FOM - University of Applied Science
Faculty of Computer Science and Business Informatics
University Center Munich

Term paper

The Silent Threat in DNS: A Hybrid Detection and Defense Concept against Cybersquatting and DGAs

presented in the module
Security Analytics & Defense

of the degree program
Cyber Security Management

Johannes Jacob Schneider

Primary Reviewer: Prof. Dr. Michael Colombo
Second Reviewer: Prof. Dr. Oliver Koch
Submitted on: July 13, 2025
Submission deadline: July 30, 2025

Contents

Acknowledgments

Special thanks are extended to Simon Ofner and the team of Fraunhofer FKIE for the generous provision of the DGArchive dataset, which served as a essential resource for this research. Access to this dataset enabled us to use state-of-the-art DGA data.

Danksagung

Besonderer Dank gilt Daniel Plohmann und dem Team vom Fraunhofer FKIE für die umfassende Bereitstellung des DGArchive-Datensatzes, das eine wesentliche Grundlage dieser Forschungsarbeit darstellte. Der Zugang zu diesem Datensatz ermöglichte die Forschung anhand aktuellster DGA-Daten.

Abstract

The Domain Name System (DNS) is a critical internet component, but its central role renders it a prime target for abuse. Malicious actors exploit DNS through techniques like Domain Generation Algorithms (DGA) and cybersquatting to enable large-scale attacks. These methods often rely on distinct domain name patterns, making them ideal for analysis by LLMs. Although previous research has successfully applied LLMs to detect these threats in isolation, little research has addressed the multi-class classification of legitimate, DGA and cybersquatting domains within a single model. To address this, we constructed a balanced dataset of 30,000 domains, combining legitimate entries from 'Majestic 1 Million', DGA samples from 'DGArchive', and synthetically generated cybersquatting domains. We then fine-tuned the character-level 'ByT5-base' transformer model, which is particularly adept at handling the raw, pattern-based nature of domain strings. Finally we developed an architectural concept to integrate the model into production environments using the Shared Signals Framework. The trained model achieved a robust overall accuracy of 86.2% and a macro-averaged F1-Score of 0.86 for the described dataset. It demonstrated exceptional reliability in identifying legitimate (0.96 recall) and DGA (0.95 recall) domains, which is crucial for minimizing operational disruption and detecting threats.

Zusammenfassung

Das Domain Name System (DNS) ist eine kritische Komponente des Internets. Aufgrund seiner zentralen Rolle ist es ein bevorzugtes Angriffsziel für böswillige Akteure. Dabei werden Angriffstechniken wie Domain Generation Algorithms (DGA) und Cybersquatting eingesetzt, um großangelegte Angriffe zu ermöglichen. Diese Methoden weisen häufig charakteristische Muster in den Domainnamen auf. Dadurch sind sie für die Analyse durch LLMs besonders gut geeignet. Während bestehende Forschungsarbeiten LLMs erfolgreich zur isolierten Erkennung dieser Bedrohungen eingesetzt haben, gibt es bislang kaum Untersuchungen zur Multiklassen-Klassifikation legitimer, DGA- und Cybersquatting-Domains innerhalb eines einzigen Modells. Um dieser Lücke zu addressieren, wurde ein ausgewogener Datensatz bestehend aus 30.000 Domains erstellt. Dieser kombiniert legitime Einträge aus der "Majestic 1 Million", DGA-Beispiele aus dem "DGArchive" sowie synthetisch generierten Cybersquatting-Domains. Folgend wurde das zeichenbasierte Transformer-Modell "ByT5-base" trainiert, das sich insbesondere durch seine Eignung zur Verarbeitung der rohen, musterbasierten Strukturen von Domain-Strings auszeichnet. Abschließend wurde ein Architekturkonzept zur Integration des Modells in Produktionsumgebungen mithilfe des Shared Signals Framework entwickelt. Das trainierte Modell erzielte beim beschriebenen Dataset eine "Accuracy" von 86,2% sowie einen "macro-averaged F1-Score" von 0,86. Es zeigte eine außergewöhnliche Zuverlässigkeit bei der Identifikation legitimer Domains (Recall: 0,96) und DGA-Domains (Recall: 0,95), was sowohl für die Minimierung von Betriebsunterbrechungen als auch für die effektive Erkennung von Bedrohungen von entscheidender Bedeutung ist.

Preface

The University of Siegen succinctly captures it: Language shapes the images in our heads and influences how we perceive the world around us (Universität Siegen, 2019). With this in mind, I would like to emphasize that predominantly masculine pronouns and nouns are used in this work. This choice is not meant to diminish the significance and presence of feminine and diverse gender identities. The preference for masculine forms is a convention aimed at enhancing readability and text comprehension, without implying a marginalization of other genders.

Vorwort

Die Universität Siegen bringt es treffend auf den Punkt. Sprache prägt die Bilder in unseren Köpfen und beeinflusst, wie wir die Welt um uns herum wahrnehmen (Universität Siegen, 2019). Vor diesem Hintergrund möchte ich betonen, dass in dieser Arbeit überwiegend maskuline Pronomen und Substantive genutzt werden. Die Bedeutung und Präsenz femininer und diverser Geschlechteridentitäten soll dadurch nicht geschmälert werden. Die Entscheidung für die maskuline Form ist eine Konvention zugunsten der Lesbarkeit und um die Textverständlichkeit zu erhöhen, ohne dabei eine Marginalisierung anderer Geschlechter zu implizieren.

