T-Cache: Efficient Policy-Based Forwarding Using Small TCAM

Ying Wan, Haoyu Song, *Senior Member, IEEE*, Yang Xu, *Senior Member, IEEE*, Yilun Wang, Tian Pan, *Senior Member, IEEE*, Chuwen Zhang, Yi Wang, *Senior Member, IEEE*, and Bin Liu, *Senior Member, IEEE*

***Abstract*— Ternary Content Addressable Memory (TCAM) is widely used by modern routers and switches to support policy-based forwarding due to its incomparable lookup speed and flexible matching patterns. However, the limited TCAM capacity does not scale with the ever-increasing rule table size due to the high hardware cost and high power consumption. At present, using TCAM just as a rule cache is an appealing solution, but one must resolve several tricky issues including the rule dependency and the associated TCAM updates. In this paper, we propose a new approach which can generate dependency-free rules to cache. By removing the rule dependency, the complex TCAM update problem also disappears. We provide the complete T-cache system design including slow path processing and cache replacement, and implement a T-cache prototype on Barefoot Tofino switches. We conduct comprehensive software simulations and hardware experiments based on real-world and synthesized rule tables and packet traces to show that T-cache is efficient and robust for network traffic in various scenarios.**

***Index Terms*— Ternary rule, cache, TCAM, forwarding.**

I. Introduction

POLICY-BASED forwarding uses a set of packet header fields as the flow ID to match against a predefined rule set and applies the associated policy of the best match to the packet. Today, as the Software-Defined Networking (SDN) [2] prevails, network operation is evolving to be more and more application-aware and service-oriented. As an indispensable component in modern routers and switches, policy-based forwarding plays the roles far beyond the conventional Access Control List (ACL) [3] and Firewall (FW). To name a few, Service Function Chaining [4] uses the rule set to classify network traffic and applies different service chains to different flows; Segment Routing [5] uses the rule set to forward packets on different paths with a source routing mechanism; Network Slicing [6] uses the rule set to assign user flows to different virtual layers which bear different QoS treatments; Network Telemetry [7] uses the rule set to pick specific packets for behavior monitoring and performance measurement.

However, policy-based forwarding is also a notoriously challenging problem. A rule is usually an aggregation of multiple flows and rules may overlap. As a result, a packet can match multiple rules and the matching rule with the highest priority needs to be found. This process must be fast enough to sustain the line-speed forwarding. Unlike address-based forwarding, traditionally policy-based forwarding lacks efficient algorithmic solutions to sustain the ever-increasing network throughput which is now in the magnitude of terabits per second per device.

The recourse to Ternary Content Addressable Memory (TCAM) is effective for now. Using TCAM for policy-based forwarding has become the *de facto* industry standard for two reasons: (1) TCAM allows a search key extracted from an incoming packet to compare with all the stored rules in parallel, thus ensuring line-speed forwarding; (2) TCAM rules support a wide range of patterns, including exact matching, Longest Prefix Match (LPM), and range matching, which are flexible enough for policy representation.

However, on the one hand, TCAM has long been criticized for high hardware cost and high power consumption. TCAM roughly consumes 100× more hardware resources and 100× more power than DRAM of the same capacity [8]. As a result, TCAM's capacity is limited. Most commercial switches can only support a relatively small number of rules in TCAM, ranging from a few hundreds to ten thousands [8].

Moreover, TCAM update, especially for rule insertion, is a slow process during which the lookup is paused and a number of existing rules in TCAM need to be relocated due to the rule dependency [9], which poses serious challenges to dynamic policy deployment in large networks [10]–[16] where the update latency requirement is stringent. Fig. 1 shows the measurements on a high-end commercial switch EdgeCore Wedge 100BF-32X [17]. For a 1K-entry TCAM-based rule table, it takes less than 40us to insert a rule that does not

Manuscript received December 16, 2020; revised April 27, 2021 and July 3, 2021; accepted July 8, 2021; approved by IEEE/ACM Transactions on Networking Editor D. Hay. Date of publication July 28, 2021; date of current version December 17, 2021. The work of Ying Wan, Yi Wang, and Bin Liu was supported in part by the National Natural Science Foundation of China under Grant 62032013, Grant 61872213, and Grant 61432009; and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019B1515120031. An early version of this work [1] was presented in part at the Annual IEEE International Conference on Computer Communications (INFOCOM), 2020. That paper has been modified and revised to reflect comments received at the conference. *(Corresponding authors: Yi Wang; Bin Liu.)*

Ying Wan, Yilun Wang, and Chuwen Zhang are with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: wany16@mails.tsinghua.edu.cn).

Haoyu Song is with Futurewei Technologies, Santa Clara, CA 95050 USA (e-mail: shykcl@gmail.com).

Yang Xu is with the School of Computer Science, Fudan University, Shanghai 200433, China (e-mail: xuy@fudan.edu.cn).

Tian Pan is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China.

Yi Wang is with the University Key Laboratory of Advanced Wireless Communications of Guangdong Province, Southern University of Science and Technology, Shenzhen 518055, China, and also with the PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: wy@ieee.org).

Bin Liu is with the Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China, and also with the PCL Research Center of Networks and Communications, Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: liub@mail.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TNET.2021.3098320