Software Architecture for a Lightweight Payload Signature-Based Traffic Classification System

Jun-Sang Park, Sung-Ho Yoon, and Myung-Sup Kim

Dept. of Computer and Information Science, Korea University, Korea
{junsang_park,sungho_yoon,tmskim}@korea.ac.kr

Abstract. Traffic classification is a preliminary and essential step for achieving stable network service provision and efficient network resource management. While a number of classification methods have been introduced in the literature, the payload signature-based classification method shows the highest performance in terms of accuracy, completeness, and practicality. However, the payload signature-based method has a significant drawback in high-speed network environments; the processing speed is much slower than that of other classification methods such as the header-based and statistical methods. In this paper, we describe various design options to improve the processing speed of traffic classification in designing a payload signature-based classification system, and we describe choices we made for designing our traffic classification system. Also, the feasibility of our design choices was proved via experimental evaluation on our campus traffic trace.

Keywords: payload-signature, traffic classification, flow analysis.

1 Introduction

As individual and corporate users are becoming increasingly dependent on the Internet, network speeds are increasing and a variety of services and applications are being developed. Thus, there is a growing need for monitoring and analyzing Internet traffic from the application perspective for achieving efficient network operation and management in various areas, for example, pay-for billing, CRM, SLA, etc. Further, the need for traffic monitoring and analysis will continue to increase. Effective methods are needed for analyzing many types of application-level traffic and handling real-time processing for the large amounts of traffic on high-speed links.

Traffic classification is a preliminary and essential step for achieving stable network service provision and efficient network resource management. While a number of classification methods have been introduced in the literature, the payload signature-based classification method shows the highest performance in terms of accuracy, completeness, and practicality. [1, 3, 4, 8, 9] However, the processing speed of the current classification system is insufficient for real-time handling of the large amounts of traffic on high-speed networks.[6, 7, 11] Given the increasing number of applications and greater usage of applications that generate large amounts of traffic, the inadequate processing speed of payload-based analysis is a challenge that must be mitigated. In this paper, we will define the factors affecting the processing speed of
the signature-based classification system. We aim to improve the processing speed by proposing an optimal classification system structure based on the experimental evaluation of possible design choices for the classification system.

This paper is organized as follows. Section 2 describes research related to this issue and Section 3 describes the design considerations needed for the current payload-based classification systems. Section 4 presents the factors affecting processing speed. An optimal solution based on the experimental results is suggested in order to improve the processing speed. In Section 5, the proposed method is applied to our classification system and its validity is proven. Finally, in Section 6, conclusions and future research directions are described.

2 Related Works

Many applications try to bypass the firewall for a seamless service by frequently changing traffic patterns, so the signature appears in a complex form. In addition, due to the increase of network-based applications and application layer protocols, the number of signatures necessary for identifying applications has been increasing. As the number of signatures and their complexity increase, the processing speed of the payload signature-based classification system has become an important element in determining the performance of the traffic classification system. Many ongoing studies on payload signature-based classification systems aim to accelerate the classification process, but most studies have focused on improving the performance of the pattern-matching algorithm.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Signature Format</th>
<th>No. of Signatures</th>
<th>bps</th>
<th>Matching Algo.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L7-filter</td>
<td>Regular Expression</td>
<td>About 70</td>
<td>Less than 10Mbps</td>
<td>NFA</td>
</tr>
<tr>
<td>Snort</td>
<td>Explicit String + Regular Expression</td>
<td>About 5000</td>
<td>Less than 100Mbps</td>
<td>DFA</td>
</tr>
</tbody>
</table>

Table 1 compares the classification speed of two popular traffic analysis systems, Snort and Linux L7-filter. The table shows the configuration of the signature-based classification system. The L7-filter is widely used for application-level traffic classification. It uses regular expressions to represent signatures and NFA (Non-deterministic Finite Automata) for pattern-matching. However, when over 70 signatures are used to classify applications, it shows a processing speed of 3.5Mbps or less. [6] DFA (Deterministic Finite Automaton) has been proposed to increase the processing speed of NFA. Snort applies the DFA-based pattern matching method, but it has a processing speed of less than 100Mbps [6, 7, 11]. These two systems have tended to focus on pattern matching algorithms to improve performance. However, the time complexity of the matching algorithm is wholly dependent on the configuration of the input data, resulting in limited performance improvement. Thus, real-time traffic analysis of high-speed links (Gbps) might be insufficient if they are