List of Figures

List of Tables

Listings

List of Abbreviations

C2	Command and Control
CSV	Comma-Separated Value
DGA	Domain Generation Algorithms
DNS	Domain Name System
DoS	Denial of Service
FN	False Negatives
FP	False Positives
LLM	Large Language Model
LoRA	Low-Rank Adaption
SIEM	Security Information and Event Management
SOAR	Security Orchestration, Automation, and Response
SOC	Security Operations Center
SSF	Shared Signals Framework
TN	True Negatives
TP	True Positives

1 Introduction

The Domain Name System (DNS) constitutes a core component of the modern internet, responsible for translating human readable domain names into IP addresses. While fundamental for global connectivity, its centrality also renders it to a high value target for adversaries. Two prevalent threats, though distinct in purpose, exploit DNS in structurally similar ways. Domain Generation Algorithms (DGA) and cybersquatting domains undermine the integrity of DNS and repurpose its functionality for malicious activities such as large scaled Denial of Service (DoS) attacks, email spamming or phishing based data theft (Plohmann et al., 2016; Bronjon and Tasiruddin, 2023).

In the case of DGAs, attackers utilize pseudo-random generated domain names to dynamically rotate Command and Control (C2) communication channels with infected devices. This strategy renders static domain blacklists largely ineffective (Plohmann et al., 2016). Cybersquatting by contrast, involves the registration of domains names that resemble legitimate brands or keywords, often relying on user error or visual deception (Marchal et al., 2014). Both threat vectors share a common reliance on the characteristics of domain patterns, highlighting their suitability for character-level analysis by modern Large Language Models (LLMs) in order to strengthen existing cyber defense frameworks.

2 Related Work

Several previous studies have demonstrated that the application of LLMs significantly enhance the detection of DGAs and cybersquatting domains (Tuan et al., 2023). Especially since traditional domain classifications largely depended on manual feature engineering and statistical techniques, which often imposed limitations on both flexibility and scalability (Tuan et al., 2023). Furthermore, as time progressed, academic methodologies evolved from binary to more sophisticated multi-class classifications (Sayed et al., 2024).

The authors of Sayed et al. fine-tuned an LLM for detecting DGAs and DNS-based data exfiltrations and achieved a multi-class classification accuracy of 77% across a dataset comprising 59 DGA malware families (Sayed et al., 2024). Bronjon and Tasiruddin pre-trained the transformer model 'CANINE-c' for binary classification of domain names as either DGA or legitimate, solely based on the raw domain strings. Their approach achieved an accuracy and precision of 99% (Bronjon and Tasiruddin, 2023). Concurrently, research into cybersquatting detection has also seen promising advancements, with one of the most notable recent contributions being the 'DomainLynx System' proposed by Chiba et al. Their framework achieved

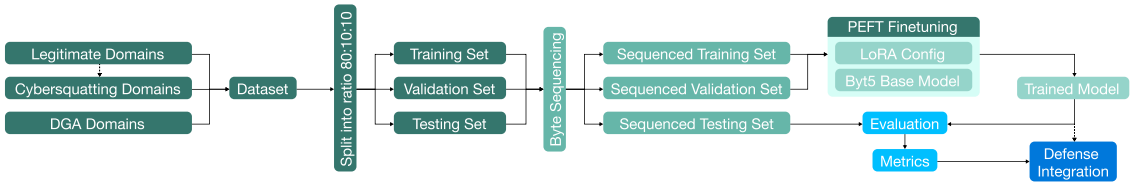
94.7% accuracy using LLaMA-3-70B, leveraging LLM-based threat scoring and semantic domain pairing (Chiba, Nakano, and Koide, 2024).

Despite growing interest in leveraging the capabilities of LLMs for advancements in cybersecurity, there is notable lack of research addressing the unified multi-class classification of DGAs, cybersquatting and legit domains within a single model. This paper seeks to close that gap by proposing a hybrid detection and response pipeline employing a lightweight transformer model and by conceptualizing its integration into established security architectures and platforms.

3 Approach

This section presents the methodological foundation of the proposed concept, detailing the dataset construction, model architecture selection and integration into operational security workflows. An overview of the proposed architecture, which is elaborated in detail in the subsequent sections, is provided in Figure 1 and Appendix A.1.

Figure 1: High-level architecture of the proposed model



Source: The author's, 2025 (based on Bronjon and Tasiruddin, 2023)

3.1 Dataset

The creation of a standardized and publicly accessible dataset remains a persistent challenge in this area of research (Tuan et al., 2023). Numerous datasets such as 'Alexa 1 Million, UMUDGA, UTL_DGA22, Bambenek' and more exist, yet vary considerably in terms of quality and scope (Kaggle, 2022; Zago, Gil Pérez, and Martínez Pérez, 2020; Tuan et al., 2023; Bambenek Consulting, 2019). To ensure scientific relevance and data integrity, we employed a hybrid dataset combining verified sources and synthetically generated samples. Full implementation details are provided in the *project repository* for transparency and reproducibility.

First, the top 20,000 **legit domains** were extracted and filtered from the 'Majestic 1 Million' dataset, which has increasingly replaced the deprecated 'Alexa 1 Million' list as a research standard (Xie and Li, 2024; Jones, 2016; Yilmaz, Siraj, and

Ulybyshev, 2021). The resulting records were stored in a structured JSONL and Comma-Separated Value (CSV) format as shown in Listing 1.

