2024 International Conference on Smart Applications, Communications and Networking (SmartNets) | 979-8-3503-8532-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/SMARTNETS61466.2024.10577699

Adaptive Segmentation: A Tradeoff Between Packet-Size Obfuscation and Performance

Mnassar Alyami,[1,2] Cliff Zou,[1] and Yan Solihin[1]

[1]*College of Engineering and Computer Science, University of Central Florida, USA*
{mnassar.alyami | changchun.zou | yan.solihin}@ucf.edu
[2]*College of Computer Science and Information Technology, Jazan University, Saudi Arabia*
malsaad@jazanu.edu.sa

***Abstract*—Connected Internet-of-Things (IoT) devices pose several privacy risks through the analysis of their encrypted network traffic. According to prior studies, packet size can be utilized to train a machine learning classifier for the identification of IoT devices because of their unique functionalities and traffic patterns. A recent defense technique aimed at addressing these privacy concerns efficiently is random segmentation [1]. This mechanism involves breaking down application messages into randomly sized chunks to obscure patterns in packet sizes. However, it leads to higher latency due to the increased number of packets and the additional packet header overhead. Furthermore, nonadaptive (or static) splitting in the original random segmentation approach is inappropriate for networks with dynamically changing conditions, which is common in smart homes. In this paper, we present an adaptive segmentation approach based on optimization, which adapts the splitting volume to changes in network usage. We formulate an optimization problem in order to maximize network traffic obfuscation while minimizing segmentation overhead. We evaluated our adaptive approach through simulations using real-world IoT data traces. Our results illustrate how the adaptive defense system adjusts its splitting parameters to enhance privacy protection, as measured by entropy, while minimizing the impact on transmission performance.**

***Index Terms*—Device Fingerprinting; Adaptive control; IoT Privacy; Traffic Analysis Countermeasure**

I. Introduction

The widespread use of IoT devices poses a privacy threat. Despite encryption, passive observers can still exploit packet size visibility for device fingerprinting (DF) attacks [1]. These attacks allow observers to identify devices and determine their operational states (e.g., whether a TV is in Active or Idle mode [2]), enabling adversaries to deduce sensitive information about user behaviors and activities. For instance, Wang et al. [3] demonstrate that an observer can infer a user's command to a smart speaker using packet length sequences and direction.

Various research efforts have focused on enhancing privacy with minimal data overhead [4], [5]. Typically, these methods aim to minimize the injected noisy data for traffic obfuscation. However, static obfuscation often faces challenges in finding the right balance between privacy protection and overhead [6], resulting in either allowing high attack accuracy or adding too much data overhead. This challenge led to the proposal of an alternative approach that distorts length-based patterns without adding noise [1]. It involves randomizing packet lengths by breaking the data stream from the application layer into random-size segments at the transport layer, instead of introducing dummy packets to hinder traffic classification. As a result, such randomization achieves anonymity with significantly reduced data overhead.

Nevertheless, this packet-splitting methodology still introduces latency due to the increase in packet count and data overhead caused by the introduction of additional packet headers. More importantly, in dynamic networks, fluctuations in network utilization can lead to prolonged queue delays, negatively impacting network performance. Therefore, a robust defense system should incorporate an "adaptive defense" feature—adaptively adjusting its configurations according to the network condition to balance privacy protection and network performance. Particularly, we employ optimization to maximize privacy protection (increasing the randomness in packet size) while considering the impact of message splitting on data transmission rate (reducing latency and overhead).

Indeed, the fundamental concept of "adaptive defense" has found application in various domains, including real-world epidemic disease control, the five-tier terrorism alert system, and others [7]. The key challenge lies in translating this foundational principle into the design of an effective defense system within the realm of traffic obfuscation. In this study, we introduce a specific adaptive defense system that estimates defense strength and its impact on performance. The adaptive parameter adjustment involves straightforward calculations and optimization, ensuring minimal computational overhead.

