MAGNET: A Hybrid Deep Learning Framework for Android Malware Detection Using Multi-Modal Feature Analysis

Alireza Iranmanesh* and Hamid Mirvaziri†

*Department of Computer Engineering, Shahid Bahonar University, Kerman, Iran

Email: alirezairanmanesh78@gmail.com

†Department of Computer Engineering, Shahid Bahonar University, Kerman, Iran

Email: h.mirvaziri@gmail.com

Abstract—Background: The proliferation of Android devices, coupled with escalating cyber threats, underscores the need for robust malware detection. Traditional single-modal approaches struggle against sophisticated obfuscation techniques. Method: This paper presents MAGNET (Multi-modal Analysis for Graph-based NEtwork Threats), a novel framework integrating tabular (static features), graph-based (function call graphs), and sequential (API call sequences) data modalities. It leverages three specialized modules—EnhancedTabTransformer, GraphTransformer, and SequenceTransformer—combined with a dynamic attention mechanism and multi-modal fusion layer. Results: Evaluations on the DREBIN dataset (6,092 samples: 4,641 training, 1,451 testing) demonstrate MAGNET's superior performance, achieving $97.24 \pm 0.65\%$ accuracy, 0.9823 ± 0.0042 F1-Score, and 0.9932 ± 0.0035 AUC, outperforming baselines (SVM: 90.6%, Random Forest: 93.5%, XGBoost: 94.8%, ANN: 96.2%). Ablation studies validate each component's contribution. Conclusion: MAGNET's multi-modal approach and advanced architectures offer a robust solution for Android malware detection, with strong potential for operational cybersecurity applications.

Keywords: Android malware detection, multi-modal learning, graph neural networks, transformer architecture, cybersecurity, DREBIN, MAGNET

Nomenclature

I. Introduction

With over 70% global market share, Android dominates the mobile ecosystem, making it a prime target for cyberattacks. Security reports document a rise in Android malware from 3.2 million samples in 2020 to over 5.8 million in 2023 ?. Conventional detection methods, reliant on static signatures, falter against advanced obfuscation, encryption, and AI-generated malware ?.

This paper introduces MAGNET (Multi-modal Analysis for Graph-based NEtwork Threats), a hybrid deep learning framework that integrates three data modalities—tabular (permissions, components), graph-based (function call graphs), and sequential (API call sequences)—to enhance detection accuracy. MAGNET employs specialized

Transformer-based modules and a novel dynamic attention mechanism, optimized via the PIRATES algorithm.

Key contributions include:

- A unified multi-modal architecture with three specialized modules.
- A dynamic attention mechanism for optimal feature fusion.
- The PIRATES algorithm for automated hyperparameter optimization.
- Comprehensive evaluation on the DREBIN dataset ?.

II. Related Work

A. Evolution of Android Malware Detection

Early static analysis methods, such as DREBIN ?, utilized features like permissions and API calls, achieving 94% accuracy with SVM. Schmidt et al. ? proposed a framework analyzing AndroidManifest.xml and DEX code, but its 87.3% accuracy diminished against obfuscated malware.

B. Deep Learning Approaches

Kim et al. ? employed Deep Belief Networks (DBNs) for API call analysis, achieving 96.5% accuracy. Wang et al. ? extended DBNs to static and dynamic features, reaching 97.8% accuracy.

C. Multi-Modal Analysis

Alzaylaee et al. ? combined static, dynamic, and textual features, achieving 98.2% accuracy. Chen et al. ? used Graph Neural Networks (GNNs) for program structure analysis, yielding 96.7% accuracy.

III. Proposed Methodology

A. MAGNET Architecture

MAGNET integrates three data streams via specialized modules, fused through a dynamic attention mechanism and a multi-modal fusion layer.

EnhancedTabTransformer: Processes static features, including:

- 128 permissions.
- App components (Activities, Services, Receivers).
- Static API calls and AndroidManifest.xml metadata.

The module employs a 6-layer Transformer with 256-dimensional embeddings.

GraphTransformer: Analyzes function call graphs (average 1,245 nodes, 3,872 edges), with:

- Node features: function type, call frequency (64 dimensions).
- Edge features: call frequency, type (32 dimensions).

It uses a 4-layer GNN with attention-based aggregation.

SequenceTransformer: Processes API call sequences (average length: 87), encoded via Word2Vec, preserving temporal order. It employs a 5-layer Transformer with 128-dimensional embeddings.

B. Dynamic Attention Mechanism

The attention mechanism integrates module outputs:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where Q, K, and V are Query, Key, and Value matrices, and d_k is the key dimension.

C. Multi-Modal Fusion

The final output combines module representations:

$$\text{Output} = \alpha \cdot h_{\text{tab}} + \beta \cdot h_{\text{graph}} + \gamma \cdot h_{\text{seq}} \tag{2}$$

Weights α, β, and γ are learned adaptively during training.

