Cost Minimization for Rule Caching in Software Defined Networking

Huawei Huang, *Student Member, IEEE*, Song Guo, *Senior Member, IEEE*, Peng Li, *Member, IEEE*, Weifa Liang, *Senior Member, IEEE*, , and Albert Y. Zomaya, *Fellow, IEEE*

Abstract—Software-defined networking (SDN) is an emerging network paradigm that simplifies network management by decoupling the control plane and data plane, such that switches become simple data forwarding devices and network management is controlled by logically centralized servers. In SDN-enabled networks, network flow is managed by a set of associated rules that are maintained by switches in their local Ternary Content Addressable Memories (TCAMs) which support high-speed parallel lookup on wildcard patterns. Since TCAM is an expensive hardware and extremely power-hungry, each switch has only limited TCAM space and it is inefficient and even infeasible to maintain all rules at local switches. On the other hand, if we eliminate TCAM occupation by forwarding all packets to the centralized controller for processing, it results in a long delay and heavy processing burden on the controller. In this paper, we strive for the fine balance between rule caching and remote packet processing by formulating a minimum weighted flow provisioning (**MWFP**) problem with an objective of minimizing the total cost of TCAM occupation and remote packet processing. We propose an efficient offline algorithm if the network traffic is given, otherwise, we propose two online algorithms with guaranteed competitive ratios. Finally, we conduct extensive experiments by simulations using real network traffic traces. The simulation results demonstrate that our proposed algorithms can significantly reduce the total cost of remote controller processing and TCAM occupation, and the solutions obtained are nearly optimal.

Index Terms—Software-defined networking, approximation algorithm, ternary content addressable memories

1 Introduction

Software defined networking (SDN) is regarded as one promising next generation network architecture [1], [2], [3]. By shifting the control plane to a logically centralized controller, SDN offers programmable functions to dynamically control and manage packets forwarding and processing in switches making it easy to deploy a wide range of network management policies and new network technologies, e.g., traffic engineering [4], [5], [6], quality of service (QoS) [7], [8], security/access control management [9], [10], [11], failure diagnosis [12] and failover mechanisms[13], [14], [15].

In SDN, each network flow is associated with a set of rules, such as packet forwarding, dropping and modifying, that should be installed at switches in terms of flow table entries along the flow path. SDN-enabled switches maintain flow rules in their local TCAMs [16], [17], [18], which support high-speed parallel lookup on wildcard patterns. A typically flow setup process [2], [19] between a pair of users, say users A and B, in SDN contains three steps. 1) User A sends out packets after connection initialization. Once a packet arrives a switch without matched flow table entries, this packet is forwarded to the controller. 2) Upon receiving the packet, the controller decides whether to allow or deny this flow according to network management policies. 3) If the flow is allowed, the controller installs corresponding rules to all switches along the path, such that consecutive packets can be processed by the installed rules locally at switches. Note that switches usually set an expiration time for rules, which defines the maximum rule maintenance time when no packet of associated flow arrives.

In practice, network flow shows various traffic patterns. For example, we show real-time traffic of four network flows [20] in Fig. 1, where some are burst transmission while the others have consecutive packet transmissions for a long time. For consecutive transmission, only the first packet experiences the delay of remote processing at the controller, and the rest will be processed by local rules at switches. However, for burst transmission, the corresponding rules cached in switches will be removed between two batches of packets if their interval is greater than the rule expiration time. As a result, remote packet processing would be incurred by the first packet of each batch, leading to a long delay and high processing burden on the controller. A simple method to reduce the overhead of remote processing is to cache rules at switches within the lifetime of network flow, ignoring the rule expiration time. Unfortunately, network devices are equipped with limited-space TCAMs because they are expensive hardware and extremely power-hungry. For instance, it is reported that TCAMs are 400 times more expensive [18] and 100 times more power-consuming [21] per Mbit than RAM-based storage. Since TCAM space is shared by multiple flow in networks, it is