This work addresses the challenge of optimizing packet splitting parameters in our previous work [1] to achieve a desirable and intelligent trade-off between overhead and protection, thereby enhancing the overall performance of the communication network. Our goal is to determine the optimal splitting percentage that maximizes randomness while taking into account the impact of the defense on latency intelligently and adaptively based on the underlying networking environment. The main contributions of this paper are as follows:

- Proposal of adaptive segmentation, a practical solution based on optimization to defend against DF with minimum latency.
- Performing simulations and verifying the effectiveness of the algorithm using real IoT device traffic.

979-8-3503-8532-8/24/$31.00 ©2024 IEEE

The remainder of this paper is organized as follows: Section II presents the related works. In Section III, we present the adversary model. Section IV covers the methodology of our proposed system. In Section V, we present our experimental results. Finally, we draw our conclusion and discuss future work in Section VI.

II. Related Work

Several studies have demonstrated that packet-length information can be used to identify IoT devices [1] and certain events [3]. The Onion Router (Tor) addresses this issue by transmitting data in a fixed packet length, thereby preventing potential side-channel leakage [8]. However, it's important to note that Tor's multi-hop architecture, while enhancing privacy, also leads to increased received traffic and introduces additional latency.

Padding packets is an effective defense, yet incurs a high data overhead. Several packet padding strategies were described in [9] to defeat attackers' classification but they often result in a substantial increase in data transmission (>500%). A more streamlined approach, suggested by Pinheiro et al. [10], involves incorporating random padding. While this method may decrease the precision of IoT device identification, it still introduces undesirable noisy traffic.

Using the data-link layer defense presented in [11], it is possible to hinder WiFi eavesdropping-based analytics. It shapes the traffic of two devices to resemble each other by simulating the behavior of one device with fake traffic. However, it is important to note that the injected noise becomes detectable for network-layer observers who leverage the encrypted flag designed exclusively for WiFi observers. As a result, this defense mechanism may not be suitable for a broader context.

Protection against traffic analysis attacks in wireless networks can be accomplished through signal-jamming strategies [12], eliminating the need for introducing dummy packets or intentional delays. Generally, this method employs antennas to disrupt traffic in potential adversary locations, thereby raising the noise level. However, it is essential to recognize that this tactic generates interference that negatively impacts the performance of nearby networks and is also deemed illegal[1].

Randomizing packet sizes has been suggested in the past to mitigate packet size leakage in secure shell (SSH) communications [13], [14]. However, such modifications are tailored specifically for the SSH protocol. IoT devices often utilize lightweight communication protocols to meet their connectivity needs [15], such as Hypertext Transfer Protocol (HTTP) and Hypertext Transfer Protocol Secure (HTTPS). Hence, the concept of random segmentation was introduced to shape traffic at the transport layer [1], encompassing data from various application-layer communication protocols. This approach streamlines deployment and is well-suited for integration with IoT architectures.

The concept of adaptive defense was previously introduced within the realm of traffic shaping defense. Pinero et al. [4] introduced an adaptive padding system that modifies the number of injected bytes based on changes in network usage. However, this system is tailored for padding packets and does not align with our random splitting method that does not introduce noise for traffic obfuscation. In contrast, our research offers an adaptive defense formulation for random segmentation. This formulation adjusts the number of splitting operations, considering factors such as randomization intensity and latency.

[1]https://www.fcc.gov/general/jammer-enforcement

III. Threat Model

We consider DF attacks that leverage packet lengths and directions in the context of encrypted WiFi traffic. Two observation points are considered for potential exploitation by an adversary collecting the traffic:

- Active Observer: The adversary can establish a deceptive access point (AP) mimicking the victim's network name, potentially redirecting IoT devices to connect to the fake AP rather than the authentic one. Once connected, the observer can analyze the network-layer traffic. In this context, the attacker can observe the packet header but lacks the capability to identify the specific device or decrypt the content.
- Passive Eavesdropper: The attacker can utilize a WiFi card in monitor mode to capture encrypted WiFi traffic. Unlike active observation, passive eavesdropping is challenging to detect since eavesdroppers do not need to access or join the network. The adversary aims to identify the device type (e.g., smart lock, baby monitor) and monitor network events based on changes in traffic patterns (e.g., a spike in smart lock traffic indicating the user is entering the apartment, or an increase in baby monitor traffic indicating the baby is awake).

