IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

A Study on Various Packet Classification Algorithm for Network Security Systems

Chanchal Pandey[1*], Dipti Verma[2]

[1,2]Department of Computer Science and Engineering, Vishwavidhyalaya Engineering College, Lakhanpur Sarguja, Chhattisgarh, India

[*]*Corresponding Author: bholupandey.bp@gmail.com*

Available online at: www.ijcseonline.org

Accepted:15/Dec/2018, Published: 31/Dec/2018

***Abstract*—** Routers can likewise function as firewalls and perform variety of operations on the incoming and outgoing packets. On the off chance that when every one of the packets share common header attributes, it is named as a packet flow. With a specific end goal to classify a packet, routers perform a query on a classifier table utilizing at least one fields from the packet header to classify the packet into its relating flow. A classifier is a set of rules which distinguish each flow and the fitting actions to be taken for any packet having belonging to that flow. The paper analyzes the problem of packet classification and different proposed systems for the same.

***Keywords*—** Packet classification, dimensions, fields, flows, prefixes.

I. INTRODUCTION

There are different qualities of services given by routers nowadays, for example, packet sending, fair-queue scheduling, resource reservation and access control [1]. Such mechanisms require the router to have the capacity to recognize between various packets and classify them into flows. Flows are characterized [2] as a group of packets which have basic header qualities yet diverse payload. Packets having a place with a similar stream obey predefined rules and are handled in a similar design by the routers.

A collection of such rules is named as a classifier. Each rule in the classifier indicates the flow of a packet may belong to. In general there are 2 stages in any packet classification algorithm [3, 4]: pre-processing stage and classification stage. Pre-processing stage extricates representative data from rules and builds optimized data structures that can capture the dependencies among rules. Such data structures help in finding the least-cost matching rules for every packet that is handled. This stage invoked at whatever new rules are included or erased or if existing rules are changed. Since these operations are rare, the preprocessing stage uses the central CPU of the router. The real parsing of the packets is done in the classification stage where headers are extracted. Utilizing the qualities in the header fields the data structures worked amid the pre-processing stage is navigated to discover the best matching rule.

II. EXECUTION METRICS

There are a few measurements based on which the classification algorithms are compared [3], for example,

A. Speed of searching:
Fast network joins require fast queries, and consequently require a fast system for the classification of packets. Speed is more often than not estimated as far as number of memory accesses required.

B. Fast Updates:
As the classifier changes because of expansion of new rules and cancellation of more seasoned ones, the data structures kept up by the algorithm should be maintained. These data structures can be additionally classified the ones which are incrementally refreshed and the ones which should be worked from the very beginning every time. This isn't a need for the core routers inside the network but the border routers where the larger part of packet filtering occurs.

C. Number of fields:
A proficient algorithm must have the capacity to deal with any number of fields.

D. Memory necessities:
The smaller the memory necessities of the algorithm, the simpler it is to utilize faster memory advances, for example, static random access memories (SRAMs).

E. Implementation flexibility:
An algorithm which can be actualized in both hardware and in addition software is more preferred than an algorithm that has just hardware or software based usage.

F. Flexibility in specification:
A classification algorithm should support general rules, prefixes, operators (less than, greater than, equal to, ranges and so forth.) and wildcards.

III. TAXONOMY OF PACKET CLASSIFICATION ALGORITHMS

The different methodologies for packet classification which have been talked about in the paper can be comprehensively ordered into the accompanying six classifications [3]:

1) Geometrical method: In these method, we see the problem of packet classification geometrically. A rule in measurements represents a d-dimensional hyper rectangle in d-dimensional space. So a classifier is seen as a gathering of hyper rectangles, every one of which is related with a priority. Classification is done in view of finding the rectangle with highest priority that contains the point representing the packet.

2) Divide and Conquer approaches: The fundamental idea behind such approaches is to divide the packet classification problem into numerous longest matching prefix problems, one for each field and subsequently combining the outcomes.

3) Decision Tree based method: A decision tree comprises of two sorts of nodes: leaf nodes and internal nodes. The internal nodes guide the searching to achieve the proper leaf nodes. The leaf nodes contain a single rule or an arrangement of best matching rules. For classification, a search key is developed from the qualities in the packet header field. The tree is navigated utilizing either individual bits or a subset of the bits from the search key for making branching decisions until the point that a leaf node is reached. On the off chance that the leaf node contains a single rule then it is considered as the best matching rule, however in the event that a rundown of rules is put away then a linear search is done to locate the best matching rule.

4) Trie based approach: Trie based methods are a unique instance of decision tree approach. These are a class of decision trees where the searching on fields is done sequentially, branching decisions are made utilizing a single bit of the search key and single rules are put away in the leaf node.