Listing 1: An excerpt from the 20.000 legitimate domains dataset, demonstrating the JSONL format used


```jsonl
{"id": 0, "domain": "google.com", "classification": "legit"}
{"id": 1, "domain": "facebook.com", "classification": "legit"}
{"id": 2, "domain": "youtube.com", "classification": "legit"}
{"id": 3, "domain": "twitter.com", "classification": "legit"}
{"id": 4, "domain": "instagram.com", "classification": "legit"}
{"id": 5, "domain": "linkedin.com", "classification": "legit"}
{"id": 6, "domain": "microsoft.com", "classification": "legit"}
{"id": 7, "domain": "apple.com", "classification": "legit"}
```


Second, 20,000 **Cybersquatting domains** were synthetically generated by applying established transformation techniques (e.g., leetspeak, homoglyphs, typosquatting, etc.) to the previously created legit domains subset (Chiba, Nakano, and Koide, 2024). Detailed descriptions of the applied techniques are provided in Appendix A.2. Consistent with the previous step, the entries were again stored in structured JSONL and CSV formats, as presented in Listing 2.

Listing 2: An excerpt from the 20.000 Cybersquatting domains dataset, demonstrating the JSONL format used


```jsonl
{"id": 9, "target_domain": "samsung.com", "domain": "5am$ung.com", "classification": "squat", "squat_type": "leetspeak"}
{"id": 3900, "target_domain": "maps.google.com", "domain": "maps.online", "classification": "squat", "squat_type": "tld_swap"}
{"id": 4477, "target_domain": "cloud.microsoft", "domain": "cloud-login.microsoft", "classification": "squat", "squat_type": "keyword_insertion"}
{"id": 10027, "target_domain": "fiverr.com", "domain": "fiverrr.com", "classification": "squat", "squat_type": "extra_char"}
{"id": 17910, "target_domain": "dropbox.com", "domain": "dropbox.com", "classification": "squat", "squat_type": "homoglyph"}
```


Third, for **DGA domains**, the 'DGArchive' dataset from the Fraunhofer Institute was requested and used. This dataset was selected, due to its recency and comprehensive coverage of 137 malware families up to late 2024 (Plohmann et al., 2016). Individual CSV files, each representing a malware family, were aggregated into a unified dataset. Stratified sampling was also applied to extract 20,000 representative domains, ensuring proportionality across different malware families. Again, the resulting entries were stored in a structured JSONL and CSV format as illustrated in Listing 3.

Listing 3: An excerpt from the 20.000 DGA domains dataset, demonstrating the JSONL format used


```jsonl
{"id": 134, "domain": "1egzjd993sb7vbz7rw1bovztj.com", "family": "gameover", "classification": "dga"}
{"id": 133, "domain": "1w9omdmkj852lw45s2k1vy755t.com", "family": "gameover", "classification": "dga"}
```

```
{"id": 12063, "domain": "storyouter.ru", "family": "suppobox", "classification": "dga"}
{"id": 12952, "domain": "thoughtapple.net", "family": "suppobox", "classification": "dga"}
{"id": 8246, "domain": "btkvcgsajmxaezfrovz.info", "family": "qakbot", "classification": "dga"}
{"id": 8247, "domain": "dvfygfaqarkxxkgiljqeu.org", "family": "qakbot", "classification": "dga"}
```

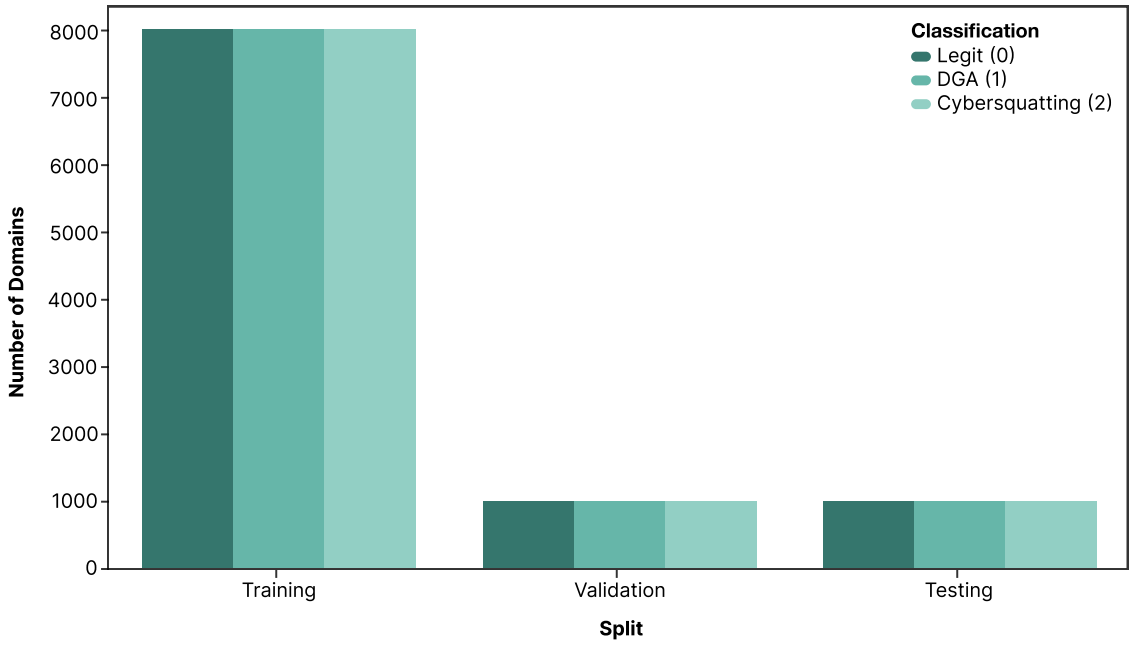

Finally, the three individually crafted datasets were merged into one corpus consisting of 30,000 equally distributed domains across the three established classes: 'Legit, DGA, Cybersquatting'. Class labels were encoded as outlined in Table 1, following a stratified split of the dataset into 80% training, 10% validation, and 10% testing subsets, as illustrated in Figure 2.

