0	1	0.9715	0.9708	0.9722	0.9714
1	1	0	0.9382	0.9376	0.9382	0.9378
0	1	1	0.9918	0.9913	0.9924	0.9918
1	1	1	**0.9992**	**0.9991**	**0.9992**	**0.9992**

Table 6. Ablation Study on benign_malicious Dataset.

HyperGraph	LTAM	GNAM	Accuracy	Precision	Recall	F1
1	0	0	0.9189	0.9272	0.9189	0.9185
0	1	0	0.8947	0.9111	0.8947	0.8936
0	0	1	0.9879	0.9880	0.9879	0.9879
1	1	0	0.9199	0.9283	0.9199	0.9195
0	1	1	0.9942	0.9943	0.9942	0.9942
1	1	1	**0.9967**	**0.9967**	**0.9967**	**0.9967**

To further analyze the learned feature representations, t-SNE visualization was performed on the raw features and the trained embeddings for both datasets as shown in Figures 5 and 6. In the raw feature space of both datasets, the two classes of samples appear dispersed with significant overlap, suggesting that the discriminative boundary between benign domains and DGA botnets is unclear when relying solely on original lexical and statistical features.

(a) raw feature (b) Embedding

Figure 5. Embedding Analysis on DGA Dataset.

After learning with HyperDGA, the samples form distinct cluster structures in the embedding space with significantly enhanced inter-class separation and more concentrated intra-class distributions, which proves the ability of the proposed method to learn discriminative high-order representations. These visualization results indicate that HyperDGA integrates local high-order relationships and global

 doi:10.20944/preprints202605.0792.v1

association information with statistical domain features to learn more powerful node representations and establish a clearer decision boundary for subsequent detection.

(a) raw feature (b) Embedding

Figure 6. Embedding Analysis on benign_malicious dataset.

6. Conclusion

Addressing the failure of current DGA botnet detection methods to capture high-order group semantic correlations and global node associations, we propose a HGNN-based scheme HyperDGA. This framework constructs semantic hypergraphs from domain feature similarity to organize related nodes into hyperedges and transcend the limitations of traditional pairwise graph structures. The Local Topology Aggregation Module uses multi-head node attention-based hypergraph convolution to strengthen local high order topological representations. Furthermore we introduce a Mamba-based Global Node Association Module to model long-range dependencies between node representations and compensate for the restricted receptive field of LTAM for synergistic modeling of local and global information. However, the current semantic hypergraph relies primarily on feature similarity and insufficiently utilizes behavioral correlations such as DNS query patterns. Future research will explore behavioral correlation modeling to more comprehensively identify latent DGA botnet patterns.

Acknowledgments: This work is supported by the National Natural Science Foundation of China (Nos.U25B2026 and 62502107), the Guangdong Basic and Applied Basic Research Foundation (No. 2026A1515011762) and the Open Fund of Xiamen Software Supply Chain Security Public Technology Service Platform (No.3502Z20231042).

Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Sood, A.K.; Zeadally, S. A taxonomy of domain-generation algorithms. *IEEE Security & Privacy* **2016**, *14*, 46–53.
2. Dange, S.; Chatterjee, M. IoT botnet: The largest threat to the IoT network. In *Data Communication and Networks: Proceedings of GUCON 2019*; Springer, 2019; pp. 137–157.
3. Selvi, J.; Rodríguez, R.J.; Soria-Olivas, E. Detection of algorithmically generated malicious domain names using masked N-grams. *Expert Systems with Applications* **2019**, *124*, 156–163. https://doi.org/https://doi.org/10.1016/j.eswa.2019.01.050.
4. Zhang, S.; Huang, L.; Zhang, Z.; Xu, W.; Yang, W.; Liu, L. RMD-Graph: Adversarial Attacks Resisting Malicious Domain Detection Based on Dual Denoising. *IEEE Transactions on Knowledge and Data Engineering* **2025**, *37*, 1394–1410. https://doi.org/10.1109/TKDE.2024.3520798.
5. Jiao, H.; Wang, Q.; Fan, Z.; Liu, J.; Du, D.; Li, N.; Liu, Y. DGGCN: Dictionary based DGA detection method based on DomainGraph and GCN. In Proceedings of the 2022 International Conference on Computer Communications and Networks (ICCCN), 2022, pp. 1–10. https://doi.org/10.1109/ICCCN54977.2022.9868932.
6. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. In Proceedings of the Proceedings of the AAAI conference on artificial intelligence, 2019, Vol. 33, pp. 3558–3565.
7. Tran, D.; Mac, H.; Tong, V.; Tran, H.A.; Nguyen, L.G. A LSTM based framework for handling multiclass imbalance in DGA botnet detection. *Neurocomputing* **2018**, *275*, 2401–2413. https://doi.org/https://doi.org/10.1016/j.neucom.2017.11.018.

