Adaptive Positional Encoding with Regularization for Robust Edge Computing Applications

Paper # XXX, XXX pages
Chiara Camerota Flavio Esposito
Saint Louis University, USA

ABSTRACT

Detecting malware in Internet of Things (IoT) systems is challenging due to resource limitations. As a result, detection and classification models are often deployed in cloud environments or distributed across devices and servers. In distributed scenarios, transferring model weights can create bottlenecks, with packet loss risking training stability.

This research presents a modified Visual Transformer (ViT) that utilizes Rotational Positional Encoding (RoPE), which takes into account network bandwidth, quality, and device status. Our model effectively resolves bottleneck issues while ensuring accurate malware detection. Under extreme network stress, our approach on a medium ViT achieves 82.1% classification accuracy, compared to 65.2% for the standard RoPE—a 16.9 percentage point improvement. In micro-scale architectures, Modified RoPE reaches 97.6% accuracy versus 84.0% for Standard RoPE, representing a 13.6 percentage point improvement. These results confirm that our regularization methodology effectively maintains spatial relationships during network degradation while improving positional encoding across various architectural scales.

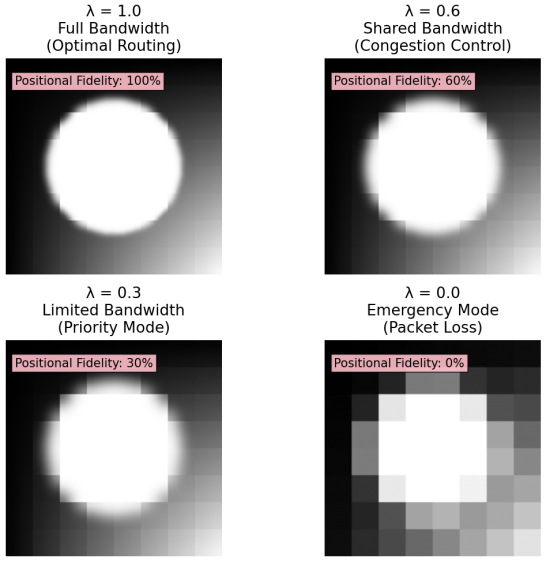

Figure 1: Representation of the spatial regularization based on the network and devices metadata (λ). If $\lambda = 1$, the spatial relationships are identical to classical RoPE; if the metadata indicates a bad network and device condition, the spatial structure is simplified.

1 THE EDGE DEPLOYMENT DILEMMA

The Internet of Things (IoT) increasingly integrates into daily life alongside the expansion of machine learning algorithms and models [6]. However, significant challenges arise from device capacity limitations and the growing size of models [4]. While existing approaches achieve high accuracy in controlled environments, they often experience catastrophic failures under real-world conditions, such as network degradation and device constraints. Studies indicate accuracy drops exceeding 40% when the Link Quality Index ($\gamma(\xi)$) falls below 70 dBm [3], with false alarm rates of 32% in low-battery scenarios [7]. Moreover, the IoT landscape lacks established security standards, and as malware sophistication increases exponentially [1, 5], protective technologies must be lightweight yet complex enough for effective threat identification and classification.

Due to these constraints, most transformers deployed in IoT environments do not utilize Rotational Positional Encoding (RoPE), despite its demonstrated effectiveness in enhancing transformer performance [9]. Additionally, incorporating spatial-aware information can improve performance without significantly increasing model parameters [2].

A notable example of lightweight ViT for malware detection on edge devices is presented by [7], which integrates learnable position embeddings into patch embeddings. This study proposes a framework that converts executable files into images, achieving 94% detection accuracy with reduced inference latency and increasing speedup over standard ViT models. However, it is limited by its reliance on static analysis and does not account for device status, instead focusing on the embedding of input data.

