Computer Networks
Volume 213, 4 August 2022, 109084

GRAIN: Granular multi-label encrypted traffic classification using classifier chain

Faiz Zaki [a], Firdaus Afifi [b], Shukor Abd Razak [c], Abdullah Gani [d], Nor Badrul Anuar [a]

https://doi.org/10.1016/j.comnet.2022.109084

Abstract

Granular traffic classification categorizes traffic into detailed classes like application names and services. Application names represent parent applications, such as Facebook, while application services are the individual actions within the parent application, such as Facebook-comment. These granular classes are still insufficient to keep pace with modern applications that offer various services. Accordingly, this paper further divides the application service class into inter-application and intra-application services to provide more insights. Inter-application service refers to a similar service between different parent applications, such as Facebook-comment and Youtube-comment, whereas intra-application service differentiates services within the same parent application, such as Facebook-comment and Facebook-post. Most studies focus on classification at the application name and inter-application service levels. In contrast, classification at the intra-application service level receives far less attention due to its complexity despite providing the highest flexibility. Therefore, this paper presents GRAIN, a granular multi-label approach to classify encrypted traffic at all three levels of granular classification: application name, inter-application and intra-application service levels using a classifier chain. GRAIN chains two random forest classifiers to produce a multi-label classification using seven novel

statistical features based on packet payload length. The utilized features are independent of the packet payload content, thus unaffected by packet encryption and preserving user privacy. Our performance evaluation showed that GRAIN achieved an average F-measure of 99% at the application name level, 93% at the inter-application service level and 88% at the intra-application service level. To test for robustness, we compared GRAIN against four baseline classifiers and the ISCX VPN-nonVPN public dataset in which GRAIN maintained its comparable performance across all tests.

Introduction

Computer networks are growing at an exponential pace in recent years. The latest Cisco's Annual Internet Report highlighted that 66% of the global population would have Internet access by 2023, a 15% surge from 2018. Furthermore, Cisco also forecasted the number of networked devices to rise to 29.3 billion devices from 18.4 billion by 2023 [1]. These significant figures cause network management tasks such as network monitoring and security to become more challenging. In order to deal with this challenge, network administrators depend on visibility into the network. They need to know the type of traffic flowing in their network before taking relevant actions such as applying network policies. However, apart from numerous networked applications available today, countless network services also exist within the applications. For example, applications like Facebook provide multiple services such as chatting, sending files and uploading images which add to the complexity faced by network administrators. To help manage these complexities, network administrators gain the appropriate visibility into the network by taking advantage of network traffic classification.

Network traffic classification plays a central role by classifying the network traffic into various meaningful traffic classes. Typical traffic classes in the early days include coarse-grained classification. Among the classes in coarse-grained classification are application protocols, generic application types, and binary classification, which classify traffic into two classes such as malware and benign traffic [2]. However, coarse-grained classification is quickly becoming ineffective to address the more complex network today. Current circumstances require network administrators to take advantage of a more granular network traffic classification. Granular network traffic classification outputs more fine-grained classification, commonly at the application name and service levels, providing network administrators more visibility and flexibility. For example, network administrators can assign specific policies targeting a particular application name or service instead of applying network policies that affect the entire network protocol. Many existing studies on granular network traffic classification are more focused on classifying traffic at the application name level [3], [4], [5]. Despite that, classifying the application name is insufficient to keep pace with the increasing complexity of modern applications, thus shifting attention to application service classification. We divide application service into two distinct categories: inter-application service (i.e., similar services from different applications such

as Facebook-chat, Twitter-chat, Skype-chat) and intra-application service (i.e., different services from the same application such as Facebook-chat, Facebook-video, Facebook-post). In the literature, the classification is more focused on inter-application services [6, 7] than intra-application, even though the latter provides more visibility and flexibility.

Therefore, this paper introduces GRAIN, a granular multi-label classification approach focusing on classifying encrypted network traffic at all three levels of granular classification: application name, inter-application and intra-application service levels. GRAIN achieves the aim by chaining two random forest classifiers by adopting the classifier chain method. A classifier chain is a well-known method to produce multi-label classification while maintaining the interdependencies between labels. A regular classifier chain implementation chains a series of binary classifiers equal to the number of labels available in the dataset. Each classifier in the series takes the preceding classifiers' output to serve as its input to model the interdependency between the labels. However, in this paper, GRAIN chains only two classifiers (i.e. random forest), reducing the total classifier compared to a regular classifier chain implementation. The first classifier classifies traffic at the application name level using seven statistical features based on payload length. The statistical features avoid any dependencies on packet payload contents, thus unaffected by packet encryption and removing any privacy concerns in obtaining the output. Furthermore, the first classifier's output (i.e., application name) adds as a new feature input for the second classifier to classify traffic at the inter-application and intra-application service levels, thus producing a multi-label classification.