D. PIRATES Optimization

The PIRATES algorithm, a custom hyperparameter optimization strategy, iteratively adjusts learning rates, batch sizes, and layer configurations over 476 trials to maximize performance.

IV. Evaluation

A. Dataset

The DREBIN dataset ? comprises 6,092 samples:

- Training: 4,641 samples.
- Testing: 1,451 samples (327 benign, 1,124 malicious).
- Period: 2010–2014.

B. Experimental Setup

Hardware:

- CPU: Intel Core i7-8700K.
- GPU: NVIDIA RTX 3080 (10GB VRAM).
- RAM: 32GB DDR4-3200.
- Storage: 256GB NVMe SSD.

Software:

- Python 3.8.10, PyTorch 1.12.0, PyTorch Geometric 2.1.0, CUDA 11.6.

TABLE I: Five-Fold Cross-Validation Results for MAGNET

Metric	Value
Accuracy	0.9722 ± 0.0065
Precision	0.9810 ± 0.0102
Recall	0.9828 ± 0.0072
F1-Score	0.9818 ± 0.0042
AUC	0.9932 ± 0.0035

Training: The model was trained for 100 epochs with a batch size of 32, using the Adam optimizer (learning rate: 0.001).

C. Reproducibility

Code and hyperparameters are available at [repository URL placeholder]. The dataset is publicly accessible ?.

V. Results

A. Overall Performance

MAGNET achieved:

- Accuracy: 97.24 ± 0.65%.
- F1-Score: 0.9823 ± 0.0042.
- Precision: 0.9796 ± 0.0102.
- Recall: 0.9849 ± 0.0072.
- AUC: 0.9932 ± 0.0035.

B. Cross-Validation

Five-fold cross-validation results are shown in Table I.

C. Comparison with Baselines

Table II compares MAGNET with baseline methods.

D. Ablation Study

Table III highlights the contribution of each component.

TABLE II: Comparison with Baseline Methods

Method	Acc.	Prec.	Rec.	F1	AUC
SVM	0.906	0.915	0.892	0.903	0.945
Random Forest	0.935	0.942	0.928	0.935	0.967
XGBoost	0.948	0.953	0.943	0.948	0.978
ANN	0.962	0.965	0.959	0.962	0.985
MAGNET	0.972	0.980	0.985	0.982	0.993

TABLE III: Ablation Study Results

Configuration	F1-Score
EnhancedTabTransformer	0.945
GraphTransformer	0.894
SequenceTransformer	0.907
Without Dynamic Attention	0.954
Without Multi-Modal Fusion	0.967
Full MAGNET	0.982

TABLE IV: Comparison with State-of-the-Art Methods

Method	Acc. (%)	F1	AUC
MAGNET	97.24	0.982	0.993
DREBIN (SVM)	92.3	0.933	0.955
PIKADROID	96.8	0.974	0.988
CrossMalDroid	95.2	0.952	0.976
DroidAPIMiner	89.7	0.891	0.927
DeepImageDroid	96.0	0.960	0.982
BERT-Graph	95.5	0.950	0.975

E. Confusion Matrix

Test results (1,451 samples):

- True Negatives: 304.
- False Positives: 23.
- False Negatives: 17.
- True Positives: 1,107.

F. Comparison with State-of-the-Art

Table IV compares MAGNET with advanced methods.

VI. Discussion

MAGNET's superior performance (97.24% accuracy, 0.9823 F1-Score) stems from:

- Multi-Modal Integration: Combining tabular, graph, and sequential data captures diverse app characteristics.
- Advanced Architectures: Transformer and GNN modules extract complex patterns.
- Dynamic Attention: Enhances focus on relevant features (Equation 1).

Compared to DREBIN's 94% accuracy, MAGNET offers a 3.24% improvement, surpassing other multi-modal and GNN-based methods. Its practical deployment potential lies in its high precision (0.980) and low false positive rate (1.58%).

Limitations include:

- Computational Complexity: Multi-modal processing demands significant resources.
- Data Dependency: Performance relies on quality feature extraction.
- Generalizability: DREBIN's 2010–2014 data may limit applicability to newer malware.

VII. Conclusion

MAGNET advances Android malware detection through a multi-modal framework, achieving 97.24 ± 0.65% accuracy and 0.9823 ± 0.0042 F1-Score on the DREBIN dataset. Its integration of EnhancedTabTransformer, GraphTransformer, and SequenceTransformer, coupled with dynamic attention (Equation 1) and PIRATES optimization, ensures robust performance.

VIII. Future Work

- Evaluate on newer datasets to enhance generalizability.
- Optimize computational efficiency for resource-constrained devices.

- Develop interpretable models to elucidate decision-making.
- Investigate resilience against adversarial attacks.

IX. Acknowledgments

The authors thank Shahid Bahonar University for computational resources. [Funding details placeholder, please provide if applicable.]

X. Conflict of Interest

The authors declare no conflicts of interest.