

ABSTRAK

Adopsi *Transport Layer Security* (TLS) yang semakin meluas telah menghambat inspeksi trafik jaringan secara langsung akibat enkripsi pada lapisan *transport*. Metode JA3 *fingerprinting* menawarkan solusi dengan menganalisis metadata TLS *ClientHello* melalui nilai *hash* MD5 guna mengidentifikasi profil klien dan ancaman. Namun, JA3 memiliki keterbatasan kritis, meliputi kerentanan terhadap *Cipher Stunting*, *false positive* akibat kolisi *hash*, serta ketidakmampuan mendeteksi variasi ancaman yang dinamis. Penelitian ini mengevaluasi pendekatan berbasis kemiripan string yang diterapkan langsung pada *String* JA3 *pre-hash* sebagai alternatif yang lebih adaptif dalam mendeteksi trafik berbahaya pada komunikasi TLS terenkripsi. Tiga algoritma kemiripan *string*, yaitu *Levenshtein Distance*, SSDEEP, dan TLSH, dievaluasi tanpa proses hashing MD5 menggunakan 9.874 paket TLS *ClientHello* dari dataset HIKARI-2021 (3.997 paket *Bruteforce*, 4.263 paket *Probing*), dataset *Malware Traffic Analysis* atau MTA (12 paket *Trickbot* dan *Emotet*), serta 1.602 paket benign hasil *live capture*.

Pengujian dilakukan dalam tiga skenario, yaitu penentuan *threshold* optimal, uji spesifisitas, dan uji ketahanan lintas keluarga trafik. Kinerja diukur menggunakan *Accuracy*, *Precision*, *Recall*, *F1-Score*, *False Positive Rate* (FPR), dan *Failure Rate*. *Levenshtein Distance* mencapai *F1-Score* sebesar 100,00% pada $threshold \leq 46$ dengan nilai *false positive* dan *false negative* sama dengan nol. SSDEEP mencapai *F1-Score* sebesar 78,60% namun mengalami *design mismatch* progresif akibat rendahnya entropi karakter pada *String* JA3. TLSH mencapai *F1-Score* sebesar 83,44% dengan FPR sebesar 99,00%, mengindikasikan kalibrasi *threshold* yang buruk. Seluruh algoritma menghasilkan *F1-Score* sebesar 0,00% pada skenario lintas keluarga, yang membuktikan spesifisitas yang benar. *Levenshtein Distance* direkomendasikan sebagai algoritma yang paling sesuai untuk sistem deteksi intrusi berbasis kemiripan *String* JA3 *pre-hash*.

Kata Kunci: Analisis trafik terenkripsi, Sistem deteksi intrusi, Keamanan jaringan, Kemiripan string, TLS *fingerprinting*.

ABSTRACT

The widespread adoption of Transport Layer Security (TLS) has hindered direct network traffic inspection due to encryption at the transport layer. The JA3 fingerprinting method offers a solution by analyzing TLS ClientHello metadata through MD5 hash values to identify client profiles and threats. However, JA3 has critical limitations, including vulnerability to Cipher Stunting, false positives due to hash collisions, and the inability to detect dynamic threat variations. This study evaluates a string similarity-based approach applied directly to the pre-hash JA3 string as a more adaptive alternative for detecting malicious traffic in encrypted TLS communications. Three string similarity algorithms, namely Levenshtein Distance, SSDEEP, and TLSH, were evaluated without the MD5 hashing process using 9,874 TLS ClientHello packets from the HIKARI-2021 dataset (3,997 Bruteforce packets, 4,263 Probing packets), the Malware Traffic Analysis (MTA) dataset (12 Trickbot and Emotet packets), and 1,602 benign packets from live capture.

Testing was conducted in three scenarios: optimal threshold determination, specificity testing, and cross-family resilience testing. Performance was measured using Accuracy, Precision, Recall, F1-Score, False Positive Rate (FPR), and Failure Rate. Levenshtein Distance achieved an F1-Score of 100.00% at a threshold of ≤ 46 , with zero false positives and false negatives. SSDEEP achieved an F1-Score of 78.60% but experienced a progressive design mismatch due to the low character entropy in the JA3 string. TLSH achieved an F1-Score of 83.44% with an FPR of 99.00%, indicating poor threshold calibration. All algorithms produced an F1-Score of 0.00% in the cross-family scenario, proving correct specificity. Levenshtein Distance is recommended as the most suitable algorithm for intrusion detection systems based on pre-hash JA3 string similarity.

Keywords: *Encrypted traffic analysis, Intrusion detection system, Network security, String similarity, TLS fingerprinting.*