To evaluate the performance of our proposed approach at both granularity levels, we utilized a private ground truth and a public dataset. We collected the ground truth of 43 different application services across ten applications from four different locations: a campus network and three home networks. We collected the data across six months using Grano-GT, a specialized tool to create a reliable granular ground truth [8]. Our evaluations showed that GRAIN achieved average F-measure scores of 99% at the application name level, 93% at the inter-application service level and 88% at the intra-application service level. Based on the scores, GRAIN demonstrated significant performance at all levels considering the complexity of intra-application service classification due to highly similar traffic characteristics. In addition, we benchmarked GRAIN's performance against four baseline classifiers selected from related works in the domain. Results showed that GRAIN outperformed three baseline classifiers, including the traditional non-hierarchical classifier (i.e., flat classifier). On the other hand, this paper also evaluated GRAIN on the ISCX VPN-nonVPN [9] public dataset and recorded *a* + 9% gain in F-measure compared to the flat classifier. As a summary, the main contributions of this paper are as follows:

a) A new categorization of application service: inter-application and intra-application service levels.

b)

A new approach to classify encrypted traffic at the application name, inter-application and intra-application service levels using classifier chain.

c) A novel set of features based on payload length to discriminate between traffic at the application name, inter-application and intra-application service levels.

We organize the remainder of the paper as follows. Section 2 discusses the background of the domain. We present the architecture of GRAIN in Section 3 and provide comprehensive interpretations for each component of the architecture. Section 4 outlines our experimental analysis, and in Section 5, we discuss the current issues in the domain and how to move forward. Finally, Section 6 concludes the paper.

Section snippets

Related work and contribution positioning

Granular network traffic classification is quickly becoming a key technology in network administration and security by leveraging the fine-grained traffic classes' high visibility and flexibility. Fig. 1 shows the taxonomy that generally divides the classification granularity into coarse and fine-grained. Coarse-grained granularity includes the application type (e.g., web, video), application protocol (e.g., HTTP, FTP) and binary classification (e.g., malware vs benign). However, its...

Methodology

In this paper, our proposed classification technique, GRAIN, adopted the classifier chain method to produce a granular multi-label network traffic classification. The classification process started by taking network traffic traces in PCAP format as the ground truth to go through the data processing phase, including input feature extraction. The input features served as the discriminators for the learning process. Finally, the learning process took advantage of two random forest classifiers...

Experimental analysis

We conducted two main experiments to evaluate our proposed approach. Namely, the first experiment evaluated GRAIN's performance when classifying traffic using the self-collected datasets based on Table 3. We combined all the self-collected datasets to introduce spatial and temporal variabilities in the data. In addition, we also compared the classification performance

with the baseline classifiers. The second experiment focused on evaluating the robustness of our proposed approach when tested...

Discussion

This paper presented GRAIN, an approach to classify network traffic with high granularity at the application name, inter-application and intra-application service levels. Granular network traffic classification is a critical technology in modern networks to manage and control the network better. Although network traffic classification has been a topic of interest for a long time, granular network traffic classification efforts have yet to receive the attention it deserves. As such, there is a...

Conclusion

This paper addressed the core issue in modern networks requiring the highest visibility and flexibility to provide better network management. In response to the issue, we presented GRAIN, an approach to classify network traffic with high granularity at the application name, inter-application and intra-application service levels. To achieve this objective, GRAIN utilized statistical features based on the packet payload length to discriminate between applications and the different services within ...

Author statement

Faiz Zaki: Conceptualization, Methodology, Software, Investigation, Writing – Original Draft; Firdaus Afifi: Validation, Visualization; Shukor Abd Razak: Writing- Reviewing and Editing, Supervision; Abdullah Gani: Writing- Reviewing and Editing, Supervision; Nor Badrul Anuar: Conceptualization, Writing- Reviewing and Editing, Funding acquisition(Table 11)...

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper....

Acknowledgements

This work was supported by Konsortium Kecemerlangan Penyelidikan (JPT(BKPI)1000/016/018/25 (49)) provided by the Ministry of Higher Education of Malaysia....

Faiz Zaki is currently a researcher at the Network Analytics Lab, Universiti Malaya. He holds a PhD in Computer Networks (Network Traffic Classification) from Universiti Malaya. Prior to his

PhD, he received his MSc. Web Science and Big Data Analytics from the University College of London. He is also a member of ACM, IEEE Computer Society and IEEE Young Professionals. His-research interests reside in the area of network traffic classification, network security and data science....

