A Configurable Packet Classification Architecture for Software-Defined Networking


Published in:
2014 27th IEEE International System-on-Chip Conference (SOCC)

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2014 IEEE
Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
A Configurable Packet Classification Architecture for Software-Defined Networking

K. Guerra Pérez, X. Yang, S. Scott-Hayward, S. Sezer
The Institute of Electronics, Communications and Information Technology (ECIT)
Queen’s University of Belfast, UK

Abstract— Network management tools must be able to monitor and analyze traffic flowing through network systems. According to the OpenFlow protocol applied in Software-Defined Networking (SDN), packets are classified into flows that are searched in flow tables. Further actions, such as packet forwarding, modification, and redirection to a group table, are made in the flow table with respect to the search results.

A novel hardware solution for SDN-enabled packet classification is presented in this paper. The proposed scheme is focused on a label-based search method, achieving high flexibility in memory usage. The implemented hardware architecture provides optimal lookup performance by configuring the search algorithm and by performing fast incremental update as programmed the software controller.

Keywords—Packet Classification; Software-Defined Network; configurable lookup architecture; lookup algorithms;

I. INTRODUCTION

The exponential growth of new network applications and services, such as virtual machine usage, is overloading network device resources. Software-Defined Networking (SDN) has arisen as a platform to reduce the complexity of network elements.

The SDN platform manages traffic loads in a flexible and more efficient manner by splitting software and hardware resource controls.

One of the great benefits of the SDN architecture is the ability to direct traffic through the network on a flow basis. This enables network service function chaining, for example, where flows are directed through a series of network services depending on the traffic or application type [1].

Packet classification is the main part of flow identification by which the action for each incoming flow at a network device is determined. This action is determined by the Highest Priority Matching Rule (HPMR) from a given filter. It is only necessary that the first packet header of a flow matches the matching rule.

In general, five tuples from packet headers are used for classification: protocol, destination and source ports and source and destination addresses from Layer 3-4 of the Open System Interconnection (OSI) model.

For the next-generation network, packet classification must support high network throughput, e.g. 40-100 Gbps, a wider range of packet fields and consequently a large rule set. It is well known that certain parameters such as scalability, flexibility, capacity, incremental update ability, memory usage and speed, are used to measure the efficiency of lookup systems.

Our hardware implementation based on a configurable search algorithm is well-suited to the programmable platform of SDN, providing greater flexibility than previous methods.

The rest of the paper is organized as follows. In section II, the background to the work is introduced and the related work is discussed. The proposed architecture is presented in section III and the design methodology is described in section IV. In section V, performance evaluation results are presented and discussed. Finally, in section VI, we conclude the paper.

II. RELATED WORK

The process of organizing packets into flows uses multiple fields of the packet header. Each of these fields is defined in diverse syntaxes, such as ranges or prefixes. As a result, each field requires a different matching method, for example Exact Matching (EM), Range Matching (RM) or Longest Prefix Matching (LPM). This presents a challenge to existing packet classification algorithms. In addition, with increasing granularity of the flow definition, an increasing number of flow match entries are held in the flow tables of network devices.

Ternary Content Addressable Memory (TCAM) is a popular method for classification due to its high lookup speed. However, this technique is tending to be replaced owing to disadvantages of high power consumption, storage limitation and the difficulty of rule ternary conversion.

Several algorithms have been proposed for packet classification in which each input packet header must be classified by comparison with a given rule set and processed according to a defined action. These packet classification algorithms can be classified into two categories: multi-field lookup algorithms and single-field lookup algorithms.

Recent research proposed improvements to packet classification efficiency based on the most popular methods, such as HyperCuts [2] or Recursive Flow Classification (RFC) [3]. RFC is of interest due to its high speed performance [4]. Distributed Cross-producting Field Labels (DCFL) [5] is a decomposition method in which individual-field lookups are performed in parallel. The individual results are combined to produce the final result using a label method. Although the lookup performance is high, the memory utilization is inefficient. A DCFL Extended (DCFLE) technique is presented in [6]. This methodology used extended TCAMs (ETCAM) and TCAMs for comparison in a reconfigurable hardware architecture.

Recent research on SDN is focused on conventional packet classification methods adapted for next generation networks.
The EffiCuts algorithm presented in [7] is based on HyperCuts. EffiCuts reduces memory space by reducing the number of overlapped rules but with increased memory access time. The same algorithm was proposed in [8] based on OpenFlow [1]. The authors group rules according to their sizes for efficient storage. The work presented in [9] proposed a HyperCuts based algorithm. These two algorithms require a large number of stored nodes and need a computing stage for multi-dimensional cutting.

