International Journal of Science, Environment and Technology, Vol. 4, No 4, 2015, 1098 – 1106

ISSN 2278-3687 (O)
2277-663X (P)

MULTI-MATCH PACKET CLASSIFICATION BASED ON DISTRIBUTED HASHTABLE

[1]Neeshma K K and [2]Renisha P Salim
[2]Assistant Professor
Department of Computer Science and Engineering
Caarmel Engineering College, Ranni, Perunad
Mahathma University, Kerala
E-mails: [1]neeshma99@gmail.com , [2]renisha.salim@bccaarmel.ac.in

Abstract: Packet classification has proved to be an important challenge in network routers. The network applications such as Network Intrusion Detection System, Packet-level accounting, Load-balancing are require packet classification. It require a ruleset database that contain set of rules and comparing each incoming packets against this database. The packet classifier perform multimatch packet classification for getting best matched rule. A hash based asynchronous pipeline architecture helps to improve throughput and memory access rate. We can also combine the intermediate result returned from single dimensional search engines. In this paper, proposed a new distributed hash based architecture. This hash tables are helps to improve packet traffic rate and reduce packet loss. This system also ensure that to avoid collision involved in hash table.
Keywords: Hash Table, Multimatch, Packet Classification, Rule set, Distributed Architecture.

I. INTRODUCTION

A computer network is a network which allows computers to exchange data. In computer networks, networked devices exchange data to each other along network links. Packets are used for data transfer. The connections between nodes are established using network media. The computer network is well-known as Internet. Nodes can include hosts such as personal computers, phones, servers and networking hardware. Two devices are networked together when one device is able to exchange information with the other device, that are they have a direct connection to each other or indirect. Computer networks differ in case of media used to carry and transmit their signals, the transmission protocols to organize network traffic, the size of the network', topology and organizational concept. Computer networks provide and support applications for accessing to the World Wide Web, application used for sharing and storage machines, printers, and fax etc, and use of email and instant messaging applications

*Received June 27, 2015 * Published Aug 2, 2015 * www.ijset.net*

Computer networks also design may differ. The two basic network designs are called client/server and peer-to-peer. Client-Server network is a centralized server computers that can store email, Web pages, files and or applications. On a peer-to-peer network, all computers tend to support the same functions and property. Client-server networks are used much more common in business and peer-to-peer networks used much more common in homes. The layout or structure of data flow of network is called network topology. All of the computers share and communicate across one common conduit called bus networks, all data flows and connections through one centralized device called star network,. Common types of network topologies include bus, star, ring network sand mesh networks [10].

In computing, a hash table (hash map) is a data structure used to perform indexing and a data structure that can map keys to values. Hash function to compute an index into an array of database in the hash table, from which the correct value can be found and each key will be assigned to a unique database by the hash key, but this situation is rarely achievable in practice (usually some keys will hash to the same bucket). Instead, most hash table design cause hash collisions different keys that are assigned by the hash function to the same bucket will occur and must be accommodated in some way. The average cost (number of instructions) in a well-dimensioned hash table, for each lookup is independent of the number of elements stored in hash table. Hash table designs allow us arbitrary insertions and deletions of key-value pairs at constant average cost per operation. Hash tables turn out to be more efficient than search trees in many situations or any other table lookup structure hence they are widely used in many kinds of computer software such as associative arrays, database indexing, caches, and sets. A hash table is made up of two parts such as 1) An array (the actual table where the data to be searched is stored) and 2) A mapping function known as a hash function and it is a mapping from the input space to the integer space that defines the indices of the array or the hash function provides a way for assigning numbers to the input data such that the data can then be stored at the array index corresponding to the assigned number [11].

A. *Packet Classification Problem*

In the packet classification application, packets are classified into flows according to policy or routing information. The policy is specified using fields in the header of a packet. Specifications of fields are called rules. So flows are specified by rules applied to incoming packets. Each rule consists of several fields, say d. A collection of rules is called a classifier.

Each field is either an exact value or a prefix or a range. The distributed hash table discussed above module are connected together by a pipeline architecture this pipeline architecture is able to achieve high packet classification speed with efficient memory utilization with the help of these architecture. Several single dimensional searching are combined to perform multidimensional search each search engine return matching characters and finally this system can return best matched rule for forwarding incoming packet [8].

