Source Identification of Encrypted Video Traffic in The Presence of Heterogeneous Network Traffic

Yan Shi , Arun Ross , Subir Biswas

PII: S0140-3664(18)30002-1
DOI: 10.1016/j.comcom.2018.07.019
Reference: COMCOM 5738

To appear in: *Computer Communications*

Received date: 1 February 2018
Revised date: 31 May 2018
Accepted date: 16 July 2018

Please cite this article as: Yan Shi , Arun Ross , Subir Biswas , Source Identification of Encrypted Video Traffic in The Presence of Heterogeneous Network Traffic, *Computer Communications* (2018), doi: 10.1016/j.comcom.2018.07.019

Source Identification of Encrypted Video Traffic in The Presence of Heterogeneous Network Traffic

Yan Shi[1], *Arun Ross*[2], *and Subir Biswas*[1]
[1] *Electrical and Computer Engineering, Michigan State University, East Lansing, MI*
[2] *Computer Science and Engineering, Michigan State University, East Lansing, MI*

Abstract: This paper uses Traffic Analysis (TA) for identifying sources of tunneled video streaming traffic. The key idea is to examine encrypted and tunneled video streaming traffic at a Soft-Margin Firewall (SMFW) that is located near the streaming client in order to identify undesirable traffic sources and to block or throttle traffic from such sources. The key contribution of the paper is the design and experimental evaluation of a novel 2-stage classifier for identifying specific video sources from heterogeneous background traffic within an encrypted tunnel. Being able to classify video sources in the presence of such traffic mixture can help the SMFW to successfully obfuscate or block undesired video browsing while allowing a user to receive traffic from legitimate applications running over the same encrypted tunnel. Using OpenVPN servers for creating encryption tunnels, experiments were conducted on a large number of popular video streaming sources with various combinations of feature extraction and data processing techniques to verify the effectiveness of the 2-stage classifier. It was experimentally demonstrated that by using the proposed 2-stage classifier, it is indeed possible to identify video streaming sources with high accuracy and low false-positive rates in the presence of non-video background traffic within an encrypted tunnel.

Index Terms: traffic analysis, video streaming, encrypted VPN, classifiers, packet size distribution, packet arrival interval, firewall.

I. INTRODUCTION

With increasing bandwidth availability, streaming video has become one of the major services of the Internet. The top 2 most heavily accessed video streaming sites Netflix and YouTube constitute roughly 50% of the North American Internet traffic according to a 2014 report [1]. As the popularity of video streaming sites grows, their usage starts to spread into private enterprises, where watching certain videos may be undesirable. For example, employers may not want certain video clips containing political, sexual, or violence related messages to be watched by their employees. Often the administrator of an enterprise network may want to block users from watching videos in order to conserve bandwidth and/or to maintain enterprise productivity. At the same time, an administrator may wish to allow users to access some of the video streaming sites due to other business reasons.

Streaming video from specific sources can be detected and subsequently blocked by adding appropriate firewall filters based on the IP packet header information. However, the problem gets complicated when the usage of encrypted tunnels, such as Virtual Private Networks (VPNs) [2] or proxies, are factored in. In this case, the information present in packet headers does not represent their actual destinations. This is because the actual headers are encrypted and replaced by the headers of the tunneling protocol instead. An example of avoiding traffic classification by going through a VPN tunnel is shown in Fig. 1. In such a case, the source, the destination, and the port numbers in a packet are encrypted, and therefore inaccessible to the firewall. As a result, such traffic cannot be blocked by the firewall using packet inspection.

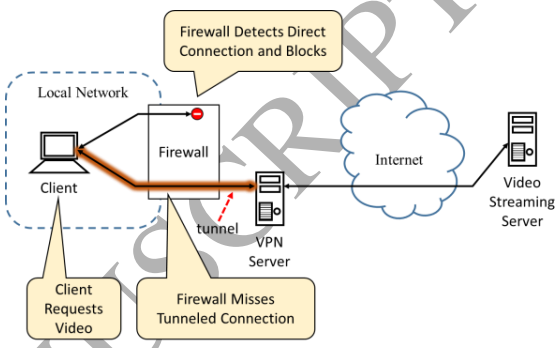

Fig. 1: Firewall circumvention via virtual private networks

This specific problem is addressed in this work by identifying video packet sources using traffic analysis as opposed to relying on packet inspection. To that end, there is a need to not only classify traffic based on the underlying video transport protocols but, also based on its source servers.

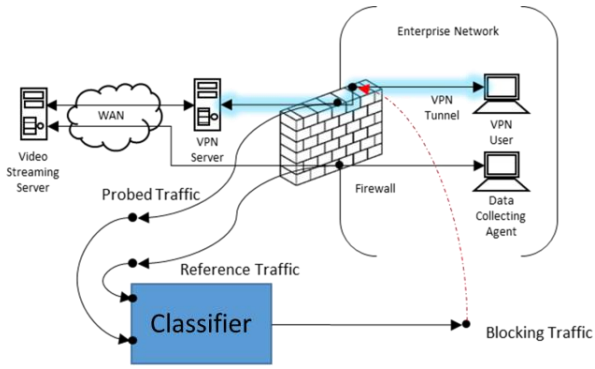

Fig. 2 Proposed System Design

Traffic Analysis (TA) [3] is a commonly used method for retrieving information from traffic flow when the traffic itself is encrypted. The basic assumption is that even if the traffic is encrypted, the underlying protocol it uses might still leave a distinctive signature in the traffic flow. By applying machine learning methods, a classifier can be trained using existing traffic flow, and new encrypted traffic can be classified using that training model. TA is effective for tunneled traffic since its classification is based on the statistical signature in traffic metadata including, packet size, timing, and direction.

The focus of TA research so far is on its usage (and negative impact on privacy) as an attack vector, while in this work the authors propose a beneficial use of the technology as a way to allow network traffic management on encrypted traffic with