- *H. Huang, S. Guo, and P. Li are with the School of Computer Science and Engineering, the University of Aizu, Japan. E-mail: {d8152101, sguo, pengli}@u-aizu.ac.jp.*
- *W. Liang is with the Research School of Computer Science, the Australian National University, Australia. E-mail: wliang@cs.anu.edu.au,.*
- *A.Y. Zomaya is with the School of Information Technologies, the University of Sydney, Australia. E-mail: albert.zomaya@sydney.edu.au.*

Manuscript received 30 Nov. 2014; revised 26 Apr. 2015; accepted 28 Apr. 2015. Date of publication 10 May 2015; date of current version 16 Mar. 2016. Recommended for acceptance by S. Yu.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPDS.2015.2431684

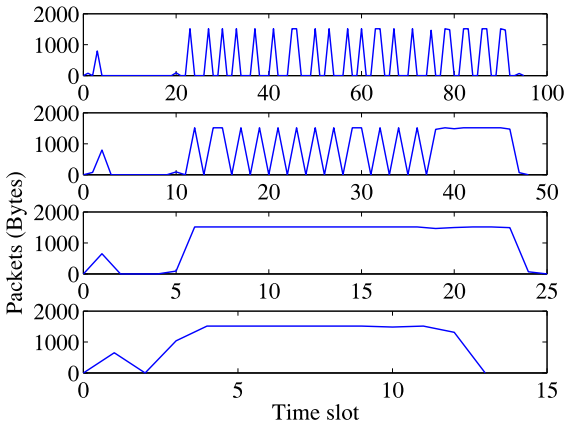


Fig. 1. The sniffed TCP traffic flow [20].

inefficient and even infeasible to maintain all rules at local switches. This dilemma motivates us to investigate efficient rule caching schemes for SDN to strive for a fine balance between network performance and TCAM usage.

The main contributions of our paper are summarized as follows.

- To the best of our knowledge, we are the first to study the rule caching problem with the objective of minimizing the sum of remote processing cost and TCAM occupation cost.
- We propose an offline algorithm by adopting a greedy strategy if the network traffic is given in advance. We also devise two online algorithms with guaranteed competitive ratios.
- Finally, we conduct extensive simulations using real network traffic traces to evaluate the performance of our proposals. The simulation results demonstrate that our proposed algorithms can significantly reduce the total cost of remote controller processing and TCAM occupation, and the solutions obtained are nearly optimal.

The reminder of the paper is structured as follows. Section 2 reviews some related work in table entries scheduling in SDN. Section 3 introduces the system model and problem formulation. An offline algorithm is proposed in Section 4. Two online algorithms are proposed and analysed in Section 5. Section 6 demonstrates the performance evaluation results. Finally, Section 7 concludes this paper.

2 Related Work

As one pioneering work of recouping control plane from the data plane, NOX [9] has been proposed to control data forwarding based on OpenFlow [1], [22].

Following this line of research, lots of efforts have been made on rule caching strategies in SDN, which can be classified into two categories: reactive way [2], [9] and proactive way [3], [22], [23].

The reactive rule caching has been widely adopted by existing work [2], [9] because of its efficient usage of TCAM space. The first packet of each "microflow" is forwarded to the controller that reactively installs flow entries in switches. For instance, Ethane [2] controller reactively installs flow table entries based on the first packet of each TCP/UDP flow. Recently, Bari et al. [19] use the on-demand approach to response flow setup requests.

On the other hand, other studies [24], [25] argue that reactive approach is time-consuming because of remote rule fetching, leading to heavy overhead in packet processing [4], [19], [23]. To reduce the response time for packets at switches without matched rules, proactive approach has been proposed to install rules in switches before corresponding packets arrive. For example, Benson, et al. [4] developed a system MicroTE that adapts to traffic fluctuations, with which rules can be dynamically updated in switches to imposes minimal overhead on network based on traffic prediction. Kang, et al. [3] have proposed to precompute backup rules for possible failures and cache them in switches in advance to reduce network recovery time.