5) Tuple space approach: This approach limits the search scope by partitioning the rules utilizing tuples. It supports packet classification on multiple fields.

6) Hardware based method: Owing to the quickly developing size of the Internet movement, there is a requirement for algorithms that scale well to a large number of packets every second and a huge number of rules are required for the same. The performance of hardware approaches is profoundly subject to the exhaustive searching capability of rules in a memory cycle.

In Table.1 below shows the various algorithms and the different classes they belong to.

Table 1. Taxonomy of Packet Classification Algorithms

Approach	***Related Algorithms***
Decision Tree	Hierarchical Intelligent Cuttings (HiCuts), HyperCuts
Trie	Hierarchical Tries, Set-pruning Tries, Grid-of-tries
Geometrical	Area-based Quadtree, Fat Inverted Segment Tree, Grid-of-tries
Divide & Conquer	Lucent Bit Vector, Aggregated Bit Vector, Cross Producting, Recursive Flow Classification
Tuple Space	Tuple Space Search, Tuple Space pruning
Hardware	Ternary CAMs, Bitmap Intersections

IV. LITERATURE SURVEY

[5] Author survey the information structures that have been proposed for one-dimensional packet classification. Survey is restricted to information structures for the situation when ties among the rules that match an approaching packet are broken by choosing the matching rule that is generally particular. For the situation when the rule channels are goal address prefixes or are nonintersecting extents, this sudden death round relates to longest-prefix or most limited range matching, individually.

[6] In this paper, author portray two new algorithms for tackling the minimum cost matching channel issue at high speeds. Our first algorithm depends on a network of-attempts development and works ideally to process channels comprising of two prefix fields, (for example, goal source channels) utilizing straight space. Our second algorithm, cross-producing, provides fast lookup times for arbitrary filters but potentially requires large storage.

[7] Author present conveyed Cross producing of Field Labels (DCFL), a novel combination of new and existing packet classification methods that Leverages key observation of the structure of genuine channel sets and exploits the abilities of current equipment innovation. Utilizing a gathering of

genuine and engineered channel sets. Creator give examinations of DCFL execution and asset prerequisites on channel sets of different sizes and pieces.

[8] Author propose the utilization of course storing to accelerate layer-4 query, and outline and actualize a reserve design for this reason. Creator examined the territory conduct of the Interenttrafflc (at layer-4) and proposed a close LRU algorithm that best tackle this conduct. In execution, to best estimated completely acquainted close LRU utilizing generally modest set-cooperative equipment, It created a dynamic set-affiliated plan that adventures the decent properties of N-widespread hash capacities.

[9] This paper considered an established algorithm that we adjusted to the firewall area. Creator call the subsequent algorithm "Geometric Efficient Matching" (GEM). The GEM algorithm appreciates a logarithmic matching time execution. In any case, the algorithm's hypothetical most pessimistic scenario space unpredictability is request of n to the intensity of 4, for a rule-base with n rules. In light of this apparent high space unpredictability, GEM-like algorithms were dismissed as unrealistic by before works. In spite of this end, this paper demonstrates that GEM is really a great decision.

[10] Author present a non specific packet classification algorithm, called Tuple Space Search (TSS). Since genuine databases regularly utilize just few unmistakable field lengths, by mapping filters to tuples even a straightforward direct search of the tuple space can give noteworthy speedup over guileless straight search over the filters. Each tuple is kept up as a hash table that can be searched in one memory get to. We at that point present strategies for additionally refining the search of the tuple space, and exhibit their effectiveness on some firewall databases.

[11] In this paper, author present another packet classification algorithm, which can generously enhance the execution of a classifier. The algorithm is based on the perception that a given packet coordinates just a couple of rules even in extensive classifiers, which recommends that the majority of rules are autonomous in any given rule base. The algorithm progressively parcels the rule base into littler free subleases in view of hashing.

TABLE 2: Comparison between existing methods

Author	Dataset Used	Approach Used	Algorithm	Classification Speed	Findings
S. Sahni et al.	Data structure based packet classification	Non-partitioning based approach	Exhaustive Searching	Order of N, Where n is the total number of rules $O(N)$	Proposed method is limited to data structure dataset only. When the rule filters are destination-address prefixes or are nonintersecting ranges, this tie breaker corresponds to longest-prefix or shortest-range matching. It searches all rules that are systematically arranged as described by its priorities.
V. Srinivasan et al.	Router database consisting large number of routing information	Non-partitioning based approach	Cross Producting	450 ns and 900ns lookup times for source and destination prefixes.	Author describe two new algorithms for solving the least cost matching filter problem at high speeds. 1. Grid-of-tries construction which works optimally for filtering rules. 2. Cross producing, which is able to look randomly distributed filters or rules in short duration.
D.E. Taylor et al.	Bloom Filter Array Data structure	Non-partitioning based approach	Cross Producting	100 million searches per second	Author presents distributed cross producing of label fields. Author presents a combination of new and existing packet classification technique that uses Bloom Filter