Table 1: Definition of the classification labels used in the unified dataset

Label	Classification	Example
0	Legitimate Domain	apple.com
1	DGA Domain	thoughtapple.net
2	Cybersquatting Domain	fiverr.com

Source: The author's, 2025

Figure 2: Distribution of the domains across the training, validation, and testing splits, broken down by classification label



Source: The author's, 2025

The unified dataset entries were finally saved again to JSONL and CSV format with the following structure, as demonstrated in Listing 4.

Listing 4: An excerpt from the 20.000 unified domains dataset, demonstrating the JSONL format used

```jsonl
{"id": 38, "domain": "amazon.nl", "classification": 0, "split": "train"}
{"id": 8234, "domain": "nintendo.co.jp", "classification": 0, "split": "val"}
{"id": 9160, "domain": "fotosearch.com", "classification": 0, "split": "test"}
{"id": 16968, "domain": "humanearth.ru", "classification": 1, "split": "train"}
{"id": 18776, "domain": "hpwiwj.com", "classification": 1, "split": "val"}
{"id": 19584, "domain": "knowfind.net", "classification": 1, "split": "test"}
{"id": 20589, "domain": "cianet.org", "classification": 2, "split": "train"}
{"id": 28664, "domain": "ubisoft.shop", "classification": 2, "split": "val"}
{"id": 29291, "domain": "amazno.jp", "classification": 2, "split": "test"}
```

This curated dataset served as the foundation for the model training and evaluation presented in the following Sections.

3.2 Model Choice

To conduct the experiments, we decided to use a controlled *Google Colab* environment utilizing a dedicated A100 GPU. This approach, in conjunction with Low-Rank Adaption (LoRA) enabled enhanced training efficiency while significantly reducing memory consumption. After initial assessments with token-based transformer models ('DistilBERT Tiny & Base') we observed that character-level representations demonstrated key advantages for raw domain classification tasks. Notably, they exhibit improved accuracy in handling obfuscation and algorithmically generated strings.

Consequently, we adopted the 'ByT5-base' model, which operates directly on byte-level input (Google Research, 2025). Featuring approximately 582 million parameters, a 12 layer transformer architecture and a hidden size of 768, it significantly outperformed its smaller sibling ('ByT5-small') with only 300 million parameters in domain classification tasks (Google Research, 2025). For full transparency, the final trained model is available in the *project's repository*.

3.3 Evaluation Metrics

To systematically assess the performance of the trained LLM on the domain classification task, a set of standard evaluation metrics was employed. These metrics are derived from the numbers of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) and are calculated as follows:

- **Accuracy**: Measures the overall fraction of correct predictions across all classes.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

- **Precision**: Quantifies the proportion of positive predictions that were actually correct.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

- **Recall**: Measures the proportion of actual positives that were correctly identified by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

- **F1-Score**: The harmonic mean of Precision and Recall, providing a single metric that balances both.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

To provide a comprehensive performance metric in the multi-class classification setting, we primarily report the **macro average**. It is computed as the unweighted mean of 'precision, recall and F1-Score' across all classes treated equally. Furthermore, since the crafted dataset is balanced (1,000 instances per class), the macro and weighted averages are effectively identical. Additionally the micro average is omitted, as it offers limited value in balanced datasets and is more appropriate when class distributions are highly imbalanced.

The application of these metrics will be detailed in Section 4, where we present and analyze the results of our model evaluation.

4 Results

This chapter presents the evaluation results of the trained ByT5-base model on the held-out test set. The analysis begins with a quantitative performance assessment, followed by a qualitative examination of the model specific error patterns to gain deeper insights regarding classification behavior, overall effectiveness, and potential practical implications.

4.1 Training Dynamics

The model achieved an overall accuracy of 86.2% and a macro-averaged F1-Score of 0.852 over 20 training epochs. Performance gains plateaued after epoch 15, suggesting that earlier stopping could yield comparable results while reducing training time and resource consumption. A detailed visualization is presented in Figure 3. Concurrently, both training and validation losses decreased steadily, indicating a stable convergence and no signs of overfitting.

Figure 3: Model performance across training epochs, demonstrating trends in accuracy and macro-averaged F1-score

Source: The author's, 2025

4.2 Overall Classification Performance

These metrics demonstrate a robust and balanced performance across the three distinct domain classes, supporting the overall effectiveness of the proposed approach. As shown in Table 2, the model performs best on legit and DGA domains. However, the comparatively lower recall of 0.67 for cybersquatting domains suggests that certain manipulative patterns are difficult to detect.

Table 2: Detailed classification report for the model on the test dataset

Class	Precision	Recall	F1-Score
DGA	0.76	0.95	0.84
Legit	0.95	0.96	0.96
Cybersquatting	0.92	0.67	0.78
Macro Average	0.87	0.86	0.86
Weighted Average	0.87	0.86	0.86

Source: The author's, 2025

The per-class results highlight distinct performance characteristics. For the 'Legit' class the model demonstrates exceptional reliability, achieving a precision of 0.95 and a recall of 0.96. This is crucial for real-world deployment in enterprise environments, as it minimizes the risk of falsely blocking legitimate DNS traffic and thereby reducing operational disruptions.