In addition, other related literatures [6], [18], [26], [27] focus on the rules scheduling considering forwarding table size utilization. For instance, Katta et al. [18] proposed a abstraction of an infinite switch based on an architecture that leverages both hardware and software, in which rules caching space can be infinite. In that case, rules can be cached in forwarding table as many as possible. This abstraction saves TCAMs space, but the packet processing speed in switch is a bottleneck. To efficiently use TCAMs space, Kanizo et al. [6], Nguyen et al. [26] and Cohen et al. [27] propose their rules placement scheduling jointly consider the traffic routing in network. However, rules updating is ignored in their optimization. To the contrast, we study both the two aspects in our optimization.

The work most related with our paper is DIFANE [23], a compromised architecture that leverages a set of authority switches serving as a middle layer between the controller in control plane and switches in data plane. The endpoints rules are pre-computed and cached in authority switches. Once the first packet of a new microflow arrives the switch, the desired rules are reactively installed, from authority switches rather than the controller. In this way, the flow setup time can be significantly reduced. Unfortunately, caching pre-computed rules all in authority switches consumes large TCAM space. In our work, we still load the flow rules into switches in a reactive way. However, rule caching period is controlled by our proposed algorithm by taking both remote processing and TCAM occupation cost into consideration.

3 System Model and Problem Formulation

We consider a discrete time model, where the time horizon is divided into T time slots of equal length l. A network flow travels along a set of SDN-enabled switches, each of which is assigned a set of rules to implement routing or network management functions. For simplicity, in this paper, we study rule caching at a switch with a set of associated rules whose maintenance cost is α per second, and our results can be directly extended to all switches along the flow path. Note that the set of rules associated with the network flow will be cached or removed as a whole at the switch. When no matched flow table entries are available at the switch for the arriving packets, these packets will be sent to the controller for processing. We model the remote processing cost at controller as β. The ratio between these

TABLE 1
Notions and Variables

Notations	Description
T	maximum time-slot range of consideration
l	length of each time slot (seconds)
a_t	a binary indicator that denotes the positive flow rate in time slot t, $t = 1, 2, \ldots T$
δ	sum of time slots where $a_i > 0$
D	amount of valid periods during $[1, T]$
V_i	the ith valid period during $[1, T]$
E_i	the ith empty period during $[1, T]$
α	occupation cost of caching table entry
β	remote processing cost
γ	$\gamma = \frac{\alpha}{\beta}$
x_t	a binary variable indicating whether cache action happens in time slot t, $t = 1, 2, \ldots T$
y_t	a binary variable indicating whether fetch action happens in time slot t, $t = 1, 2, \ldots T$
C_C	total cost of cache action during $[1, T]$
C_F	total cost of fetch action during $[1, T]$

two kinds of cost is denoted by γ, i.e., $\gamma = \frac{\alpha}{\beta}$. We consider arbitrary traffic pattern of the network flow. The set of time slots with ($a_t > 0$) and without ($a_t = 0$) packet transmission are referred to as valid period and empty period, respectively. All symbols and variables used in this paper are summarized in Table 1.

We define a binary variable x_t to denote whether a flow entry is cached at the tth time slot:

$$x_t = \begin{cases} 1, & \text{if rules are cached at the } t\text{th time slot,} \\ 0, & \text{otherwise.} \end{cases}$$

The occupation cost can be calculated as follows:

$$C_C = \alpha l \cdot \sum_{t=1}^{T} x_t. \tag{1}$$

We also define a binary variable y_t to indicate whether remote packet processing is conducted.

