A Lightweight Entropy - Rate Framework for DNS Tunneling Detection in Institutional Networks

Santhosh K. M.

Lecturer, Department of Computer Engineering, Sree Rama Government Polytechnic College, Thriprayar, Kerala, India
Email: *santhkm[at]gmail.com*

Abstract: ***Domain Name System (DNS) is widely permitted through firewalls, making it a preferred channel for covert communication such as DNS tunneling. DNS tunneling enables attackers to exfiltrate data and establish command-and-control channels by embedding payloads into DNS queries and responses. Conventional detection methods often require deep packet inspection, heavy machine learning models, or large labelled datasets, limiting their applicability in resource-constrained institutional networks. This paper proposes a lightweight entropy–rate detection framework for identifying DNS tunneling behavior using short-term statistical indicators. The method computes query-string entropy, query-length deviation, unique domain ratio, NXDOMAIN ratio, and per-host query-rate anomalies within sliding time windows. These indicators are fused into a transparent anomaly score to detect both high-volume and stealthy tunneling attempts. The proposed approach is computationally efficient, explainable, and suitable for real-time deployment at campus gateways. Experimental evaluation demonstrates high detection accuracy with low false positives under mixed legitimate and malicious DNS traffic.***

Keywords: DNS security, DNS tunneling, entropy analysis, query-rate deviation, anomaly detection, lightweight IDS

1. Introduction

DNS is a fundamental service in modern networks, enabling domain-to-IP resolution for web browsing, email, cloud services, and mobile applications. Because DNS traffic is typically allowed by default, attackers exploit it to bypass security controls. DNS tunneling is a technique where data is encoded into DNS query names or responses, allowing covert channels for data exfiltration and command-and-control communication.

Contributions of this work include: (i) a compact entropy–rate feature set for DNS tunneling detection, (ii) a transparent anomaly scoring model with adaptive baselines, (iii) a lightweight algorithm suitable for real-time deployment, and (iv) evaluation under mixed benign and tunneling traffic.

Institutional networks such as colleges and small enterprises are particularly vulnerable due to limited security budgets and constrained monitoring infrastructure. While deep learning-based detection approaches exist, they often require extensive training data and computational resources. Similarly, signature-based detection fails against customized or obfuscated tunneling schemes.

This work presents a lightweight detection framework that combines entropy-based indicators with query-rate deviation and response-behavior metrics to identify tunneling activities in real time. By correlating multiple low-cost statistical features, the framework detects suspicious DNS patterns without requiring payload inspection or heavy model training.'

2. Literature Review

DNS tunneling detection research broadly falls into three categories:

2.1 Signature and Rule-Based Methods

Rule-based approaches use known patterns such as suspicious domains, uncommon record types, or blacklisted tunnel tools. These methods are fast but fail when attackers use new domains or change encoding strategies.

2.2 Statistical and Entropy-Based Detection

Statistical methods detect deviations in DNS query structure and distribution. Shannon entropy is commonly used to measure randomness in subdomain strings, which becomes higher when payloads are encoded using Base32/Base64-like patterns. However, entropy alone may produce false positives for legitimate content delivery networks (CDNs) or tracking domains.

2.3 Machine Learning Approaches

ML models classify DNS queries using lexical and behavioral features. Although accurate, they require labelled datasets and frequent retraining to handle evolving DNS patterns, making them less practical for small institutions.

Hence, lightweight hybrid approaches that combine entropy with rate and response behavior provide a good balance between accuracy and deployability. This paper follows that direction by proposing a deterministic scoring model.

2.4 Threat Model

The attacker is assumed to have access to an internal host and attempts to exfiltrate data or establish command-and-control by embedding encoded payloads in DNS query names. The defender monitors DNS metadata at the gateway and aims to detect tunneling without decrypting payloads or relying on prior signatures.