IV. Adaptive Segmentation

A. Background

Random segmentation involves intercepting application messages and splitting them into random chunks at the transport layer. For reasons of space, we refer readers to more detailed descriptions in [1]. The quantity of messages subject to segmentation is determined by a predetermined splitting percentage P. This variable serves as a control mechanism, specifying the proportion of messages to be segmented—such as 10% or 50%.

Notably, as P escalates, the degree of packet size randomization, quantified as R, increases. We assess the randomness in packet size, denoted as R, using the same metric proposed in a comparable network obfuscation system [16]. Specifically, we employ Shannon entropy to measure the effectiveness of masking through randomization. Shannon entropy, an information-theoretic metric, evaluates the uncertainty inherent in a random variable. Given a random variable X with values in the finite set $\{x_1, x_2, \ldots, x_M\}$ and l_i denoting the likelihood of $X = x_i$, Shannon entropy is expressed as follows:

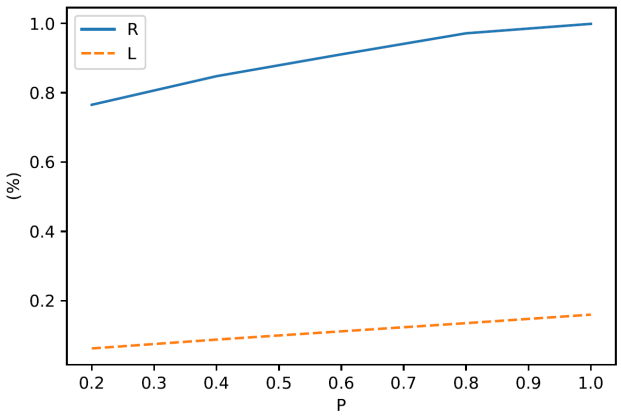

Fig. 1. Packet size randomness R versus latency L under different defense splitting percentage P.

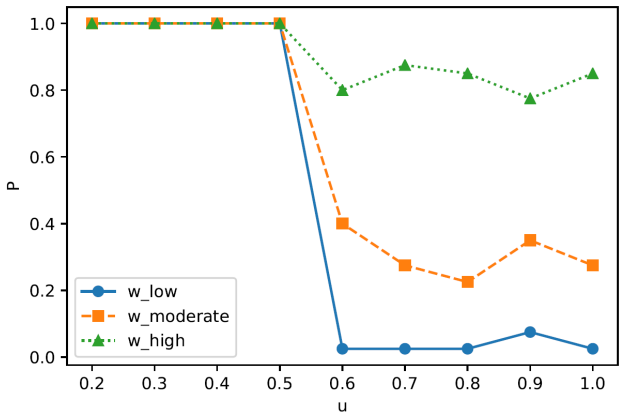

Fig. 2. Performance of our adaptive defense system using different weights w.

$$R(X) = -\sum_{i=1}^{M} l_i \log(l_i) \tag{1}$$

where:

- $R(X)$ is the entropy of the random variable X,
- M is the number of possible outcomes in the discrete random variable X,
- x_i represents each possible outcome, and
- $l(x_i)$ is the likelihood/probability of the occurrence of x_i.

However, the increase in randomness is likely to lead to higher latency L, stemming from the construction of additional packets and data overhead introduced by extra packet headers. To validate this relationship, we utilized a program to emulate an IoT application sending an array of data ($\approx$ 976 KB) to a cloud server, where each message in the array has a length of 1000 bytes. The program sent the same array using various P values, ranging from 0 to 1, and calculated the elapsed time for each P. We conducted five sets of experiments, calculating the average latency overhead normalized by the maximum value (i.e., when $P = 1$).