In contrast, the 'DGA' class similarly shows a high recall of 0.95 but a low precision of 0.76. In the context of cybersecurity, this implies that the model is effective at accurately detecting the majority of DGA-based threats. However, the relatively low precision of 0.76 indicates a tendency for generating a considerable number of False Positives by misclassifying non-DGA domains as malicious. This contributes to the already high alert fatigue on Security Operations Center (SOC) analysts.

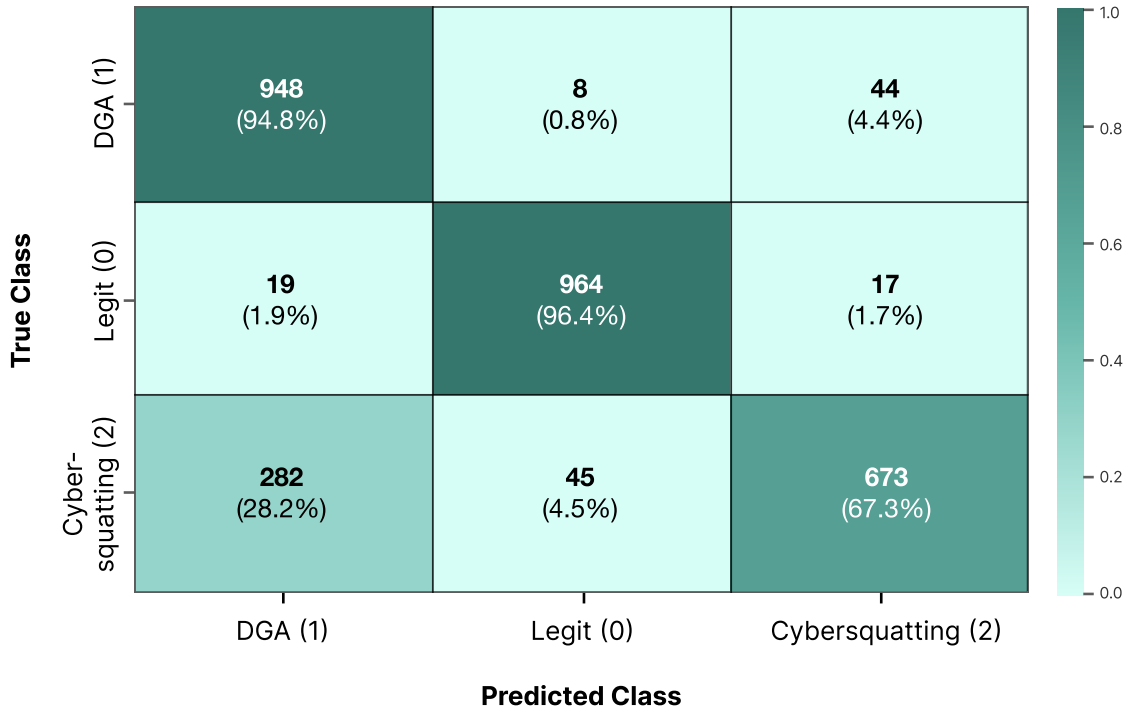
The 'Cybersquatting' class exhibits an inverse pattern, characterized by a high precision of 0.92 and a recall as low as 0.67. This means that when the model classifies a domain as cybersquatting, it is highly likely to be correct. Conversely, it fails to detect approximately one-third of all cybersquatting instances, revealing a clear area for improvement.

To break down the underlying causes of these performance trade-offs, particularly the misclassifications between the two malicious classes (28.2% of cybersquatting domains were labeled as DGA), a qualitative analysis of the error patterns is necessary. This can be effectively facilitated through the use of a confusion matrix as presented in Figure 4.

The models strength lies in the reliable identification of DGA domains. However, a notable limitation is the misclassification of 282 cybersquatting domains as DGAs. This behavior is likely attributed due to the architecture of the ByT5 model. As a byte-level transformer it is inherently sensitive to subtle, non-semantic character patterns. While this trait enhances its ability to detect algorithmic irregularities inherent in many DGA families, it may also cause the model to interpret the intentional misspellings and character manipulations as analogous statistical anomalies.

To illustrate this phenomenon, we identified representative misclassified examples in Table 3.

Figure 4: Confusion Matrix and performance of the three classes in absolute numbers and percentages



Source: The author's, 2025

Table 3: Examples of model misclassifications

Domain	True Label	Predicted Label	Potential Reason for Error
goog1e.c0m	Cybersquatting	DGA	Interpreted as random characters
fiverrr.com	Cybersquatting	DGA	Unusual letter combination, appears random
thoughtapple.net	DGA	Legit	Two valid English words appearing plausible

Source: The author's, 2025

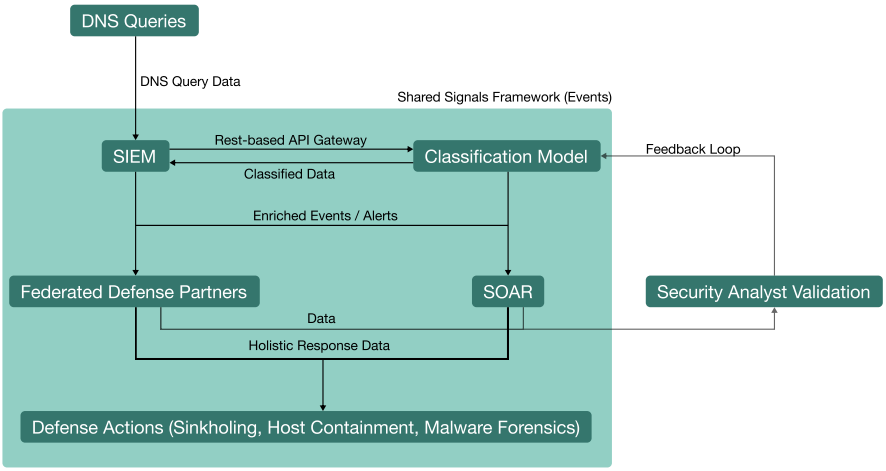
In summary, the results demonstrate that the proposed model serves as a powerful tool for distinguishing malicious domains from legitimate DNS traffic. The primary limitation remains in the nuanced differentiation between DGAs and sophisticated cybersquatting variants. This detailed understanding of the model is fundamental for designing an effective defense integration strategy.