With the above definitions, a rule can be considered as a hyper rectangle in the d-dimensional space. A classifier is a set of such hyper rectangles. Hyper rectangles in the classifier might be overlapped. A packet is then a point in the d-dimensional space. Thus, packet classification is equivalent to finding all hyper rectangles which contain the query point. This resembles the point location problem in computational geometry. The difference between the packet classification and the point location problem is that hyper rectangles in the point location problem are no overlapping, while hyper rectangles in the packet classification problem may overlap. Hence, the packet classification is more complex than the point location problem. However, structures and characteristics in classifiers could be exploited to develop high performance packet classification algorithms. Such packet classification algorithms may sidestep the performance upper bound achieved in the point location background. The algorithm to be developed in this paper serves as one such example.

B. ***Signature Tree***

The signature file method is a most popular indexing technique used in information retrieval and databases. It proficient in lower space overhead and efficient index maintenance. Different approaches for organizing signature files have been designated such as S-trees, sequential signature files, bit-slice files and its different variants such as signature trees. In this paper, extends the structure of signature trees by introducing multiple-bit checking [1][13]. That is, during the searching of a signature tree against a query signature “sq”, more than one bit in “sq” will be checked each time when a node is encountered. This is not only reducing significantly the size but also increases the filtering ability of the signature tree. Experiments showing that the general signature tree uniformly outperforms the signature tree approach.

The emergence of latest network applications techniques, such as packet-level accounting and the network intrusion detection system needs packet classification to report all best matched rules .several schemes have been developed to address the multimatch packet classification problem, majority of them require either big memory or expensive ternary content addressable memory (TCAM) to store the intermediate data structure, or they may suffer from steep performance degradation under certain types of classifiers. Here decompose the operation of multimatch packet classification from the complicated multidimensional search to several single-dimensional search and present an asynchronous pipeline architecture based on a signature tree structure to combine the intermediate results returned from single-dimensional searches.

II. RELATED WORK

Many schemes have been proposed in literature to address the best-match packet classification problem such as TCAM-based schemes, hicut, hyper cut, set splitting algorithm, multi dimensional cutting. TreeCAM, wire speed TCAM. To provide high throughput and efficient memory packet classification by using hash-based pipeline architecture. This existing method using multi dimensional packet classification using several single dimensional searches. It describe a centralized hashtable based scheme [1].

Ternary content-addressable memories (TCAMs) is popular in the industry for storing and searching Access Control Lists (ACLs). Hence, when using a TCAM proposing algorithms for addressing two important problems that are help to reducing range expansion and multi-match classification. The problem of expansion of rules with range fields to represent range rules in TCAMs are the problem of first algorithm. This system can reduces the utilization of TCAMs by mapped a single range rule to multiple TCAM entries .Here propose a new method called Database Independent Range PreEncoding (DIRPE) the main comparison to earlier approaches that, in a single rule maps it reduces the worst-case number of TCAM entries . DIRPE works without prior knowledge of the database, and good incremental update properties, scales when a large number of ranges is present [5].

Internet routers that provide services such as firewalls, a variety classes, perform different flows in different operations. All the packets sharing common header characteristics called flow; for example a flow may be defined as all the packets between two specific IP addresses. When classifying a packet, a router check a table (or classifier) using one or more fields from the packet header to search for the corresponding flow. The classifier means a list of rules that identify each flow and the actions to be performed on each. With the performance of router increases, develop an algorithm that help packet classification very quickly and efficient memory utilization that allow new flows to be frequently added and deleted. Sometime the packet classification is hard requiring routers when using heuristics that exploit structure present in the classifiers [9]. Multi match packet classification are performed some new network applications such as intrusion detection systems and packet-level accounting, it means that all matching fields are need to be reported. Ternary Content Addressable Memories (TCAMs) have been proposed to solve the multi-match classification problem because their ability to perform fast parallel matching. However, the main limitation of TCAM is it is very expensive and consumes large amounts of power. The previously published all multi-match classification schemes are not provide memory and power efficient.

This system introduces an algorithm called HyperCuts that used for classification. It is similar to the existing algorithm such as HiCuts, were decision tree structure based algorithm is called HyperCuts. Then in case of HiCuts, however, a hyperplane is used for representing the decision tree, in case of HyperCuts k–dimensional hypercubes are used to represents each node in the decision tree. A new set of heuristics and the degree of freedom help to find optimal hypercubes for a given amount of storage memory, HyperCuts has an advantages that it can provide an order of magnitude improvement over existing classification algorithms[6]. The first algorithm is a renovated TCAM design that can find all or the first r matches in a packet classifier set. The second algorithm is a novel partitioning scheme based on filter intersection properties allowing us to use off-the-shelf TCAMs for multimatch packet classification. In compare to existing method the classifier engine finds all matches in exactly one conventional TCAM cycle therefore this system can reduce the power consumption, so this method is far better than the existing hardware-based designs