$$y_t = \begin{cases} 1, & \text{if rules are remotely fetched in the } t\text{th} \\ & \text{time slot,} \\ 0, & \text{otherwise,} \end{cases}$$

and the corresponding cost can be calculated by:

$$C_F = \beta l \cdot \sum_{t=1}^{T} y_t. \tag{2}$$

With the global information of the given traffic flow, the minimum weighted flow provisioning (**MWFP**) problem can be formulated as follows:

$$\mathbf{MWFP}: \min_{x_t, y_t} C_{Total} = C_C + C_F \quad \text{s.t.} \quad x_t - x_{t-1} \le y_t,\ t = 2, 3, .., T, \tag{3a}$$

$$\sum_{j=1}^{t} y_j \ge x_t,\ t = 1, 2, .., T, \tag{3b}$$

$$(x_t + y_t) \ge a_t,\ t = 1, 2, .., T, \tag{3c}$$

$$x_t, y_t \in \{1, 0\},\ t = 1, 2, .., T. \tag{3d}$$

The trigger of remote processing is represented by constraint (3a): the fetch must be made ($y_t = 1$) if the required rules are not available at the $(t-1)$th time slot ($x_{t-1} = 0$) and will be in the catch on the tth time slot ($x_t = 1$). Constraint (3b) indicates that the switch needs to fetch rules at least one time before caching. Constraint (3c) claims that arriving packets must be processed by local cached rules or remote controller. Note that the input of this problem are a_t, α and β, and the output is the scheduling solution, i.e., x_t and y_t, $t = 1, 2, ...T$.

4 Offline Algorithm

In this section, we propose a heuristic algorithm to solve the offline version of the problem. As shown in Algorithm Offline Greedy, we first generate a collection of time slot sets that represents possible rule cache periods as $\mathcal{S}$ ={ {1},{2},. . ., {T}, {1, 2}, {2, 3}, . . ., {T-1, T}, {1, 2, 3}, {2, 3, 4}, . . ., {T-2, T-1, T}, . . ., {1, 2, . . ., T-1}, {2, 3, . . .,T}, {1, 2, . . ., T}}. Each set is assigned a weight according to (4) in line 2

$$w(S_j) = \begin{cases} \beta l + \alpha l \cdot |S_j|, & \text{if } |S_j| > 1; \\ \beta l \cdot a_t, & \text{if } |S_j| = 1, t \in S_j, \forall S_j \in \mathcal{S}. \end{cases} \tag{4}$$

Then, we select time slot sets for rule caching in an iterative manner from line 4 to 8. In each iteration, we choose the set X that minimizes the value of weight $w(X)$ divided by number of elements not yet covered.

Algorithm 1. Offline Greedy Algorithm to solve **MWFP**

Input: flow indicator set $\mathcal{F}$ = {$a(t)$, $t \in [1, T]$}, and $\mathcal{T}$={1, 2, . . ., T}
Output: The collection C of subsets of $\mathcal{T}$
1: generate sample collection $\mathcal{S}$ with $\bigcup_{S \in \mathcal{S}} S = \mathcal{T}$
2: generate the weight function w: $\mathcal{S} \to \mathbb{R}_+$ by invoking (4)
3: $C \leftarrow \emptyset$, and $R \leftarrow \mathcal{T}$
4: **while** $R \neq \emptyset$ **do**
5: $X \leftarrow \arg\min_{X \in \mathcal{S}} \frac{w(X)}{|X \cap R|}$
6: $C \leftarrow C \cup X$
7: $R \leftarrow R \setminus X$
8: **end while**

Remark 1. The computing complexity of Algorithm 1 is $O(T \log T)$, if we use binary search tree in line 5 of Algorithm 1.

Proof. In the first, the main loop iterates for $O(T)$ time. Then, X in line 5 can be found in $O(\log\ m)$ time while using binary search tree, where m is the number of sets in instance $\mathcal{S}$. Because $|\mathcal{S}| = \frac{1}{2}(T^2 + T)$, we obtain the total computational time $O(T) \times\ O(\log(\frac{1}{2}(T^2 + T))) = O(T)O(\log(T^2)) = O(T)O(2 \log T) = O(T \log T)$. □

Theorem 1. *Algorithm 1 is* $(\ln T + 1)$*-approximation to the optimal solution of* ***MWFP****, where* T *is the maximum time slot.*

Proof. For each element (time slot) $t_j \in \mathcal{T}$, let S_j be the first picked set that covers it while applying Algorithm 1, and $\theta(t_j)$ denote the amortized cost of each element in S_j,

$$\theta(t_j) = \frac{w(S_j)}{|S_j \cap R|}.$$

Obviously, the cost of Algorithm 1 can be written as $\sum_{t_j \in \mathcal{T}} \theta(t_j)$.

