2019 32nd International Conference on VLSI Design and 2019 18th International Conference on Embedded Systems (VLSID)

Self-Organizing Maps-based Flexible and High-Speed Packet Classification in Software Defined Networking

Shih-Chang Hung[1], Nick Iliev[2], Balajee Vamanan[3], and Amit Ranjan Trivedi[2]

[1]Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA, Email: hungshih@msu.edu; [2]Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA, Email: niliev4@uic.edu, amitrt@uic.edu; [3]Department of Computer Science, University of Illinois at Chicago, Chicago, IL, USA, Email: bvamanan@uic.edu

***Abstract*—This work studies the application of growing hierarchical self-organization map (GH-SOM) for high performance and flexible packet classification in the context of software-defined networking (SDN). Highly flexible packet classification is necessary for SDN since SDN applications enable fine-grained policies (i.e., increased number of rules) and install packet flow classification rules on-the-fly during flow setup. We show that a hierarchical tree of SOM is fast, flexible, and retraining of SOM-tree is not required when only a small number of rules are updated. Therefore, SOM-tree adeptly absorbs updates in rulesets facilitating a flexible packet classification, a key requirement of SDN. Our results for rulesets generated using *ClassBench* [1] shows high accuracy in packet classification. Classification accuracy is also characterized when network rules are updated on-fly.**

Keywords— *Network packet classification; Software-defined networking; Self-organizing maps*

I. Introduction

Software-Defined Networking (SDN) is a new paradigm that is becoming popular to manage large data centers and enterprise networks [2]. Although, SDN has revolutionized the management of large networks, the architecture of switch data-plane for packet classification has remained virtually unchanged. Packet classification is the problem of identifying the highest priority rule that matches a given network packet out of a set of rules. Each rule specifies the desired action (e.g., drop, forward, tag, etc.) on a collection of packets identified by a combination of packet header fields. Today's switches employ Ternary Content Addressable Memories (TCAM) to match every incoming packet against a set of flow rules at fast line rates. Because TCAMs perform brute-force search against all rules for every incoming packet in the hardware data-plane, TCAMs do not scale well in throughput and power as line-rates and number of rules increase. Moreover, because TCAMs physically order rules based on priority, TCAMs perform poorly with frequent updates to rules. Addressing the present limitations of packet classification in SDN, we discuss the potential of self-organizing maps (SOM) for *scalable* and *flexible* packet classification.

II. Background and Related Works

Previous packet classification approaches fall into two categories – TCAM and algorithmic. A TCAM uses parallel, hardware-based, brute-force search of all rules for every incoming packet, incurring high energy, area overheads and high update complexity. Algorithmic schemes fundamentally try to reduce the search scope overcoming the above limitations of TCAM. The current state-of-the-art algorithmic approaches use decision trees [3]. The algorithms recursively partition the sub-spaces until the sub-spaces only contain a handful of rules that can be manually searched (using TCAMs). While decision trees reduce search complexity, they suffer from some key shortcomings. First, decision trees do not scale well with increasing number of rules, the number of match fields, and line rate. Second, decision trees suffer from high update complexity. Rule updates may increase tree-imbalance to the point where the decision tree needs to be rebalanced to keep the search time under control [4]. Last, the structure of decision trees is highly sensitive to how the rules are spatially distributed in each ruleset.

III. Network Packet Classification by Growing Hierarchical Self-Organizing Map (GH-SOM) Tree

Self-organizing maps (SOMs) are a class of neural network that can cluster a high dimensional input space to a low dimensional grid of discrete indices. A SOM identifies similarity of input patterns and clusters similar patterns onto the same index of its output grid. However, the output grid size in the traditional SOM is fixed. Therefore, the traditional SOM does not adapt to the characteristics of input sample space. Meanwhile, in the network packet classification problem, rule-sets at a router can have different distributions, and the distribution may dramatically change over time due to rule updates from SDN applications. To handle rule updates, we specifically adopt a growing hierarchical self-organizing map (GH-SOM) as presented in [5]. The approach for GH-SOM-based packet classification is shown in Figure 1. A hierarchical GH-SOM-tree is utilized. The top layer in the tree has a fixed size of output grid (i.e., it is a traditional SOM). While the remaining layers in the tree have a GH-SOM architecture, i.e., the output grid size is adapted depending on the applied rule-set at these nodes. Packet rules can be represented as a hyperrectangle where each dimension denotes a field of the rule filter. Each GH-SOM layer in Figure 1 uses volume or centroid of the hyper-rectangle as a feature for rule clustering.