Figure 1 illustrates the interrelationship between R and L. On the x-axis, 0.2 indicates a defense split of 20% of the total packets, while 1 signifies a defense split involving all messages (100%). R was calculated by simulating our defense on data traces collected from four IoT devices in our previous study [1]. We normalized R on the y-axis by each device's maximum, aggregating the performance of all devices into a single curve for comparable results with L. The latter is calculated based on our client-server experiment described above.

In Figure 1, a clear linear upward trend is observed for both R and L with the increase in P; therefore, it is reasonable to assume they linearly increase with P for the sake of simulation. This assumption is made because the linear or exponential increase will have the same effect on the reward and penalty scheme of our optimization, given that R is the only dimension we aim to maximize, and L is the only dimension we aim to minimize.

B. Obfuscation Optimization Problem

At a conceptual level, our goal is to randomize the packet sizes of a device, thereby thwarting potential classifiers from identifying patterns that facilitate DF. The model determines the optimal P for maximizing R while considering its impact on L. Hence, the objective function defines a score S, combining the defense benefit measured by traffic randomness R, with a penalty term associated with the increase in latency L. Thus, our goal is to maximize the objective function, defined as:

$$\text{Maximize} \quad S = w \cdot R - (1 - w) \cdot L \tag{2}$$

where w is a weight constant that predetermines the tradeoff between R and L before running the adaptive optimization. It could be predetermined by a field expert to express how much we value privacy protection compared with the network performance degradation.

Based on the analysis reported in [17], the latency shows a slight increase with the rise in network usage u. However, when the usage reaches 60%, the latency starts to increase rapidly. Hence, we set our defense to ensure maximum protection when u is below 60%; otherwise, our optimization model is triggered to find the optimal parameter P_{optimal} for the desired tradeoff.

$$P = \begin{cases} 1 & \text{if } u < 0.6 \\ P_{\text{optimal}} & \text{otherwise} \end{cases} \tag{3}$$

V. Experiments and Results

We evaluate the performance of the adaptive obfuscation system using data traces collected from four IoT devices in our previous study [1]. We partition the packet length traces within each device's traffic into sequences within 30-second time windows. Next, we deploy our defense across a range of u from 0 to 100%, calculating the corresponding P for each setting.

As shown in Figure 2, our optimization is triggered when u exceeds the threshold (i.e., $u \geq 0.6$), returning the optimal solution for the maximum score S. We demonstrate the effectiveness of our approach using different weights, w_{high}, w_{moderate}, and w_{low}. To ensure high protection, a user can adjust w similarly to w_{high} to place more emphasis on privacy than performance. Alternatively, w_{moderate} provides lower obfuscation for reduced overhead, or one can opt for w_{low} for weaker protection but improved network performance.

VI. Conclusion and Future Work

Our adaptive segmentation approach provides a systematic and effective means of optimizing privacy with controllable costs. By considering both entropy and latency, the method offers a balanced solution that enhances the overall efficiency of communication networks. Unlike previous defenses that prioritize traffic obfuscation at any cost, our approach is designed to give users the flexibility to balance protection and performance according to their preferences.

For future work, our aim is to integrate network traffic obfuscation with machine learning techniques to enhance defense effectiveness. We envision creating a machine learning model to optimize obfuscation quality while mitigating performance degradation. Our plan includes implementing reinforcement learning, where the agent is rewarded for effectively confusing traffic analysis tools and penalized in the event of a decline in system performance.

Acknowledgment

This research was sponsored by the U.S. National Science Foundation (NSF) under Grant DGE- 2325452. We also acknowledge that this work is based on a chapter of the author's dissertation [18].

References

[1] Mnassar Alyami, Abdulmajeed Alghamdi, Mohammed A Alkhowaiter, Cliff Zou, and Yan Solihin. Random segmentation: New traffic obfuscation against packet-size-based side-channel attacks. *Electronics*, 12(18):3816, 2023.

[2] Mnassar Alyami, Ibrahim Alharbi, Cliff Zou, Yan Solihin, and Karl Ackerman. Wifi-based iot devices profiling attack based on eavesdropping of encrypted wifi traffic. In *2022 IEEE 19th Annual Consumer Communications Networking Conference (CCNC)*, pages 385–392, 2022.