Then, let $\widehat{\mathcal{T}} = \{t_1, t_2, \ldots, t_T\}$ denote the ordered set of elements in [1,T] that each is covered. Note that, when t_j is to be covered, apparently we have $R \supseteq \{t_j, t_{j+1}, \ldots, t_T\}$. We can see that R contains at least $(T - j + 1)$ elements. Therefore, the amortized cost in S_j is at most the average cost of the optimum solution (denoted by OPT), i.e.,

$$\theta(t_j) = \frac{w(S_j)}{|S_j \cap R|} \leq \frac{\texttt{OPT}}{T - j + 1}.$$

By summing the $\theta(t_j)$ in all time slots, we get:

$$\sum_{t_j \in \mathcal{T}} \theta(t_j) \leq \texttt{OPT}\left(\frac{1}{T} + \frac{1}{T-1} + \cdots + \frac{1}{2} + 1\right).$$

That is $\texttt{Greedy} \leq \texttt{OPT} \cdot H(T) \leq \texttt{OPT} \cdot (1 + \ln T)$, where $H(T)$ is called the harmonic number of T. □

5 Online Algorithms

In this section, we consider the **MWFP** problem assuming that the packet traffic information is not given in advance. We first present several important observations in the optimal solution, followed by two proposed online algorithms with low computational complexity to approximate the optimal solution.

5.1 Typical Actions in Optimal Solutions

By carefully examining the optimal solutions of several problem instances, we find that there exists several typical actions as follows.

- *Fetch and cache (*FnC): for valid periods with at least two time slots, flow rules are first fetched from the remote controller, and then they are cached at local switches.
- *Successive fetch* (SuF): for a valid period with at least two time slots, all packets are forwarded to the remote controller for processing.
- *Cache in empty period* (CiE): rules are cached in switches during the empty period between two valid periods.
- Only Forward (NF): Packets are processed by the controller in the empty periods of one time slot.

Above typical actions are illustrated in Fig. 2. The optimal solutions can be categorized into following three cases.

- **OPT-A**: there are only SuF actions in the optimal solution.
- **OPT-B**: there are only CiE actions in the optimal solution.
- **OPT-C**: both SuF and CiE actions exist in the optimal solution.

5.2 Online Exactly Match the Flow (EMF) Algorithm

Our first online algorithm, which is referred to as EMF (Exactly Match the Flow Algorithm), is shown in Algorithm

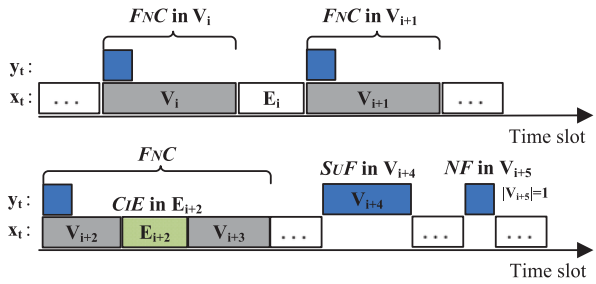


Fig. 2. Illustration of typical actions in optimal solution.

2. In each time slot, each switch makes a decision, caching or fetching, according to observed network traffic. When there are packets arriving in the current time slot, i.e., $a_t = 1$, if no matched rules are cached, i.e., $x_{t-1} = 0$, switch fetches rules from the controller. Otherwise, we keep caching them in switches.