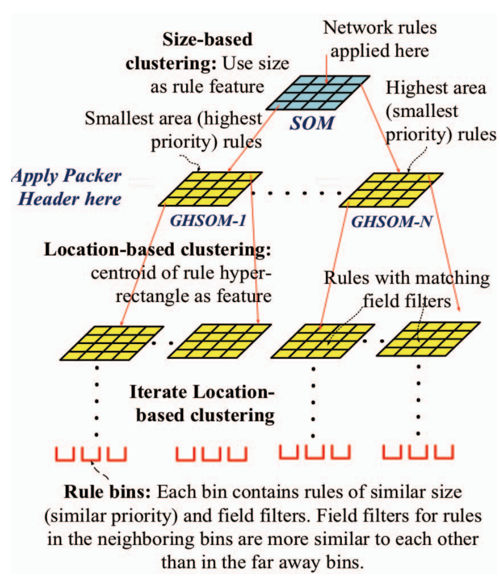

Figure 1: Scheme for SOM-based Network Packet Classification.

2380-6923/19/$31.00 ©2019 IEEE
DOI 10.1109/VLSID.2019.00128

IEEE computer society

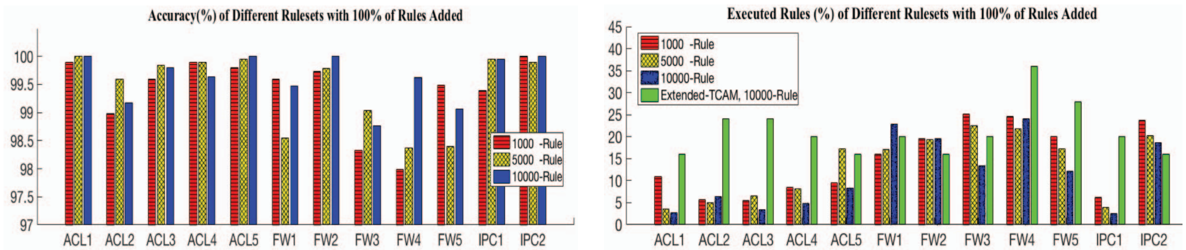

Figure 2: GH-SOM Tree-based Packet Classification for various benchmark rulesets: (a) Accuracy of Different Rulesets. All (100%) rules in the respective ruleset are used for accuracy characterization. (b) Percentage of Executed Rules of Different Rulesets with 100% of Rules Added of SOM algorithm. Executed Rules are those which were searched using TCAM to find the matching rule including radial and backward search. The performance of Extended TCAM is also listed for comparison.

In the training phase, the network rules are applied at the top of the tree in Figure 1. The first layer in the tree clusters rules using their sizes as a feature. The size of a rule is given by volume of the hypercube, whereas the vertices of the hypercube are determined by the rule range across various header fields. Since, lower size rules have a higher priority, the SOM at the top in Figure 1 clusters rules based on priority. The subsequent GH-SOM layers develop finer clusters of rules having a similar centroid location (i.e., similar filters for header fields). Centroid-based clustering is iterated to develop finer clusters of rules of matching centroid locations. Tree for each parent unit can expand both horizontally and vertically depending on two controlling parameters τ_1 & τ_2 as well as the mean quantization error (MQE) of input data [5]. At the leaf of the tree, the network rules are associated with an edge bin. The network rules in a bin match each other regarding their priority and filters for header fields.