A decomposition-field approach is evaluated in [10] on a multi-core processor based on OpenFlow. This research performs parallel field search using a balanced range tree or hash function. The search result from each field is stored in a Bit Vector and the final result is obtained by computing each bit vector. Although this method is proposed to handle a large number of fields, it is not suitable for high-speed lookup in current network systems. Furthermore, the authors do not provide evaluation results for the update operation, which is a key process in packet classification.

### III. PROPOSED ARCHITECTURE

The SDN architecture consists of three layers/planes: Application, Control and Infrastructure/Data plane [1]. This architecture is illustrated in Fig. 1. Note the distinction between a network application, which provides instructions to the control plane (e.g. load-balancing), and a user application, which determines the packet/flow type across the network (e.g. video). For the purpose of this work, we consider the user application with respect to the traffic to be classified and the network application with respect to the rules, which determine how to handle the user traffic. The rules generated at the controller are pushed to the network devices by means of an open protocol such as OpenFlow.

Two different approaches, multi-field and single-field lookup, were analyzed using the most popular packet classification algorithms in our previous work [17]. Table I summarizes the results from the comparison analysis. The most popular algorithms based on different approaches are evaluated in order to compare them against the measures of memory usage and average number of lookup memory access. Lower memory usage is preferable and a lower number of memory accesses implies a higher lookup speed. HyperCuts and RFC are considered as the most popular multi-field lookup algorithms in two different approaches: decision-based tree and decomposition. DCFL focuses on another approach based on the combination of one-dimensional lookup algorithms. Finally, two options are considered as the best results from the comparison between different levels of trie based algorithms. These options are based on a combination of one-dimensional algorithms according to the syntax required for each field. Thus, Option 1 is represented by a 5-level Multi-bit trie, a 4-level Segment trie and a register-based lookup table for IP address fields, port fields and protocol field respectively. Option 2 is formed by a 4-level Multi-bit trie, a 5-level Segment trie and a register-based lookup table.

Our proposed system takes advantage of the flexibility and programmability introduced by SDN. A hardware design based on the optimum lookup approach for a given application is proposed in this paper. It is then possible for the appropriate lookup process to be selected by the Controller. This configurable architecture is explored in this section from the programmability perspective, which is represented in Fig. 2.
A. Controller Functionality

The algorithm structure requires to be updated incrementally to add or delete rules. In our proposed system, the update process is controlled by software (SDN Controller). The architecture is configured with selected algorithms for each dimension according to the specifications of the application. For example, speed is the critical parameter for a Multi-end videoconferencing application supporting real-time connection [11]. The software controller chooses the optimal algorithm combination. In our system, this decision is to select between two possible IP algorithms, which are controlled by a simple signal shown in Fig. 2 (IPalg_s).

Subsequently, the software controller transmits the required information to configure the relevant memory blocks of the hardware platform.

B. Lookup Process

Four pipelined phases are identified in the lookup process. The first phase is stimulated by the Lookup_s signal, which enables the search process in each algorithm. At the same time, the packet header is split into segments, which are sent to the corresponding algorithm selected by the software.

In the second phase, the selected algorithms perform parallel lookup. The result from each algorithm is a pointer to a list of matching labels, which are passed to the next phase.

The third phase shows the combination of the result labels in order to find the HPMR address. This combination is the product of the highest priority label stored in the first position in the list of each output algorithm.

The last phase is to access the memory and obtain the HPMR. The HPMR and, consequently, the associated action are determined by the priority label. Fig. 3 shows the lookup process phases in more detail.

C. Label Method

The label-method proposed in [5] tags each unique rule field avoiding rule duplication. Table II shows an example of the number of unique rule fields. In this case, three ACL filters [12] with different sizes are analyzed for each independent field. Avoiding rule field repetition, the storage requirement can be reduced by more than 50%. The label method is an efficient technique for algorithms with fixed structures such as Multi-bit Trie (MBT), and is not applicable to dynamic structures, which require a re-built algorithm structure.

<table>
<thead>
<tr>
<th>Packet Header Field</th>
<th>acl1 1K (916 rules)</th>
<th>acl1 5K (4415 rules)</th>
<th>acl1 10K (9603 rules)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source IP Address</td>
<td>103</td>
<td>805</td>
<td>4784</td>
</tr>
<tr>
<td>Destination IP Address</td>
<td>297</td>
<td>640</td>
<td>733</td>
</tr>
<tr>
<td>Source Port</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Destination Port</td>
<td>99</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>Protocol</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

That is, the algorithm must reorganize the trie nodes when a new node is created or deleted. This technique reduces the memory space by avoiding replicated rules in the trie.