 doi:10.20944/preprints202605.0792.v1

8. Chen, Y.; Pang, B.; Shao, G.; Wen, G.; Chen, X. DGA-based botnet detection toward imbalanced multiclass learning. *Tsinghua Science and Technology* **2021**, *26*, 387–402. https://doi.org/10.26599/TST.2020.9010021.
9. Tuan, T.A.; Long, H.V.; Taniar, D. On Detecting and Classifying DGA Botnets and their Families. *Computers & Security* **2022**, *113*, 102549. https://doi.org/https://doi.org/10.1016/j.cose.2021.102549.
10. Sidi, L.; Mirsky, Y.; Nadler, A.; Elovici, Y.; Shabtai, A. Helix: DGA Domain Embeddings for Tracking and Exploring Botnets. In Proceedings of the Proceedings of the 29th ACM International Conference on Information & Knowledge Management, New York, NY, USA, 2020; CIKM '20, p. 2741–2748. https://doi.org/10.1145/3340531.3416022.
11. Liang, J.; Chen, S.; Wei, Z.; Zhao, S.; Zhao, W. HAGDetector: Heterogeneous DGA domain name detection model. *Computers & Security* **2022**, *120*, 102803. https://doi.org/https://doi.org/10.1016/j.cose.2022.102803.
12. Wang, Q.; Dong, C.; Jian, S.; Du, D.; Lu, Z.; Qi, Y.; Han, D.; Ma, X.; Wang, F.; Liu, Y. HANDOM: Heterogeneous Attention Network Model for Malicious Domain Detection. *Computers & Security* **2023**, *125*, 103059. https://doi.org/https://doi.org/10.1016/j.cose.2022.103059.
13. Hamilton, J.D. State-space models. *Handbook of econometrics* **1994**, *4*, 3039–3080.
14. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In Proceedings of the First Conference on Language Modeling, 2024.
15. Wang, T.; Chen, L.C.; Genc, Y. A dictionary-based method for detecting machine-generated domains. *Information Security Journal: A Global Perspective* **2021**, *30*, 205–218.
16. Marques, C. Benign and malicious domains based on DNS logs, 2021. https://doi.org/10.17632/623sshkdrz.5.
17. Taud, H.; Mas, J.F. Multilayer perceptron (MLP). In *Geomatic approaches for modeling land change scenarios*; Springer, 2017; pp. 451–455.
18. Zhou, S.; Lin, L.; Yuan, J.; Wang, F.; Ling, Z.; Cui, J. CNN-based DGA detection with high coverage. In Proceedings of the 2019 IEEE international conference on intelligence and security informatics (ISI). IEEE, 2019, pp. 62–67.
19. Ke, W.; Zheng, D.; Zhang, C.; Deng, B.; Yao, H.; Tian, L. CGFMD: CNN and GRU Based Framework for Malicious Domain Name Detection. In Proceedings of the International Conference on Artificial Intelligence and Security. Springer, 2022, pp. 564–574.
20. Zang, X.; Cao, J.; Zhang, X.; Gong, J.; Li, G. BotDetector: a system for identifying DGA-based botnet with CNN-LSTM. *Telecommunication Systems* **2024**, *85*, 207–223.
21. Koroteev, M.V. BERT: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943* **2021**.
22. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* **2016**.
23. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903* **2017**.