5 Defense Integration

To implement the classification model into an existing security infrastructure composed of Security Information and Event Management (SIEM) and Security Orchestration, Automation, and Response (SOAR) platforms, we propose an encapsulated REST-based microservice approach. The model processes real-time DNS query data from the SIEM to enrich events and trigger correlation rules.

A central element of this architecture - as presented in Figure 5 - is the use of the Shared Signals Framework, which provides a standardized and interoperable communication layer between security entities. To reduce False Positives, only high-confidence detections (e.g. >90%) are encoded as Shared Signals Framework (SSF) events and transmitted to internal enforcement systems and federated defense partners. This enables a comprehensive defense posture.

Figure 5: High-level defense integration of the DNS classification model into security architecture platforms with the Shared Signals Framework



Source: The author's, 2025

This architecture facilitates the proactive sharing of threat intelligence throughout the defense stack, enabling the detection of large-scale domain abuse patterns. Moreover, enriched alerts are forwarded to the SOAR platform to automatically

initiate predefined playbook actions. For instance, cybersquatting detections may trigger domain takedown requests or temporary DNS sinkholing, while DGA detections prompt host containment or malware forensics.

Finally, a feedback loop is established and closed via analyst validation. To this end, flagged domains and their corresponding labels are periodically reintegrated into the training set. This process ensures continuous model recalibration and performance enhancements, especially with regard to the accurate detection and classification of DGA and cybersquatting activities.

Overall, this integration concept transforms static and isolated domain detection into an adaptive, holistic, and collaborative defense mechanism.

6 Conclusion & Future Research

This paper outlines an architectural concept to detect and defend against DNS-based threats, by leveraging a LLM and integrating it into a collaborative security infrastructure using the Shared Signals Framework. The achieved results reflect the models capabilities to distinguish between the three domain types with an overall accuracy of 86.2% and a macro-averaged F1-score of 0.86. Moreover the approach confirms robustness in distinguishing between DGA and legitimate domains while revealing areas of improvement in DGA and cybersquatting differentiation.

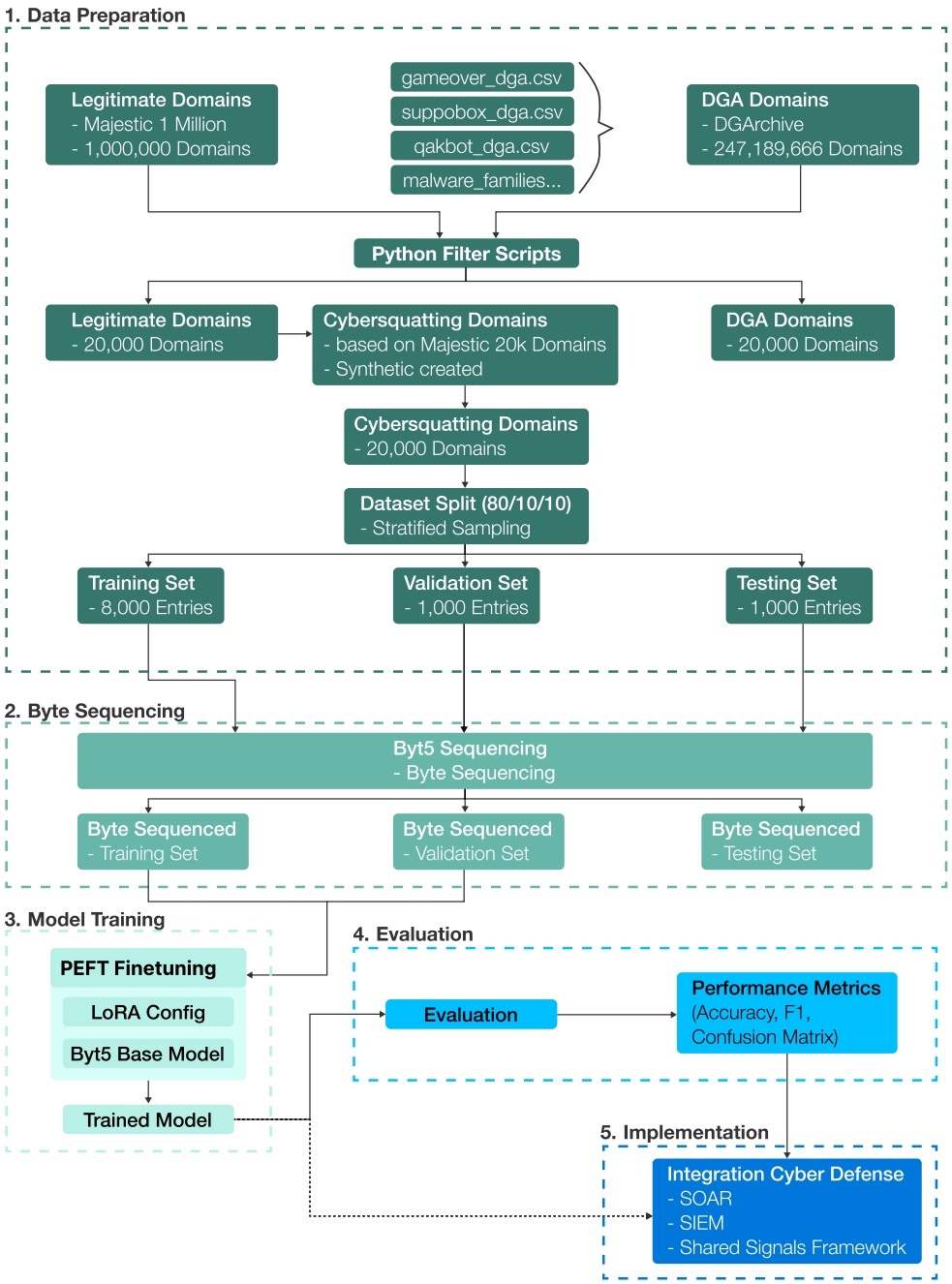
By embedding the model into SIEM and SOAR platforms via a microservice architecture and the Shared Signals Framework, real-time enrichment of security events, federated intelligence sharing and automated response actions were enabled. This design evolved static detection into an adaptive and collaborative defense mechanism.