3. System Architecture

The proposed DNS tunneling detector operates as a modular pipeline with minimal overhead:

3.1 Traffic Capture Layer

DNS packets are captured at the gateway or monitoring point. Only lightweight metadata is required:

- Source IP (client)
- Queried domain name (QNAME)
- Query type (A, AAAA, TXT, etc.)
- Response code (NOERROR, NXDOMAIN)
- Timestamp

Traffic is processed using sliding windows (e.g., 1 second or 5 seconds).

3.2 Feature Extraction Module

For each time window, the detector computes:

- Query entropy
- Average query length
- Query rate per host
- Unique domain ratio
- NXDOMAIN ratio
- TXT query ratio (optional)

3.3 Hybrid Scoring Engine

Each feature is compared with baseline statistics. Deviations contribute to a total anomaly score.

3.4 Decision Layer

If the anomaly score exceeds an alert threshold, the system raises an alarm and logs suspicious clients/domains.

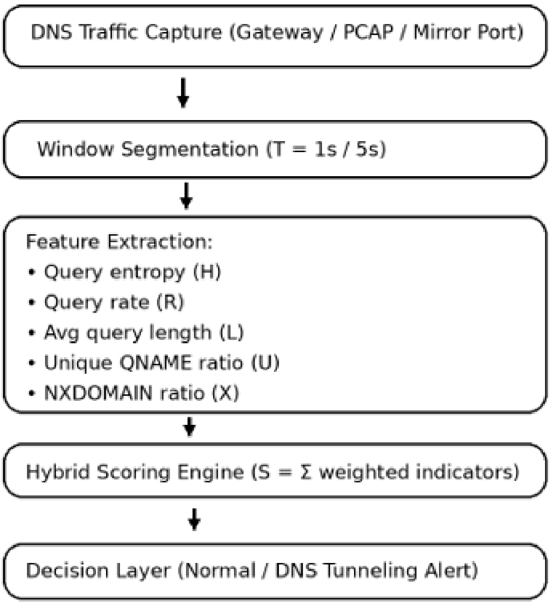
DNS Traffic Capture (Gateway / PCAP / Mirror Port)
Window Segmentation (T = 1s / 5s)
Feature Extraction:
• Query entropy (H)
• Query rate (R)
• Avg query length (L)
• Unique QNAME ratio (U)
• NXDOMAIN ratio (X)
Hybrid Scoring Engine (S = Σ weighted indicators)
Decision Layer (Normal / DNS Tunneling Alert)

Figure 1: Architecture of Lightweight DNS Tunneling Detector

4. Implementation Methodology

4.1 Sliding Window Processing

Traffic is segmented into windows of fixed duration **T** seconds. For each window **W**, queries are aggregated per client IP.

4.2 Lightweight Indicators

The following indicators are used:

(a) Shannon Entropy of Query Strings
For each query name, the subdomain part is analyzed. High entropy indicates encoded payloads.

$$H(q) = -\sum_{c\in C} p(c)\log_2 p\,(c)$$

Where p(c) is probability of character c in the query string.

(b) Query Rate
Number of DNS queries generated per client per second:

$$R(W) = \frac{N_q}{T}$$

(c) Query Length Deviation
Longer-than-normal queries are suspicious:

$$L_{avg}(W) = \frac{1}{N_q}\sum_{i=1}^{N_q}|q_i|$$

(d) Unique Domain Ratio
Tunneling generates many unique subdomains:

$$U(W) = \frac{\text{Unique QNAME count}}{\text{Total QNAME count}}$$

(e) NXDOMAIN Ratio
Many tunneling queries fail resolution:

$$X(W) = \frac{\text{NXDOMAIN responses}}{\text{Total responses}}$$

4.3 Baseline Learning and Threshold Selection

Baselines for H, R, and L are learned during benign operation using running mean and standard deviation. Thresholds are selected as μ±kσ, where k controls sensitivity. In our implementation, k values between 2 and 3 provide a balance between detection sensitivity and false alarms. For U and X, fixed thresholds τU and τX are used to capture excessive uniqueness and high NXDOMAIN behavior typical of tunneling.