The process of categorizing packets into "flows" in an Internet router called packet classification. All packets bound up to the identical flow obey a pre-defined rule and are processed in an identical manner by the router. For example, all packets with the same root and destination IP addresses may be defined to form a flow. Packet classification is requisite for non "best-effort" services like firewalls and quality of service that require the capability to distinguish and isolate traffic in divergent flows for appropriate processing. In general, on multiple fields - packet classification is a challenging problem. Hence, researchers have planned a variety of algorithms which broadly speaking can be categorized as heuristic algorithms or hardware-specific search algorithms, basic search algorithms, geometric algorithms. Algorithms describe that are representative of each category and discuss which type of algorithm might be desirable for different applications.

Packet Classification is a key functionality offered by modern routers. Former approaches TCAM and algorithmic perform well in either lookup efficiency (power and number of accesses) or update effort but not both. To execute well in both, suggest TreeCAM, which pursue three novel ideas [7]. Dual versions of TreeCAMs decision tree to decouple search and updates: A coarse version with a few thousand rules per leaf achieves efficient lookups and a fine version with a few tens of rules per leaf reduces update effort. The Interleaved layout rules in the TCAM: Combined with the fine version's few rules per leaf, the layout enables us to bound our worst-case update effort. Path-by-path updates to enable update work to be interspersed with packet lookups (i.e., non-atomic updates), eliminating packet buffering. Usage of simulations of 100,000-rule classifiers reveal that Tree CAM performs well in both lookups and updates: 6- 8 TCAM subarray approaches per packet, matching with modern TCAMs. Nearer to an idealized TCAM in worst-case update effort while obviating little buffering of the packets.

Here we propose TreeCAM [2], which employs three novel ideas. Packet classification is a function which is increasingly used in a number of networking appliances and applications widely. Typically, it consists of a set of abstract classifications and a set of rules which sort packets into the various classifications. The Ternary Content-Addressable Memories, (TCAMs) have become a norm in most of the present networking hardware system. However, TCAMs are very costly and power-hungry thus, a packet classification rule set need to be minimized before populating the TCAM. Here we formulate the Rule set minimization Problem for TCAM as an abstract and optimization problem based on two-level logic minimization, and propose an exact solution of heuristics. Here the present experimental, results with using Class Bench tool suite and a real firewall Access Control List (ACL) from a big enterprise, shows an average decrease in 41% in case of artificial filter sets and 72.5% decrease in the firewall ACL using these proposed heuristics.

III. METHODOLOGY

In this section, the major implementation techniques of the system are described.

A. System Overview

In the packet classification application, packets are classified into flows according to policy or routing information. The policy is specified using fields in the header of a packet.

Specifications of fields are called rules. So flows are specified by rules applied to incoming packets. These packets are classified without using predetermined classifier. Using this classifier create a signature tree Edges of signature tree divided into a hash table Therefore create a pipelined architecture for perfect hash table. The main system architecture of this system shows figure.

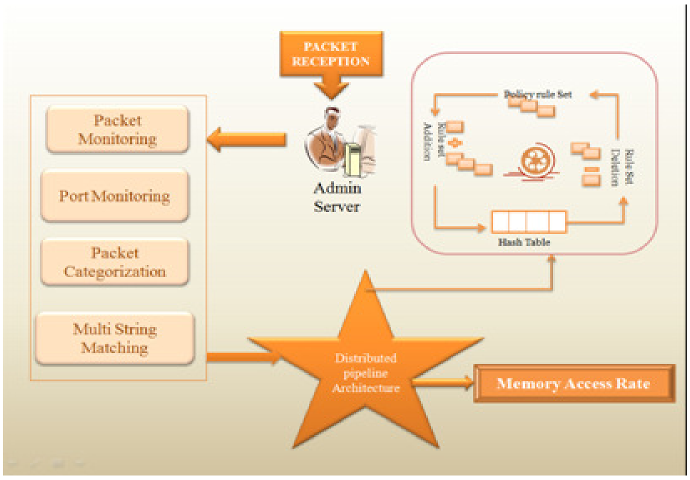

Fig 1: System Overview

This system architecture shows that a administrative system recieve all incoming packets and monitoring packets and their ports, then categorize the packets using classifier. There is an option for adding new ruleset and delete ruleset. All rulesets are arranged into a hashtable. Access corresponding packets and there is a chance for anomaly detection.