					Array Data structure for filtering rules.
J. Xu et al.	FIX-WEST network traces	Non-partitioning based approach	Caching based algorithm	Miss ratio is only 8.04 %	Author proposed FIX-WEST traces for analysis with caching algorithm, to speed up layer-4 lookup. Author has designed and implemented the cache architecture for analysis.
D. Rovniagin et al.	GEM database of more than 5000 rules	Partitioning based approach	Decision tree based algorithm	$Best\ case: O(Log(N))$ Worst case: $O(N^4)$	This paper consider traditional algorithm that are implemented to the firewall domain. The proposed method is GEM sorts for Geometric Efficient Matching. It perform well and gives logarametic complexity.
V. Srinivasan et al. (2)	Real firewall database	Partitioning based approach	Tuple space based algorithm	Search time: $O(W^{k-1})$ For k-dimensional filters.	Author a traditional packet classification algorithm, named Tuple search space. It maps the filters to tuples in a linear way. It is possible only because of small size of database. The linear search is efficient which performing searches on small database. Each tuple is maintained as a hash table that can be searched in one memory access.
L. Choi	Network trace database	Partitioning based approach	Hash based algorithm	4.2 access for 5,000 rules, 5.6 access for 10,000 rules, 207 access for 500 thousand rules.	This paper used a new packet classification algorithm, which improves the performance of classifier at some extent. The algorithm is built on the observation that a given packet matches only a few rules even in large classifiers, which suggests that most of rules are independent in any given rule base.

V. CONCLUSION

The problem of packet classification gives a methods for recognizing packets and enabling various servicing for every one of them. The key performance prerequisites of a packet classification algorithm are the quantity of memory accesses required and the capacity prerequisites. There are a few factors that determined the selection of a proper packet classification algorithm, for example, the quantity of rules, size of the network, memory and bandwidth available and limit of processing components. The fitting algorithm can be chosen based on the above criteria.

REFERENCES

[1] T.Y.C. Woo, "A modular approach to packet classification algortithms", Bell Laboratories, Lucent Technologies.

[2] P. Gupta and N. McKeown, "Packet Classification using Hierarchical Intelligent Cuttings", IEEE Micro, pp 34-41, vol. 20, no.1 , January/February2000.

[3] D.Medhi, K.Ramasamy, "Network Routing Algorithms Protocols and Architectures", Morgan Kauffman Series on Networking, pp. 567-579; pp. 704 March 2007.

[4] Kim, K.C. Claffy, M. Formenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic demysitfied : Myths, caveats, and best practices," ACM Conext, 2008.

[5] S. Sahni, K.S. Kim, and H. Lu, "Data structures for onedimensional packet classification using most specificrule matching," Proc. Parallel Architectures, Algorithms and Networks

(I-SPAN '02), pp. 1-12, May 2002, doi:10.1109/ISPAN.2002.1004254.

[6] V. Srinivasan, G. Varghese, S. Suri, and M. Waldvogel, "Fast and scalable layer four switching," Proc. ACM Conf. Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '99), pp. 191-202, Sep. 1999, doi:10.1145/285237.285282.

[7] D.E. Taylor and J.S. Turner, "Scalable Packet Classification using Distributed Crossproducting of Field Labels," Proc. IEEE Conf. Computer and Communications Societies (INFOCOM '05), pp 269-280, Mar. 2005, doi:10.1109/INFCOM.2005.1497898

[8] J. Xu, M. Singhal, and J. Degroat, "A novel cache architecture to support layer-four packet classification at memory access speeds," Proc. IEEE Conf. Computer and Communications Societies (INFOCOM '00), pp. 1445-1454, Mar. 2000, doi:10.1109/INFCOM.2000.832542.

[9] D. Rovniagin and A. Wool, "The Geometric Efficient Matching Algorithm for Firewalls,", IEEE Trans. Dependable and Secure Computing, vol. 8, iss. 1, Jan./Feb. 2011, pp. 147-159, doi:10.1109/TDSC.2009.28.

[10] V. Srinivasan, S. Suri, and G. Varghese, "Packet classification using tuple space search," Proc. ACM Conf. Applications, technologies, architectures, and protocols for computer communication (SIGCOMM '99), pp. 135-146, Aug. 1999, doi:10.1145/316188.316216.

[11] L. Choi, H. Kim, S. Kim, and M.H. Kim, "Scalable Packet Classification Through Rulebase Partitioning Using the Maximum Entropy Hashing," IEEE/ACM Trans. Networking, vol 17, iss. 6, Dec. 2009, doi:10.1109/TNET.2009.2018618.