Lemma 1. *Suppose there are D valid periods (denoted by V_i, $i = 1, 2, \ldots, D$) including δ valid time slots within $[1, T]$, the total cost of EMF is*

$$C_{\text{EMF}} = \beta l D + \alpha l \delta. \tag{5}$$

Proof. In Algorithm 2, flow rules are maintained in switches only when there are network traffic passing through, such that TCAM occupation cost can be easily calculated by $\alpha l \delta$. Since rule fetching action happens in the beginning of each valid period, we have fetching cost of $\beta l D$. By summing them up, the total cost of EMF can be calculated by (5). □

Algorithm 2. Online Exactly Match the Flow Algorithm

```
for each time slot t ∈ [1, T] do
  if a_t=1 and x_{t-1}=0 then
    fetch flow rules from the controller
    y_t ← 1, x_t ← 1
  else if a_t=1 and x_{t-1}=1 then
    keep caching entries in switches
  else if a_t=0 and x_{t-1}=1 then
    remove the corresponding flow table entries
    x_t ← 0
  end if
end for
```

Lemma 2. *When the optimal solution belongs to OPT-A, SuF is adopted by any valid period V_i, we have $\gamma > \frac{|V_i|-1}{|V_i|}, \forall i = 1, 2, \ldots, D$.*

Proof. Since SuF is adopted in the optimal solution, its cost must be less than FnC, i.e.,

$$|V_i|\beta l < \beta l + |V_i|\alpha l$$
$$\Rightarrow \beta < \frac{|V_i|\alpha}{|V_i| - 1} \Rightarrow \gamma > \frac{|V_i| - 1}{|V_i|}, \forall i = 1, 2, \ldots, D.$$

□

Theorem 2. *When the optimal solution belongs to OPT-A, the EMF algorithm is $(\gamma + \frac{D}{\delta})$-competitive.*

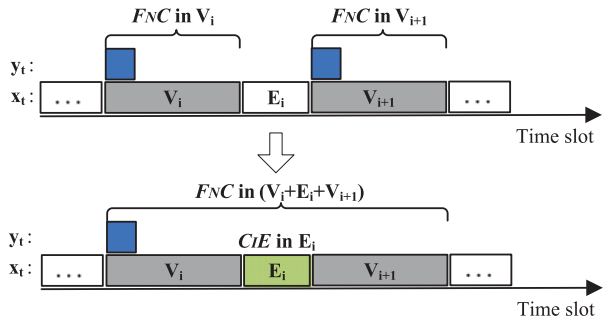


Fig. 3. Rules are cached in E_i because of CIE action.

Proof. Since the optimal solution belongs to OPT-A, i.e., all packets are sent to the controller for processing, it is easy to see that the total cost of optimal solution is $\beta l\delta$. Thus, the competitive ratio λ_A is:

$$\widehat{\lambda}_A = \frac{C_{\text{EMF}}}{\beta l\delta} = \frac{\alpha\delta + \beta D}{\beta\delta} = \gamma + \frac{D}{\delta}.$$

□

Lemma 3. *When the optimal solution belongs to OPT-B, rules are always maintained by switches once downloaded, i.e., CIE is adopted by the period E_i between any two valid periods V_i and V_{i+1}, and we have $\gamma < \frac{1}{|E_i|}, \forall i = 1, 2, \ldots, D-1$.*

Proof. If switches continue to cache rules once they're downloaded, the TCAM occupation cost during E_i must be less than the fetching cost at the beginning of V_{i+1}, i.e.,

$$\begin{aligned} &\alpha l(|E_i| + |V_{i+1}|) < \beta l + |V_{i+1}| \cdot \alpha l \\ &\Rightarrow \beta > \alpha|E_i| \Rightarrow \gamma < \frac{1}{|E_i|}, \forall i = 1, 2, \ldots, D-1. \end{aligned}$$

□

Theorem 3. *When the optimal solution belongs to OPT-B, the EMF algorithm is $(\frac{D+\gamma\delta}{1+\gamma(\delta+D-1)})$-competitive.*