References (37)

S.E. Gómez *et al.*
Exploratory study on class imbalance and solutions for network traffic classification
Neurocomputing (2019)

S.E. Gómez *et al.*
Ensemble network traffic classification: algorithm comparison and novel ensemble scheme proposal
Comp. Networks (2017)

G. Aceto *et al.*
Multi-classification approaches for classifying mobile app traffic
J. Netw. Comput. Appl. (2018)

Y.-n. Dong *et al.*
Novel feature selection and classification of Internet video traffic based on a hierarchical scheme
Computer Networks (2017)

F. Zaki *et al.*
Grano-GT: a granular ground truth collection tool for encrypted browser-based Internet traffic
Comp. Networks (2021)

G. Aceto *et al.*
DISTILLER: encrypted traffic classification via multimodal multitask deep learning
J. Netw. Comput. Appl. (2021)

Cisco Annual Internet Report (2018–2023) White Paper
(2020)

W. Wei *et al.*
Malware traffic classification using convolutional neural network for representation learning

Z. Bu *et al.*
Encrypted network traffic classification using deep and parallel network-in-network models

IEEE Access (2020)

M. Lotfollahi *et al.*
Deep packet: a novel approach for encrypted traffic classification using deep learning
Soft Computing (2020)

View more references

Cited by (0)

Recommended articles (6)

Research article

Improving the attribute retrieval on ABAC using opportunistic caches for Fog-Based IoT Networks

Computer Networks, Volume 213, 2022, Article 109000

Show abstract

Research article

Resource-efficient seamless transitions for high-performance multi-hop UAV multicasting

Computer Networks, Volume 213, 2022, Article 109051

Show abstract

Research article

Software-Defined Networking in wireless ad hoc scenarios: Objectives and control architectures

Journal of Network and Computer Applications, Volume 203, 2022, Article 103387

Show abstract

Research article

Defending saturation attacks on SDN controller: A confusable instance analysis-based algorithm

Computer Networks, Volume 213, 2022, Article 109098

Show abstract

Research article

Secure medical data management with privacy-preservation and authentication properties in smart healthcare system

Computer Networks, Volume 212, 2022, Article 109013

Show abstract

Research article

A scalable approach for smart city data platform: Support of real-time processing and data sharing

Computer Networks, Volume 213, 2022, Article 109027

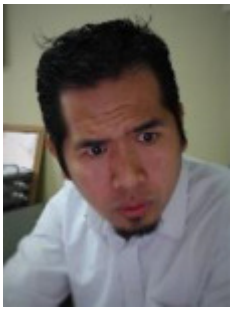
Show abstract

Faiz Zaki is currently a researcher at the Network Analytics Lab, Universiti Malaya. He holds a PhD in Computer Networks (Network Traffic Classification) from Universiti Malaya. Prior to his PhD, he received his MSc. Web Science and Big Data Analytics from the University College of London. He is also a member of ACM, IEEE Computer Society and IEEE Young Professionals. His-research interests reside in the area of network traffic classification, network security and data science.

Firdaus Afifi received bachelor's and master's degree in computer science from the University of Malaya, Malaysia, in 2015 and 2017 respectively. He is currently a PhD student at the Security Research Group, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur. He has published a number of journal papers internationally. His-research interests include information security, data sciences and internet of things.

Shukor Abd Razak is currently a Professor with Universiti Teknologi Malaysia. His-research interests include the security issues for mobile ad hoc networks, mobile IPv6, vehicular ad hoc networks, and network security. He also actively conducts several types of research in digital forensic investigation, wireless sensor networks, and cloud computing. He is the author or coauthor of many journals and conference proceedings at national and international levels.

Abdullah Gani is currently serving as the Dean at the Faculty of Computing and Informatics, Universiti Malaysia Sabah. He is also a Distinguished Professor at the Faculty of Computer Science and Information Technology, Universiti Malaya, Malaysia. He obtained tertiary academic qualifications from the University of Hull, UK – BPhil, and MSc (Information Management) and the University of Sheffield, UK for Phd in Computer Science. His-interest in research kicked off in 1983 when he was chosen to attend the 3-Month Scientific Research Course in RECSAM by the Ministry of Education, Malaysia. Since then, more than 150 academic papers have been published in proceedings and respectable journals internationally within top 10% ranking. Internationally, he serves as a Visiting Professor at the King Saud University, Saudi Arabia, Adjunct Professor at the COMSATS Institute of Information Technology, Islamabad, Pakistan. Currently, he serves as a reviewer to several high-quality journals. He is a senior member of IEEE and was elected as a Fellow of the Academy of Sciences Malaysia (ASM) for Engineering and Computer Science discipline.



Nor Badrul Anuar obtained his Master of Computer Science from University of Malaya in 2003 and a PhD at the Centre for Information Security & Network Research, University of Plymouth, UK. He is an Associate Professor at the Faculty of Computer Science and Information Technology at University of Malaya, Kuala Lumpur. He has published a number of journal papers related to security areas locally and internationally. He has a good profile of publications in renowned Journals. His-research interests include Intrusion Detection System (Intrusion Detection Systems, Intrusion Response Systems, Security Event and Management, Digital Forensic and Network Security), High Speed Network

(Switching, Routing, IPV6, and Multicast) and Management Information System (E-thesis, Library Systems and Online Systems). He is also an associate member of Cisco Systems, Inc. 2008–2016, member of IEEE Communications Society, IEEE Young Professionals and IEEE Computer Society.

View full text