B. packet monitoring

This system considered as a administrative tool.It accept all the incoming packet coming from the intranet and internet. The system monitor the packets such as IP packets, TCP packets, UDP packets, DNS packets. Ip packets are packet coming from own system.tcp packets are Packet from external network. UDP packets are packet from home network and DNS packets are packets coming directly from servers. It also monitor all the port that the packet coming from. There is a histogram that show the travelling rate of packets and mean variation entropy that shows any changes occur the system. This module also report all the network activity and UDP/TCP packet information.

C. Ruleset Generation

This module describe the generation of ruler, The administrator create a ruler database that contain set of rules that specify the content of specific packet header fields to result in a match. Each rule is comprised of different fields such as source IP, destination IP, source port, destination port, protocol fields and actions. Also add new rules in rulerset database and can delete existing rule from rulerset database, a packet classifies compare every incoming packet against a set of rulerset based on header field of each packet and identify a flow. this proposed system perform multimatch packet classification that is the packet classification is required to report all matched rule.

D. Hashtable creations

This module describes the creation of hashtable. This system already create a ruleset database and every incoming packet compare and search the rule hard on a signature tree, To make this comparison easier here propose a hash based packet classification. This system propose a distributed hashtable, that help to reduce packet loss and traffic rule in the network. this method consider the classifies as a set of strings with the characters. Searching can be performed using signature tree and divide edges of signature tree into different hashtables. The hash function used here are "parent node id:character: the result will be child node id. this proposed work implemented the hashtable into a distributed networks, therefore it can reduce the packet loss and also help to make traffic easier.

E. Packet classification

The distributed hashtable discussed above module are connected together by a pipeline architecture this pipeline architecture is able to achieve high packet classification speed with efficient memory utilisation. Using these architecture, several single dimensional searching are combined to perform multidimensional search each search engine return matching characters and finally this system can return best matched rule for forwarding incoming packet.

IV. Conclusion

Here model the multimatch packet classification as a combined multistring matching problem, which can be solved by using a pipelined architecture based on a hashtable. Hash tables are generated by traversing a flat signature tree. To speed up the classification and improve traffic rate propose a distributed hash table. These hash tables are then connected together by a pipeline architecture, and they work in parallel when packet classification operations are performed. The proposed pipeline architecture is able to reduce packet loss with a very low storage requirement.

ACKNOWLEDGMENT

The first author would like to thanks all those people, who guided and supported and also likes to thank colleagues for their discussions and suggestions.

REFERENCES

[1] Yang Xu, IEEE, Zhaobo Liu, Zhuoyuan Zhang, and H. Jonathan Chao, "High-Throughput and Memory-Efficient Multimatch Packet Classification Based on Distributed and Pipelined Hash Tables" in IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 22, NO. 3, JUNE 2014.

[2] M. Faezipour and M. Nourani, “Wire-speed TCAM-based architectures for multimatch packet classification,” IEEE Trans. Comput., vol. 58, no. 1, pp. 5–17, Jan. 2009.

[3] R. McGeer and P. Yalagandula, “Minimizing rulesets for TCAM implementation,” Proc. IEEE INFOCOM, pp. 1314–1322, Apr. 2009.

[4] S. Kumar, J. Turner, and J. Williams, “Advanced algorithms for fast and scalable deep packet inspection,” in Proc. ACM/IEEE ANCS, New York, NY, USA, 2006, pp. 81–92.

[5] K. Lakshminarayanan, A. Rangarajan, and S. Venkatachary, “Algorithms for advanced packet classification with ternary CAMS,” in Proc. ACM SIGCOMM , New York, NY, USA, 2005, pp. 193–204.

[6] F. Yu, T.V. Lakshman, M.A. Motoyama and R.H. Katz, “SSA: a power and memory efficient scheme to multi-match packet classification,” in Proc. ACM ANCS, New York, NY, USA, 2005, pp. 105–113.

[7] S. Singh, F. Baboescu, G. Varghese, and J.Wang, “Packet classification using multidimensional cutting,” in Proc. SIGCOMM, New York, NY, USA, 2003.

[8] P. Gupta and N. McKeown, “Algorithms for packet classification,” IEEE Netw., vol. 15, no. 2, pp. 24–32, Mar.–Apr. 2001.

[9] P. Gupta and N. McKeown, “Classifying packets with hierarchical intelligent cuttings,” IEEE Micro, vol. 20, no. 1, pp. 34–41, Jan.–Feb. 2000.

[10] https://en.wikipedia.org/?title=Computer_network

[11] https://en.wikipedia.org/wiki/Hash_table

[12] https://en.wikipedia.org/wiki/Ternary Content Addressable Memory

[13] https://en.wikipedia.org/wiki/Signature_Tree.