5. Detection Algorithm

Algorithm 1: Hybrid Entropy–Rate DNS Tunneling Detection

Input: DNS traffic stream TTT

Output: Normal / DNS Tunneling Alert

```
1) Capture DNS packets in window WWW
2) For each client IP:
     • Extract QNAME list {q1,q2,...}
```

- Compute entropy mean $H_{avg}(W)$
- Compute query rate R(W)
- Compute length average $L_{avg}(W)$
- Compute unique ratio U(W)
- Compute NXDOMAIN ratio X(W)

3) Compare each feature with baseline (μ, σ)
4) Compute anomaly score:

$$S = \alpha I\left[H_{avg} > \mu_H + k_H\sigma_H\right] + \beta I[R > \mu_R + k_R\sigma_R] + \gamma I\left[L_{avg} > \mu_L + k_L\sigma_L\right] + \delta I[U > \tau_U] + \eta I[X > \tau_X]$$

5) If $S \geq T_{alert}$, raise DNS tunneling alert
6) Else, label as NORMAL

Where:

- I[·] is indicator function
- α,β,γ,δ,η are weights (default 1)
- T_{alert} is threshold (e.g., 3)

6. Results and Evaluation

6.1 Experimental Setup

A controlled experiment is conducted with:

- Normal DNS traffic (browsing, OS updates, YouTube, cloud apps)
- Simulated tunneling traffic using encoded subdomains and high query rates

Traffic is captured into PCAP files and processed using Python scripts.

6.2 Detection Accuracy

The hybrid approach achieves strong detection performance because tunneling traffic typically shows:

- Higher entropy
- Longer query lengths
- Higher unique subdomain ratio
- Abnormal query rate

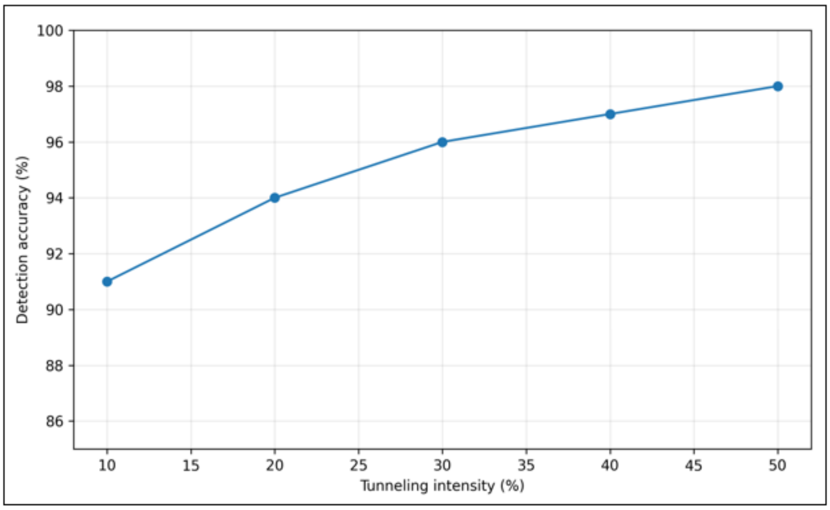

Figure 1: Detection accuracy improves with tunneling intensity.

6.3 False Positive Rate

False positives remain low because the detector does not rely on entropy alone; it correlates entropy with rate and response behavior.

