Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation ▮ (▮▮▮▮) ▮▮▮–▮▮▮
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Branch tire packet classification algorithm based on single-linkage clustering

Xia-an Bi*, Xianhao Luo, Qi Sun

College of Mathematics and Computer Science, Hunan Normal University, Changsha 410081, PR China

Received 20 September 2017; received in revised form 1 November 2017; accepted 8 November 2017
Available online xxxxx

Abstract

As the core of network devices is the packet classification technology, the performance influences the development of computer networks. Therefore, researches on the packet classification are of principal theoretical and practical significance. This paper presents a branch trie packet classification algorithm based on the single-linkage clustering (BTSLC). The algorithm includes the preprocessing stage and the matching stage. In the preprocessing stage, packets and rules are firstly mapped to rectangular areas in the two-dimensional space by the formalization method. Then the single-linkage algorithm is used to cluster the formalized rules. In the matching stage, a branch trie is firstly constructed then the matching process is followed. The structure of the branch trie in the algorithm not only wipes out the backtracking by employing the trie path compression but also solves the trie update inefficient problem, which would enhance the performance of the algorithm to a large extent. Finally, the simulation experiments and the actual environmental experiments are carried out to evaluate our algorithm's performances, and some algorithms are used as the comparison groups. The experimental findings indicate that our algorithm has better performance in the searching speed, memory requirement and rule update.

Keywords: Packet classification; Clustering; Single-linkage; Branch trie

1. Introduction

The technology of packet classification has become a vital component of the network equipment [4]. The packet classification acts as a tool to divide data packets into distinguishing types based on rules and to handle these packets with diverse actions such as packet forwarding. The packet classification technology could be applied to not only policy-based routing, SPD (Security Policy Database) of VPN (Virtual Private Network), and access control in network security systems, but also firewall policy, QoS (quality of services) [18] and intrusion detection [15]. Therefore, the performance of the packet classification has a direct impact on the improvement of the network

* Corresponding author.
E-mail address: bixiaan@hnu.edu.cn (X. Bi).

https://doi.org/10.1016/j.matcom.2017.11.003

Please cite this article in press as: X. Bi, et al., Branch tire packet classification algorithm based on single-linkage clustering, Mathematics and Computers in Simulation (2017), https://doi.org/10.1016/j.matcom.2017.11.003.

technology. Researches on the efficient packet classification algorithms have become a hot topic in the computer network. A packet classification system in the real-time operation needs to promptly decide how to handle each incoming data packet, such as whether to forward, where to forward, etc. [21]. A complete rule consists of two main parts: fields and actions. Generally, the field includes the source IP address prefix, the destination IP address prefix, the source port, the destination port and the protocol [2]. The process of the packet classification is to searching for the rule set to find out the highest priority matching rule. Thus, the indicators to reflect an algorithm's performance include the commonly used searching speed and the memory storage [12], and also include the scalability and rule update that have attracted growing attentions in the existing studies [8].

Extant literature on the packet classification could be summarized into the hardware-based algorithms [9,10,19] and software-based algorithms [3,7,11,16,17]. In the hardware-based algorithms, the usage of the hardware such as TCAMs helps to improve the searching speed, but also brings the high cost and low flexible scalability [21]. Meanwhile, as the energy consumption of such algorithms is relatively high, the application of these algorithms is restricted to small-scale rule sets. By contrast, the software-based algorithms are characterized by the low cost, easy implementation and deployment. But this type of algorithms does not have good performance and could not effectively support large-scale rule set. Therefore, it is urgent to propose a new solution to improve the extensibility as well as achieve the high packet matching performance [22]. In this paper, we employed the single-linkage cluster theory to build a branch trie to realize the packet matching. Firstly, the formalization method was used to map rules and packets. Next, the single-linkage clustering method was introduced to divide the rules into different clusters according to the aggregate characteristics. Thirdly, a branch trie based on the diversified single-linkage clustering was proposed. Lastly, the experiments were respectively conducted to evaluate the performance of the proposed algorithm and typical algorithms.

By combining the single-linkage clustering method and branch tries, our paper contributes to the literature on packet classification as follows. Theoretically, the method of formalization based on the geometric space is introduced in this paper. We map the packets and the rules into the two-dimensional areas, and the appropriateness of this method's application to the packet matching process has also been proved. Moreover, we proposed a branch trie structure by adopting the path compression method to avoid backtracking. This branch trie structure helps to enhance the update performance to a large extent. Practically, our proposed algorithm was deployed in the network system which is beneficial to further evaluate and improve the algorithm. The findings indicate that our proposed algorithm has the better performance of high packet matching speed, low memory requirement, and easy deployment.

The remaining part of the paper has 6 sections. Section 2 concludes the related works of the packet classification research area. Section 3 details the process of the packet and rules formalization. A branch trie packet matching algorithm based on the single-linkage algorithm is introduced in Section 4. Finally, the experimental evaluation is presented in Section 5 and the conclusions are given in Section 6.

2. Related works

Existing studies on packet classification algorithms mainly fall into two research streams. They are respectively the software-based algorithms and the hardware-based algorithms. The software-based stream uses the complex data structures to construct packet classification algorithm, and is featured with the performance constraint. In comparison, the hardware-based stream aims to accelerate the process of packet classification through designing dedicated hardware architecture, but is featured with high cost, uneasy implementation and deployment.

The algorithms based on hardware could be summarized as follows: TCAM-based algorithms [19], GPU-based algorithms [10], FPGA-based algorithms [9], DRAM/SRAM-based algorithms [14]. In general, the method of parallel search has been introduced to speed up the searching process. The parallel operation helps to accelerate the searching speed but brings huge energy consumption. Meanwhile, the hardware resources such as GPU are always expensive. Therefore, this type of algorithms is unable to effectively deal with rule sets in large scale [2]. For example, most hardware-based algorithms adopted the FPGA platform. As developing such a dedicated hardware chip requires high cost and long time, it is not easy to modify and extend the chips to support new functions when customers' needs change [15]. On the other hand, the algorithms dealing with rule sets in small scale usually utilize SRAM and general DRAM. These algorithms have faster searching speed, but SRAM and general DRAM are expensive and lack scalability [14]. In short, the algorithms based on hardware have high classification performance but low flexibility.

The algorithms based on software could be summarized as many types, such as Hierarchical Trie based algorithms [3], Grid of Trie based algorithms [17], Tuple Pruning based algorithms [7], Area-Based Quad Tree

Please cite this article in press as: X. Bi, et al., Branch tire packet classification algorithm based on single-linkage clustering, Mathematics and Computers in Simulation (2017), https://doi.org/10.1016/j.matcom.2017.11.003.