[3] Chenggang Wang, Sean Kennedy, Haipeng Li, King Hudson, Gowtham Atluri, Xuetao Wei, Wenhai Sun, and Boyang Wang. Fingerprinting encrypted voice traffic on smart speakers with deep learning. In *13th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 254–265, 2020.

[4] Antônio J. Pinheiro, Paulo Freitas de Araujo-Filho, Jeandro de M. Bezerra, and Divanilson R. Campelo. Adaptive packet padding approach for smart home networks: A tradeoff between privacy and performance. *IEEE Internet of Things Journal*, 8(5):3930–3938, 2021.

[5] Noah Apthorpe, Danny Yuxing Huang, Dillon Reisman, Arvind Narayanan, and Nick Feamster. Keeping the smart home private with smart (er) iot traffic shaping. *arXiv preprint arXiv:1812.00955*, 2018.

[6] Gaofeng He, Xiancai Xiao, Renhong Chen, Haiting Zhu, Zhaowei Zhang, and Bingfeng Xu. Secure and efficient traffic obfuscation for smart home. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 6073–6078. IEEE, 2022.

[7] Cliff C Zou, Nick Duffield, Don Towsley, and Weibo Gong. Adaptive defense against various network attacks. *IEEE Journal on Selected Areas in Communications*, 24(10):1877–1888, 2006.

[8] Tao Wang and Ian Goldberg. Walkie-talkie: An efficient defense against passive website fingerprinting attacks. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 1375–1390, 2017.

[9] Sean Kennedy, Haipeng Li, Chenggang Wang, Hao Liu, Boyang Wang, and Wenhai Sun. I can hear your alexa: Voice command fingerprinting on smart home speakers. In *2019 IEEE Conference on Communications and Network Security (CNS)*, pages 232–240, 2019.

[10] Antônio J Pinheiro, Jeandro M Bezerra, and Divanilson R Campelo. Packet padding for improving privacy in consumer iot. In *2018 IEEE Symposium on Computers and Communications (ISCC)*, pages 00925–00929. IEEE, 2018.

[11] Mnassar Alyami, Mohammed Alkhowaiter, Mansour Al Ghanim, Cliff Zou, and Yan Solihin. Mac-layer traffic shaping defense against wifi device fingerprinting attacks. In *2022 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–7, 2022.

[12] Fan Zhang, Wenbo He, and Xue Liu. Defending against traffic analysis in wireless networks through traffic reshaping. In *2011 31st International Conference on Distributed Computing Systems*, pages 593–602. IEEE, 2011.

[13] Martin R Albrecht, Kenneth G Paterson, and Gaven J Watson. Plaintext recovery attacks against ssh. In *2009 30th IEEE Symposium on Security and Privacy*, pages 16–26. IEEE, 2009.

[14] Mihir Bellare, Tadayoshi Kohno, and Chanathip Namprempre. Authenticated encryption in ssh: provably fixing the ssh binary packet protocol. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 1–11, 2002.

[15] Jasenka Dizdarević, Francisco Carpio, Admela Jukan, and Xavi Masip-Bruin. A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. *ACM Computing Surveys (CSUR)*, 51(6):1–29, 2019.

[16] Louma Chaddad, Ali Chehab, and Ayman Kayssi. Opriv: Optimizing privacy protection for network traffic. *Journal of Sensor and Actuator Networks*, 10(3):38, 2021.

[17] Robert Underwood, Jason Anderson, and Amy Apon. Measuring network latency variation impacts to high performance computing application performance. In *Proceedings of the 2018 ACM/SPEC International Conference on Performance Engineering*, pages 68–79, 2018.

[18] MNASSAR ALYAMI. *INTERNET-OF-THINGS PRIVACY IN WIFI NETWORKS: SIDE-CHANNEL LEAKAGE AND MITIGATIONS*. PhD thesis, University of Central Florida, Orlando, Florida, 2024.