In the packet classification phase, a network packet is correlated across hierarchical SOM nodes in the GH-SOM-based tree and eventually mapped to a bin at the leaf of the tree. Subsequently, the TCAM activates only those rules contained in the bins predicted by the SOM-tree. TCAM determines the highest priority matching rule for each network packet based on the rule priority index. If TCAMs do not find any match for the rules predicted by the sub-tree, then only a neighboring sub-tree at the same layer of incrementally larger rules is processed.

IV. Results on Benchmark Rule Sets

The GH-SOM-based packet classification is implemented functionally and examined by different rulesets from *ClassBench* [1]. There are two major advantages of the discussed GH-SOM–based packet classification over the state-of-the-art packet classification method. First, for an incoming packet, the average number of rules executed by TCAM in the SOM-tree algorithm is significantly smaller than in a brute-force search method against all rules. Second, as compared to the software-based algorithm and pure TCAM-based searching scheme, the discussed SOM-tree algorithm is more efficient in dealing with rule updates. Figure 2 presents packet classification accuracy and the number of executed rules by TCAM bins for 36 different rulesets such as access control list (ACL), firewall (FW), and IP chains (IPC) from the *ClassBench*. The largest ruleset contains around 10000 rules and 100% rules are added to the original rule sets. The newly added rules are classified into existing SOM bin without regenerating new SOM-tree and neuron weights. Each ruleset matched against 2000 packets and high accuracy is observed across all the tested rulesets.

Since searching for all rules during each packet lookup would cause exorbitant energy overhead, modern TCAMs are implemented not as a single monolith but as a collection of subarrays. While a naive implementation would search all subarrays sequentially, there are proposals to prune the search among TCAM subarrays. To quantify the reduction in the number of rules searched, we compare our proposal with Extended TCAM [6], a state-of-art TCAM proposal, regarding the fraction of rules searched in large 10,000 rulesets. Figure 2(b) also compares searched rules using extended TCAM, where SOM can significantly reduce the number of searched rules as compared to the alternative approach. Moreover, as emerging technologies-based SOMs achieve ultralow power operation [7], our approach is also expected to harness the benefits of the novel technologies.

V. Conclusions

We proposed a SOM-tree algorithm which is flexible, efficient, and scalable. Notably our proposed SOM-tree does not require re-training for updating a small subset of rules. Because SDN applications require large rulesets, as well as frequent updates to rules, our proposed SOM-tree based classification, is well-suited for SDN.

References

[1] D. E. Taylor and J. S. Turner, "Classbench: A packet classification benchmark," *IEEE/ACM Transactions on Networking (TON),* vol. 15, no. 3, pp. 499-511, 2007.

[2] C.-Y. Hong *et al.*, "Achieving high utilization with software-driven WAN," in *ACM SIGCOMM Computer Communication Review*, 2013, vol. 43, no. 4, pp. 15-26: ACM.

[3] B. Vamanan, G. Voskuilen, and T. Vijaykumar, "EffiCuts: optimizing packet classification for memory and throughput," in *ACM SIGCOMM Computer Communication Review*, 2010, vol. 40, no. 4, pp. 207-218: ACM.

[4] B. Vamanan and T. Vijaykumar, "TreeCAM: decoupling updates and lookups in packet classification," in *Proceedings of the Seventh COnference on emerging Networking EXperiments and Technologies*, 2011, p. 27: ACM.

[5] M. Dittenbach, A. Rauber, and D. Merkl, "Uncovering hierarchical structure in data using the growing hierarchical self-organizing map," *Neurocomputing,* vol. 48, no. 1, pp. 199-216, 2002.

[6] E. Spitznagel, D. Taylor, and J. Turner. Packet classification using extended TCAMs. In Proceedings of IEEE International Conference on Network Protocols - 2003, pages 120 – 131, 2003.

[7] S. Manasi and A. R. Trivedi, "Gate/source-overlapped heterojunction Tunnel FET-based LAMSTAR neural network and its Application to EEG Signal Classification," International Joint Conference on Neural Networks (IJCNN), 2016.