However, several limitations warrant further research. Primarily due to the length constraints of this paper the practical implementation of the proposed model remains open and theoretical. Integrating the trained model into operational environments and further tuning it with real-world business data could yield valuable insights. Furthermore, future work should focus on improving the classification accuracy between DGA and complex cybersquatting domains. Longitudinal studies covering emerging domain abuse techniques, incorporate more diverse languages and assess cost-effectiveness in production environments are also recommended.

In summary this paper confirms the considerable potential of the proposed technology to combat modern DNS-based attacks, while also outlining key next steps for further research and exploration.

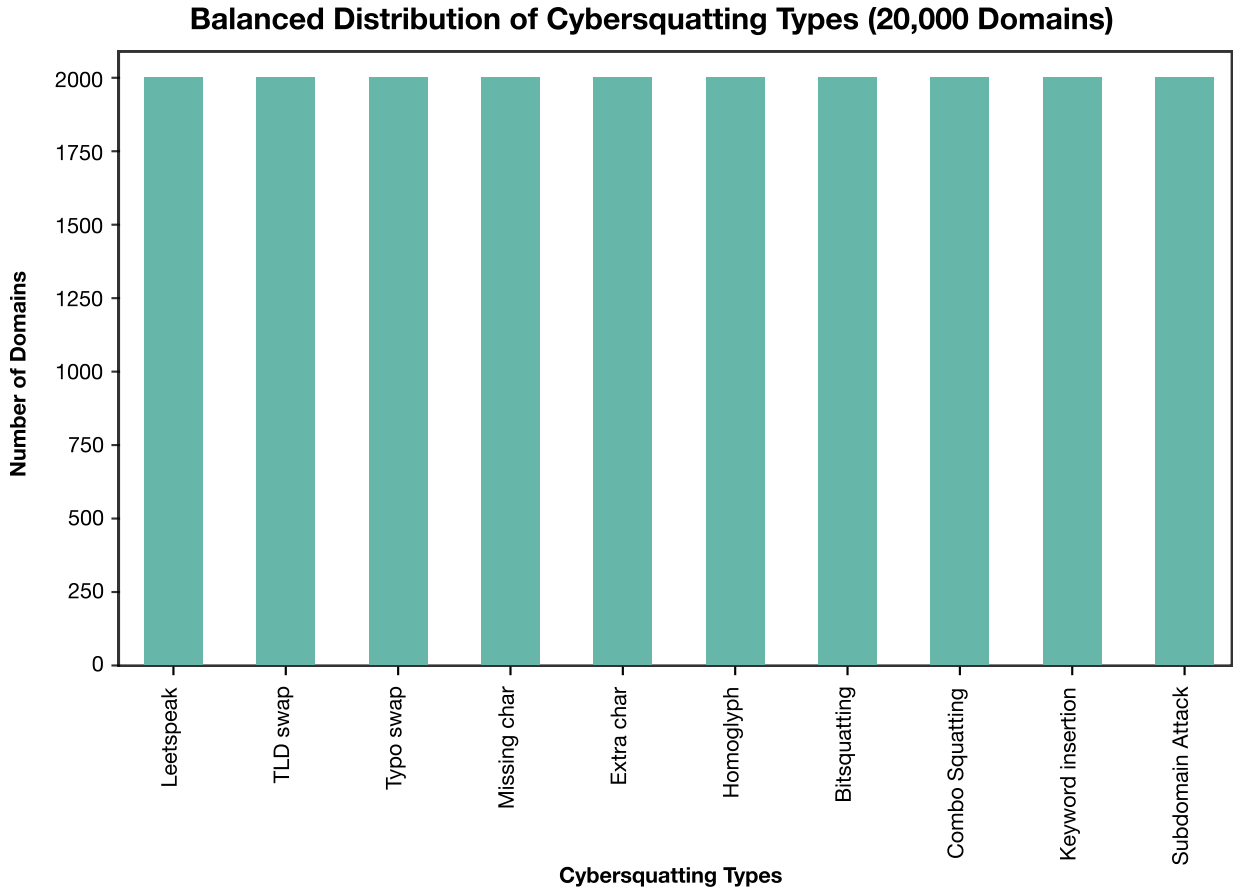
Appendix

A.1 Detailed architecture of the proposed model



Source: The author's, 2025

A.2 Distribution details of the applied Cybersquatting Types



Source: The author's, 2025

Bibliography

Bronjon, Gogoi and Ahmed Tasiruddin (June 2023). “DGA domain detection using pretrained character based transformer models”. In: *2023 IEEE Guwahati Subsection Conference (GCON)*. DOI: 10.1109/GCON58516.2023.10183602. URL: https://ieeexplore.ieee.org/document/10183602/metrics (visited on 06/16/2025).