Taking into account the fact that the single-field lookup algorithms yield better performance, the next logical step is the design of a configurable hardware architecture harnessing this optimized packet classification system for SDN.

D. Memory Blocks

Three different memory blocks can be identified in the proposed architecture: Algorithms, Labels and Rule Filter memory blocks. A set of memory blocks are shared among the selected algorithms. Once the IP labels are determined, the final result is produced, incorporating the Port and Protocol labels attained in earlier stages, which are stored in storage-capacity buffers.

IV. DESIGN METHODOLOGY

A. Update Methodology

A set of binary files are created using C++ with the data needed for the hardware architecture, simulating a control plane of SDN. All cited algorithms are implemented in order to obtain hardware information, such as addresses, data nodes and label list of nodes.

The algorithm combination focuses on obtaining the optimal lookup performance required for each field. Unlike the majority of lookup algorithms, such as TCAM or EffiCuts, the conversion and adaptation of the different rule fields to a lookup method with specific syntax is not required.

The proposed work focusses on the field rule repetitions, providing high flexibility by grouping the single field rules. Thus, when one or more new rules must be inserted in the system, the Controller searches the unique labels for each field in lookup tables (Label Tables). The label tables also contain a counter for each label to support fast incremental update. When a label is not found in the table, three steps are performed: a new label is created, the counter is incremented by 1 and the new rule information is inserted.

However, if the label is found in the Label Table, only the incremental value of the counter is required. Rule deletion is performed according to the same process, deleting the corresponding labels from the table, and only when the counter is zero, the label is deleted from the hardware architecture.

The lists of labels are reorganized according to the priority rule in order to ensure the highest priority matching label (HPML) is in the first position in the list.
The final address to store each rule in the Rule Filter block is performed using a hash function implemented in hardware. A pseudo-code for the design of hardware information files for each algorithm is presented in Fig 4.

This information is based on available on-line filter sets [12] Access Control Lists (ACL), Firewalls (FW) and IP Chains (IPC) with a range of rule sizes, as summarized in Table III.

![Pseudo-code of Rule insertion](image)

**Fig. 4.** Pseudo-code of Rule insertion

<table>
<thead>
<tr>
<th>Filter Types</th>
<th>Names</th>
<th>1 K rules</th>
<th>5 K rules</th>
<th>10 K rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACL</td>
<td></td>
<td>916</td>
<td>4415</td>
<td>9603</td>
</tr>
<tr>
<td>FW</td>
<td></td>
<td>791</td>
<td>4653</td>
<td>9311</td>
</tr>
<tr>
<td>IPC</td>
<td></td>
<td>938</td>
<td>4460</td>
<td>9037</td>
</tr>
</tbody>
</table>

**TABLE III. ANALYSIS OF RULE FILTERS**

### B. Lookup Methodology

Multi-bit Trie (MBT) and Binary Search Tree (BST) algorithms are currently used for IP lookup. The aim of this work is to build a configurable platform choosing between fast IP lookup algorithm (MBT) and efficient-memory-space algorithm (BST).

Multi-bit trie is a special tree structure where the trie depth is defined by the number of bits stored in each level and the branches are characterized by a fixed number of wildcards and exact match nodes. BST is a binary data structure where the left branches contain lower values than the right branches. The tree depth is defined by input prefixes.

It can be seen in Table II that the number of unique rule fields for Port is low. These label search methodologies are explained in section D.

Flow table lookup in SDN requires between 11 and 15 fields from packet headers. However, currently, large real-life rule sets are available with no more than five fields for analysis and performance. The authors in [8] and [9] created a scenario with 12-field rules, which has not been practically applied, while there is no evidence of the validity of the 15-field rules used in [10].

The work presented here focuses on IP field algorithm configuration as the IP address field, due to its length, potentially becomes the bottleneck in lookup performance.

### C. Memory Management

In this section, the architecture is explained in more detail. In terms of memory allocation, all single-field lookup algorithms are implemented in hardware. The major challenge identified is to handle different algorithms without memory explosion. In the Algorithm memory block, a simple Look-Up Table is utilized for Protocol. The protocol value addresses the table where the label is contained.

Registers utilized for Port field lookup contain information about the port values defined in range, high value and low value of port field rule, and the corresponding label. Table IV shows an example of exact matching and range matching of port values. Each unique range is tagged with a unique label.

**TABLE IV. EXAMPLE OF PORT FIELD AND LABELING**

This architecture partitions the IP address field into two 16-bit segments. In other words, two MBT (and BST) algorithms are implemented, one for the high 16-bit source (or destination) IP address field and the other for the low 16-bit source (or destination) IP address field.