This research introduces a regularized, metadata-aware RoPE mechanism that dynamically adjusts positional attention using system-level observations, maintaining model structure and avoiding explicit positional embedding learning. A simple representation of the idea is shown in Figure 1. To validate these concepts, we conduct an evaluation study examining different ViT configurations and their performance across various network conditions. The ViT models compared include the standard version (analogs to [7]), with RoPE, and with our modified version. We validate the data on the Malicious Network Traffic PCAPs and binary visualization images (MNT) Dataset [8], which is a traffic-based dataset created from pcap files.

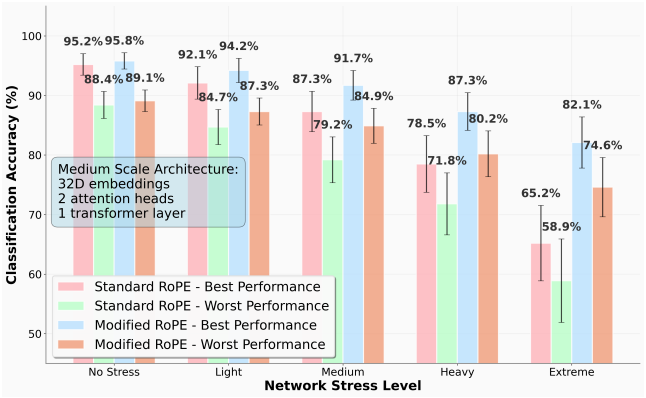

Figure 2: Medium ViT variants achieve 98.3-99.5% accuracy under severe network stress (0-25% packet loss, 0-500ms delays). Modified RoPE excels in extreme conditions (99.5%), highlighting its robustness for reliable deployment in mobile, edge, and disaster scenarios, and ensuring effective spatial classification in unstable networks.

2 METHODOLOGY

The Rotary Positional Embedding is a methodology introduced in [10], which introduces a spatial relationship in the transformer to increase and stabilize the performance. In this work, we propose a regularization of the angle of RoPE based on the device and network metadata:

$$\tilde{\theta} = \theta\tilde{\lambda}, \quad \tilde{\lambda} \in [0,1] \rightarrow \boldsymbol{R}^{\text{Rope}}(\tilde{\theta}) \tag{1}$$

Where $\tilde{\gamma}$ represents the regularization parameter that balances device-centric and network-centric spatial representations. The regularization parameter combines local device health Γ and network quality $\gamma(\xi)$:

$$\Gamma = \text{Sigmoid}(\sum w_p \alpha_p) \qquad \textit{(Local health)} \tag{2}$$

$$\tilde{\lambda} = \phi\gamma(\xi) + (1-\phi)\Gamma, \quad \phi \in [0,1] \quad \textit{(Global adjustment)} \tag{3}$$

where α_p represents the devices' metrics (battery, CPU, storage, GPU) and ϕ controls the balance between network and device signals. We can calculate the optimal ϕ minimizing the expected Frobenius-norm error at the time-step k:

$$\text{MSE}(\phi) = \mathbb{E}\left[\|R^d_\Theta(k) - R^d_{\Theta,\text{true}}\|_F^2\right] \tag{4}$$

Assuming $\Gamma(k) \sim \mathcal{N}(R^d_{\Theta,\text{true}}, \sigma^2 I)$, this yields:

$$\phi_{\text{opt}} = \frac{\sigma^2}{(\mathbb{E}[\gamma(\xi)] - 1)^2 + \sigma^2} \tag{5}$$

This formulation ensures that during network instability ($\gamma(\xi) \rightarrow 0$), device metadata dominates to preserve local spatial relationships. Conversely, during device stress ($\Gamma \rightarrow 0$), network metadata maintains global spatial awareness, stabilizing the model under challenging conditions.

