

Research Article

Edge-Based Detection and Classification of Malicious Contents in Tor Darknet Using Machine Learning

Runchuan Li,¹ Shuhong Chen,¹ Jiawei Yang,¹ and Entao Luo²

¹School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China

²School of Electronics and Information Engineering, Hunan University of Science and Engineering, Yongzhou, Hunan 425199, China

Correspondence should be addressed to Shuhong Chen; shuhongchen@gzhu.edu.cn

DWV[hW \$ EVWfW TVd \$" \$ #- 3UWbfW \$` A UfaTVd \$" \$ #- BgT'feZW \$" \$ @ahW TVd \$" \$ #

Academic Editor: Ke Gu

Copyright © 2021 Runchuan Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increase of data in the network, the load of servers and communication links becomes heavier and heavier. Edge computing can alleviate this problem. Due to a sea of malicious contents in Darknet, it is of high research value to combine edge computing with content detection and analysis. Therefore, this paper illustrates an intelligent classification system based on machine learning and Scrapy that can detect and judge exactly categories of services with malicious contents. Because of the nondisclosure and short survival time of Tor Darknet domain names, obtaining uniform resource locators (URLs) and resources of the network is challenging. In this paper, we focus on a network based on the Onion Router (tor) anonymous communication system. We designed a crawler program to obtain the contents of the Tor network and label them into six classes. We also construct a dataset which contains URLs, categories, and keywords. Edge computing is used to judge the category of websites. The accuracy of the classifier based on a machine learning algorithm is as high as 89%. The classifier will be used in an operational system which can help researchers quickly obtain malicious contents and categorize hidden services.

1. Introduction

The Darknet has a huge amount of data. Edge computing can process massive data on terminal devices and transfer the processed results to the server, which alleviates the computing pressure of servers and the load of communication links [1]. Tor (the Onion Router) Darknet [2], which is also known as onion network and dark web, is a network using anonymous communication technology [3]. It is hard to access hidden services and obtain resources from it without using specific software or a proxy agency. Their sites are not only not indexed by Google or other standard search engines but also invalidate quickly [4]. Due to the good concealment of Tor Darknet a lot of illegal contents exit from it, such as drugs, guns, and hacking technology. After the outbreak of COVID-19, many medical products and supplies also appeared in the Darknet market [5] that is not good for the stability of the society.

AlQahtani and El-Alfy [6] conducted extensive research on anonymous technology and the onion network. The strong concealment of Tor network was illustrated, but there were some defects in the design and implementation of Tor hidden service technology [7–9]. Furthermore, due to the special network structure and the characteristics of hiding identities on both sides of communication, the improved onion network technology was also applied to other applications such as the Internet of Vehicles (IoV) ad hoc network [10]. Because of a large number of high-quality resources and the difficulty of obtaining them from the dark web, the mining and analysis of Tor network resources has been a major research hotspot in the academic community.

There are many research methods for Tor network resources. Web crawler technology can improve the efficiency of obtaining network resources. Iliou et al. [11] and Monterrubio et al. [12] proposed a general crawling framework for automatically obtaining web resources in