Between the two selected IP lookup algorithms, MBT supplies faster lookup. Each MBT algorithm is composed of three memory blocks corresponding to the three levels using 5-bit, 5-bit and 6-bit partitions. A pipelining MBT structure is designed improving the search throughput.

In contrast, BST is implemented in order to achieve more efficient memory usage. Therefore, a simple memory block is designated for each 16-bit segmented IP field.

The assumption is that not all tree/trie nodes will be stored in the memory so that a reduction in storage is possible. The data node contains children node pointers, a counter of labels stored and the pointer to the list of labels.

The IP lookup tree algorithms are implemented in unbalanced memory distribution. Although the lookup process accesses a large memory in some cases, the main limitation is the update process. It can be seen that the number of memory accesses and memory size increase in proportion to the number of tree nodes for rule insertion, which results in sub-optimal memory consumption. However, this methodology implies re-built structure. The perfect balanced memory cannot be achieved. Additionally, a balanced tree algorithm can be easily implemented in software and the information with the new structure can be applied in the architecture for each rule insertion.

1) Label Methodology: The label sizes are 13 bits, 7 bits and 2 bits for IP address, Port and Protocol fields respectively. The label sizes support the number of unique rule fields shown in Table II (e.g. 108 unique destination port fields can be represented by 7 bits).

It is important to maintain the labels stored in priority order in the memory. The priority is included in the given
information with rule and mask fields. This priority value defines the order in the IP label list deposited in the Label memory block. The priority of Port labels is given by exact matching label following by the tightest range matching label. Using Table IV in an example, for an input packet with a destination port field equal to 7812, the labels of Port lookup will be ordered as B, C and A. The priority label for Protocol lookup is determined by the exact matching value.

According to the order, the first label in the list of each lookup algorithm corresponds to the highest priority matching rule address. The first labels are merged in one large data segment (68 bits) in which a hash function is used to obtain the HPMR address.

2) Memory Sharing: The information provided by the software corresponds to all the data that must be stored in a specific address for an explicit memory including the three main memory blocks.

Using the label method, the Rule Filter memory block is treated independently of the chosen algorithm. Moreover, the Label memory block for one field can also be stored without any effect on the chosen algorithm combination. However, any single-field lookup algorithm must obey the following condition:

As the label represents a completed or segmented unique rule field, the packet header must be split into equal size segments for any chosen algorithm.

The benefit is that, the HPML lookup methodology is maintained for any chosen algorithm combination using a simple signal (IPalg_s) to enable one precise input data.

As both IP lookup algorithms are implemented in the hardware architecture, there is no benefit from memory space reduction due to the fact that the hardware synthesis must consider all the different memory blocks for each algorithm. Nevertheless, this research proposes a new approach by sharing memory blocks. The main advantage is to use the same memory for two different input data depending on the selected algorithm. In this system, the MBT level-2 memory requires the same characteristics of dimension and output and input size as the simple BST memory. The rest of the memory determined for MBT can be used to collect more rules. The decision about which data is stored is determined by IPalg_s signal. The consequence is flexibility of memory storage and with efficient allocation data. Fig. 5 represents an example of memory sharing for one segmented IP field.

Data 1 and Data 2 represent the node information corresponding to MBT and BST respectively and Data 3 represents the rule information. According to the IPalg_s signal the node information of Data 1 or Data 2 is stored in the Level 2 memory and Data 3 or Data 2 is selected to store in the rest of the memory block.

V. PERFORMANCE EVALUATION

A. Memory Accesses For Update

One of the benefits of this architecture based on SDN is fast update. Rule Insertion and deletion are achieved by a simple memory upload in two clock cycles per rule; one cycle to store source information and one clock cycle to store destination information due to the limitations of number of input-output pins available. As previously mentioned, an additional clock cycle is required to obtain the rule address using hash function implemented in hardware.

B. Memory Accesses For Lookup

Protocol and Port field methods realize very fast lookup. The protocol label search is executed in a single clock cycle. The Port lookup process produces the labels in two clock cycles.

The MBT structure is executed in the pipelining stage and, thus supports high throughput. MBT performs the lookup in the latency of 6 clock cycles. As expected, the BST algorithm requires a large number of clock cycles per input packet.

The algorithms provide the pointer to labels demanding one more cycle for the entire lookup process.

Finally, two more cycles are required for the final result processing in each case. This final phase is pipelined.