Chiba, Daiki, Hiroki Nakano, and Takashi Koide (2024). *DomainLynx: Leveraging Large Language Models for Enhanced Domain Squatting Detection*. en. URL: https://arxiv.org/html/2410.02095v1 (visited on 06/17/2025).

Marchal, Samuel, Jérôme François, Radu State, and Thomas Engel (2014). “PhishStorm: Detecting Phishing With Streaming Analytics”. In: *IEEE Transactions on Network and Service Management* 11.4, pp. 458–471. ISSN: 1932-4537. DOI: 10.1109/TNSM.2014.2377295. URL: https://ieeexplore.ieee.org/document/6975177 (visited on 06/16/2025).

Plohmann, Daniel, Fraunhofer Fkie, Khaled Yakdan, Michael Klatt, Johannes Bader, and Elmar Gerhards-Padilla (2016). “A Comprehensive Measurement Study of Domain Generating Malware”. en. In: Austin, TX. ISBN: 978-1-931971-32-4. URL: https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_plohmann.pdf (visited on 06/16/2025).

Sayed, Abu, Asif Rahman, Christopher Kiekintveld, and Sebastián García (2024). *Fine-tuning Large Language Models for DGA and DNS Exfiltration Detection*. en. URL: https://arxiv.org/html/2410.21723v1 (visited on 06/17/2025).

Tuan, Tong Anh, Nguyen Viet Anh, Tran Thi Luong, and Hoang Viet Long (2023). “UTL_DGA22 - a dataset for DGA botnet detection and classification”. en. In: *Computer Networks* 221. ISSN: 13891286. DOI: 10.1016/j.comnet.2022.109508. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389128622005424 (visited on 06/16/2025).

Xie, Qinge and Frank Li (2024). “Crawling to the Top: An Empirical Evaluation of Top List Use”. en. In: *Passive and Active Measurement: 25th International Conference, PAM 2024, Virtual Event, March 11–13, 2024, Proceedings, Part I*. Berlin, Heidelberg: Springer-Verlag, pp. 277–306. ISBN: 978-3-031-56248-8. DOI: 10.1007/978-3-031-56249-5_12. URL: https://doi.org/10.1007/978-3-031-56249-5_12.

Yilmaz, Ibrahim, Ambareen Siraj, and Denis Ulybyshev (Jan. 2021). *Improving DGA-Based Malicious Domain Classifiers for Malware Defense with Adversarial Machine Learning*. en. arXiv:2101.00521 [cs]. DOI: 10.48550/arXiv.2101.00521. URL: http://arxiv.org/abs/2101.00521 (visited on 06/19/2025).

Zago, Mattia, Manuel Gil Pérez, and Gregorio Martínez Pérez (2020). “UMUDGA: A dataset for profiling algorithmically generated domain names in botnet detection”. en. In: *Data in Brief* 30. ISSN: 23523409. DOI: 10.1016/j.dib.2020.105400. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352340920302948 (visited on 05/31/2025).

Internet Sources

Bambenek Consulting (2019). *Bambenek Consulting Feed*. en. URL: https://osint.bambenekconsulting.com/feeds/ (visited on 06/18/2025).

Google Research (June 2025). *google-research/byt5*. original-date: 2021-05-26T17:37:59Z. URL: https://github.com/google-research/byt5 (visited on 06/20/2025).

Jones, Dixon (Nov. 2016). *Alexa top 1 Million sites is retired. Here's the Majestic Million.* URL: https://blog.majestic.com/development/alexa-top-1-million-sites-retired-heres-majestic-million/ (visited on 06/19/2025).

Kaggle (2022). *Alexa Top 1 Million Sites*. en. URL: https://www.kaggle.com/datasets/cheedcheed/top1m (visited on 06/18/2025).

Universität Siegen (2019). *Hinweise zur geschlechtergerechten Sprache*. URL: https://www.uni-siegen.de/gleichstellung/geschlechtergerechte_sprache/hinweise_geschlechtergerechte_sprache.pdf (visited on 06/16/2025).

Declaration under Oath

I hereby affirm that I have completed the registered examination assignment independently, without assistance from third parties, and have not used any sources or aids other than those specified in the examination assignment. I will clearly indicate any direct or paraphrased quotations, including AI-generated content.

At the time of submission, this examination assignment has not been previously submitted in the same or similar form, even in part, to any examination authority; exceptions apply only to assignments for which explicitly different regulations are stated in the module description.

I am aware that violating the content of this declaration constitutes an attempt to deceive, which will result in failure of the examination and may be prosecuted under criminal law according to §156 of the German Criminal Code (StGB). Furthermore, I am aware that in cases of severe misconduct, I may be expelled from the university and fined up to 50,000 EUR in accordance with the applicable examination regulations.

I consent to this examination assignment being uploaded to external providers' servers for plagiarism checking. The plagiarism check does not constitute public disclosure.

Munich, July 13, 2025

Location, Date