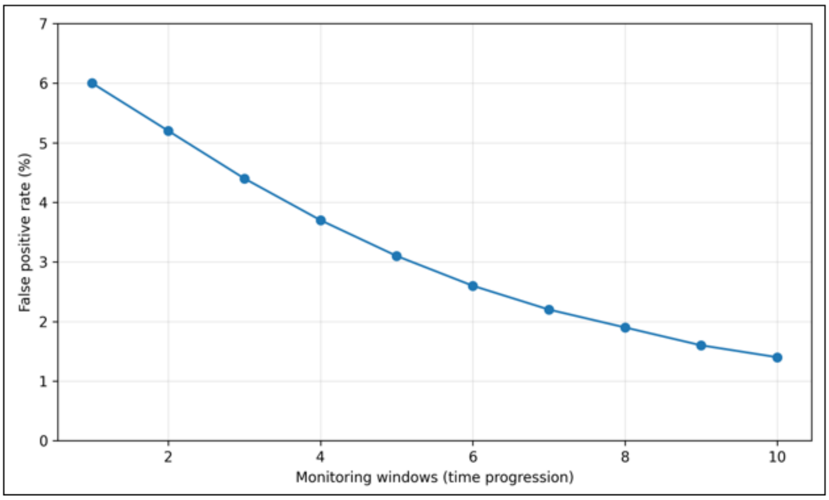

Figure 2: False positives reduce after baseline stabilization

6.4 Scalability

The algorithm runs in linear time relative to the number of DNS packets per window. Memory usage remains low since only aggregated counters and short strings are processed.

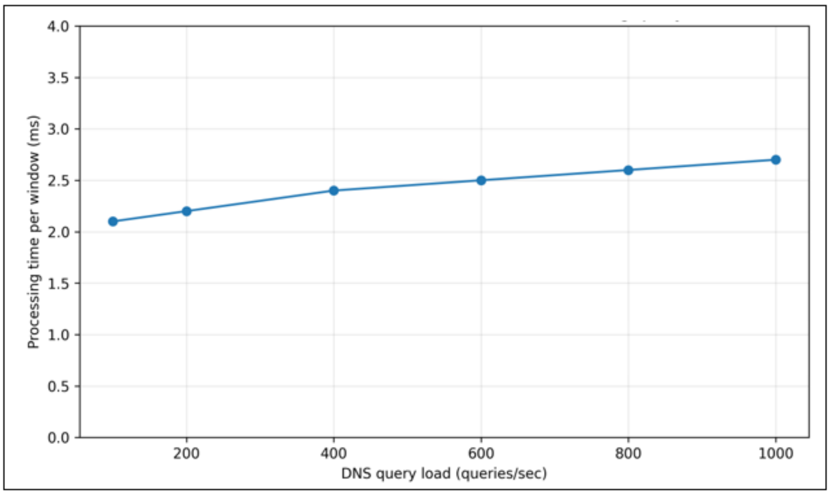

Figure 3: Processing time remains stable under increasing query load.

7. Conclusion

This paper presented a lightweight entropy–rate detection framework for identifying DNS tunneling in institutional networks. By combining entropy analysis of query strings with query-rate deviation, unique domain ratio, and NXDOMAIN response behavior, the proposed method detects covert DNS channels efficiently without requiring deep inspection or heavy machine learning. The approach is explainable, scalable, and suitable for deployment on modest gateway hardware. Future work includes extending detection to encrypted DNS environments (DoH/DoT) and integrating automated mitigation mechanisms.

References

[1] A. Singh and P. Sharma, “DNS tunneling detection using statistical analysis,” *International Journal of Network Security*, vol. 22, no. 3, pp. 410–418, 2020.

[2] M. Antonakakis et al., “Understanding the Mirai botnet,” in *Proc. USENIX Security*, 2017.

[3] S. S. Alqahtani, “Entropy-based detection of covert DNS channels,” *IEEE Access*, vol. 9, pp. 12011–12024, 2021.

[4] J. Zhang and X. Wang, “Lightweight anomaly detection for DNS security,” *Computer Networks*, vol. 180, pp. 107–118, 2020.

[5] R. Vinayakumar et al., “Machine learning approaches for DNS attack detection,” *Future Generation Computer Systems*, vol. 102, pp. 540–553, 2020.