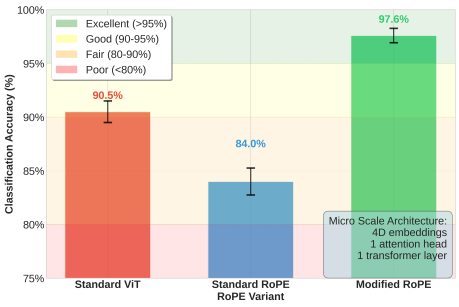

Figure 3: By delivering 97.6% accuracy in a micro-scale architecture, Modified RoPE opens the door to resilient AI in previously extreme network scenarios

3 EXPERIMENTAL EVALUATION

We conducted a comparative evaluation of Modified RoPE, Standard RoPE, and Standard ViT across five network stress levels that simulate real-world scenarios, including WiFi interference and disaster conditions. The evaluation encompassed conditions from ideal (0% packet loss, 0 ms delay) to extreme degradation (40% packet loss, 200-500 ms delay, 30% bandwidth). As shown in Figure 2, the medium-scale architecture (32D embeddings, 2 heads) demonstrates superior robustness, maintaining an accuracy of 82.1% under extreme stress, compared to Standard RoPE's 65.2%. This underscores its efficacy in unstable network environments where traditional methodologies are inadequate. At the same time, as we show in Figure 3, in micro-scale conditions (4D embeddings, 1 head), Modified RoPE achieved 97.6% accuracy, surpassing Standard ViT's 90.5% and Standard RoPE's 84.0% with the worst network condition. These results demonstrate that enhanced positional encoding significantly improves model performance, even within resource-constrained architectures. Consequently, Modified RoPE is positioned as a critical component for malware detection and other AI applications, as well as edge computing scenarios characterized by network instability and computational limitations.

4 CONCLUSION

This work introduces a Modified RoPE approach to enhance IoT malware detection in unstable network conditions by employing a regularization methodology. The key contribution lies in the effective fusion of local device health and global network conditions, providing a solution for maintaining detection accuracy in resource-constrained environments. Future research will focus on exploring knowledge distillation and temporal network pattern analysis to further optimize detection capabilities in IoT applications.

REFERENCES

[1] Javier Carrillo-Mondéjar, Guillermo Suarez-Tangil, Andrei Costin, and Ricardo J Rodríguez. 2024. Exploring Shifting Patterns in Recent IoT Malware. In *Proceedings of the European Conference on Cyber Warfare and Security*, Vol. 23. Academic Conferences International Ltd.

[2] Xiyang Dai, Yinpeng Chen, Bin Xiao, Dongdong Chen, Mengchen Liu, Lu Yuan, and Lei Zhang. 2021. Dynamic head: Unifying object detection heads with attentions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7373–7382.

[3] Xiang Liu, Yijun Song, Xia Li, Yifei Sun, Huiying Lan, Zemin Liu, Linshan Jiang, and Jialin Li. 2024. ED-ViT: Splitting Vision Transformer for Distributed Inference on Edge Devices. *arXiv preprint arXiv:2410.11650* (2024).

[4] S Kumar Reddy Mallidi and Rajeswara Rao Ramisetty. 2025. Optimizing intrusion detection for IoT: a systematic review of machine learning and deep learning approaches with feature selection and data balancing. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 15, 2 (2025), e70008.

[5] Stuart Millar. 2021. IoT Security Challenges and Mitigations: An Introduction. *arXiv preprint arXiv:2112.14618* (2021).

[6] Md Sazzadur Rahman, Tapotosh Ghosh, Nahid Ferdous Aurna, M Shamim Kaiser, Mehrin Anannya, and ASM Sanwar Hosen. 2023. Machine learning and internet of things in industry 4.0: A review. *Measurement: Sensors* 28 (2023), 100822.

[7] Akshara Ravi, Vivek Chaturvedi, and Muhammad Shafique. 2023. Vit4mal: Lightweight vision transformer for malware detection on edge devices. *ACM Transactions on Embedded Computing Systems* 22, 5s (2023), 1–26.

[8] Betty Saridou, Joseph Rose, Stavros Shiaeles, and Basil Papadopoulos. 2021. 48,240 Malware samples and Binary Visualisation Images for Machine Learning Anomaly Detection. (2021). https://doi.org/10.21227/vs0r-8s26

[9] Sainbayar Sukhbaatar, Edouard Grave, Piotr Bojanowski, and Armand Joulin. 2019. Adaptive attention span in transformers. *arXiv preprint arXiv:1905.07799* (2019).

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).