C. Memory Consumption

The proposed architecture was implemented on Altera’s StratixV FPGA which contains a memory capacity of 54 Mbits. The memory usage of the designed architecture consumes 4% of total memory for the mentioned algorithms, lists of labels and the rules filter. The proposed circuit is able to operate 42.73 Gbps with 40 bytes per packet as minimum size, overcoming the OC-768 rate. This is not achieved in software-based systems such as [8] and [10]. Table V summarizes the synthesis results.

TABLE V. SYNTHESIS RESULT ON ALTERA STRATIX V DEVICE

<table>
<thead>
<tr>
<th>Logical Utilization</th>
<th>79,835 / 225,400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total block memory bits</td>
<td>2,097,184 / 54,476,800</td>
</tr>
<tr>
<td>Total registers</td>
<td>129,273</td>
</tr>
<tr>
<td>Maximum Frequency</td>
<td>133.51 MHz</td>
</tr>
<tr>
<td>Total Number Pins</td>
<td>500 / 908</td>
</tr>
</tbody>
</table>

Table VI summarizes the lookup performance from the simulation for the IP algorithm selected, MBT or BST, in terms of memory accesses per packet. Note that, for a comparison, this measure is according to a unique packet, while the lookup process in MBT is a pipelining design to obtain higher throughput. The memory requirement and the number of rules that can be stored using the same number of memory blocks is also presented in the Table VI.
It can be seen that MBT realizes a fast lookup process while BST can support a greater number of rules. If the network application requires MBT for fast lookup, an external memory can be used for a rule filter with more than 8K rules. Although BST has notable disadvantages in search speed, a 12K rule filter can be stored using embedded memory, which is attractive for applications with large rule filters.

Our approach holds high flexibility capacity and the architecture can be adapted with different single-field lookup algorithms that obey certain conditions for label method and where the packet header can be segmented.

The work presented in [9] proposed a HyperCuts based algorithm. In this case, the need for syntax alteration and the linear search is critical. However in our system the algorithms are chosen according to the field characteristics.

Although the architecture in [9] supports incremental update, this is performed in hardware and the complexity is high in comparison with the proposed architecture in this paper. According to the results in [9], HyperCuts requires large memory usage. However, it is able to achieve high throughput for 5 fields and is a candidate technique to be implemented in our configurable lookup architecture.

DCFLE groups unique rule field value sets according to the exact matching values. For that, this algorithm uses TCAMs for IP address and ETCAM for the rest of the fields for range comparison. This architecture is implemented in pipelined stages. However, it cannot achieve target throughputs for the current line rate.

For further evaluation, Table VII compares our design against different hardware designs based on 5-field packet classification, assuming a packet size of 40 bytes. In the future, the system must be able to handle a minimum number of 15 header fields to support the current OpenFlow protocol.

### TABLE VI. PERFORMANCE EVALUATION FOR IP ALGORITHM

<table>
<thead>
<tr>
<th>Lookup Memory Accesses (clock cycle)</th>
<th>Memory Space Required</th>
<th>Number of Stored Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBT</td>
<td>1 per packet</td>
<td>543 Kbits</td>
</tr>
<tr>
<td>BST</td>
<td>16 per packet</td>
<td>49 Kbits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Memory Space (Mb)</th>
<th>Number of Stored Rules</th>
<th>Throughput (Gbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our system with MBT</td>
<td>2.1</td>
<td>8K</td>
<td>42.73</td>
</tr>
<tr>
<td>Our system with BST</td>
<td>2.1</td>
<td>12K</td>
<td>2.67</td>
</tr>
<tr>
<td>Optimizing HyperCuts [9]</td>
<td>4.90</td>
<td>10K</td>
<td>80.23</td>
</tr>
<tr>
<td>DCFLE [4]</td>
<td>1.77</td>
<td>128</td>
<td>16</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

The principal contribution of this research is a highly configurable parallel lookup architecture for IP flow classification. The proposed and prototyped lookup architecture targets primarily SDN systems providing a fast update through software programmability. It offers optimal lookup performance by configuring the best performing set of algorithms for a given type of flow entries. Efficient memory utilization is also achieved by sharing memory resources among multiple lookup algorithms. An example lookup configuration of a multi-bit trie lookup algorithm achieves 133 million lookups per second. Assuming an average IP packet size of 100 bytes, this is equivalent to over 100 Gbit/s link throughput, surpassing lookup targets of emerging high-performance network technology. If configured as binary search trie, a higher rule storage capacity can be achieved. The proposed architecture can be extended with new custom-purpose algorithms in the configuration-set so that the scope of the lookup circuit can be widened to new SDN applications and flow entries. This solution is ideally suited for the evolution of SDN with granular flow classification for emerging SDN applications.

REFERENCES