Proof. As shown in Fig. 3, since rules are maintained in switches once downloaded under OPT-B, its total cost can be easily calculated by $\beta l + \alpha l(\delta + \sum_{i=1}^{D-1}|E_i|)$, where the first term is the cost of the only fetching that happens in the beginning of the first valid period, and the second term is TCAM occupation cost. Combined with Lemma 1, the competitive ratio is:

$$\begin{aligned} \widehat{\lambda}_B &= \frac{C_{\text{EMF}}}{\beta l + \alpha l(\delta + \sum_{i=1}^{D-1}|E_i|)} \leq \frac{\beta D + \alpha\delta}{\beta + \alpha(\delta + D - 1)} \\ &= \frac{D + \gamma\delta}{1 + \gamma(\delta + D - 1)}. \end{aligned}$$

□

Lemma 4. *When the optimal solution belongs to OPT-C, there exists at least one valid period $V_i, i \in [1, D]$ and one empty period $E_j, j \in [1, D-1]$, such that $\frac{|V_i|-1}{|V_i|} < \gamma < \frac{1}{|E_j|}$.*

Proof. This lemma can be easily proved following similar analysis in Lemmas 2 and 3. □

Theorem 4. *When the optimal solution belongs to OPT-C, the EMF algorithm is $(\frac{D+\gamma\delta}{D+\gamma(\delta-D+2)})$-competitive.*

Proof. Without loss of generality, we suppose CIE is adopted in empty periods $\{E_1, E_2, \ldots, E_x\}$, and SUF is adopted in valid periods $\{V_1, V_2, \ldots, V_y\}$. There are z NF actions in the optimal solution. The total cost of optimal solution belongs to OPT-C can be calculated by:

$$\begin{aligned} C_{OPT-C} &= \beta l\left[D - x + \sum_{i=1}^{y}(|V_i| - 1)\right] \\ &\quad + \alpha l\left(\delta + \sum_{j=1}^{x}|E_j| - \sum_{i=1}^{y}|V_i|\right) - \alpha l z \\ &= \beta l D + \alpha l\delta + h(x, y, z), \end{aligned}$$

where

$$\begin{aligned} h(x, y, z) &= \alpha l\sum_{j=1}^{x}|E_j| - \beta l x \\ &\quad + \left[\beta l\sum_{i=1}^{y}(|V_i| - 1) - \alpha l\sum_{i=1}^{y}|V_i|\right] - \alpha l z. \end{aligned}$$

Referring to **Lemma 4**, $\gamma < \frac{1}{|E_j|} \Rightarrow \alpha - \beta < 0$ and $\frac{|V_i|-1}{|V_i|} < \gamma \Rightarrow \beta l\sum_{i=1}^{y}(|V_i| - 1) - \alpha l\sum_{i=1}^{y}|V_i| < 0$. Therefore, $h(x, y, z) < 0$ and the ratio $\lambda_C > 1$.

Obviously, we have $x \geq 1$, $y \geq 1$ and $z \geq 0$. We then consider two extreme cases. In the first case, there exists multiple CIE actions but only one SUF action. Since there must be one empty period between these two types of patterns, CIE actions cover at most $D-2$ empty periods, i.e., $x \leq D-2$. In the second case, there are only one CIE action and multiple SUF actions. The only one CIE occupies at least two valid periods. As a result, SUF actions use at most $D-2$ valid periods, i.e., $y \leq D-2$. Furthermore, the constitution of y SUF actions occupy y valid periods, and the other $D-y$ ones are left to CIE and NF actions, which cover x empty periods and z tiny valid periods, respectively. If there is only one CIE pattern, in which all the $D-y$ valid periods are crossed with x empty periods, we have $x_{max} + 1 + z = D - y$, i.e., $x + y + z \leq D - 1$. Finally, we obtain a feasible region of $h(x, y, z)$, which is denoted by $\Lambda = \{(x, y, z) | 1 \leq x \leq D-2, 1 \leq y \leq D-2, z \geq 0, x + y + z \leq D - 1\}$, where x, y and z are all integers. The lower bound of $h(x, y, z)$ is derived as follows.

$$\begin{aligned} h(x, y, z) &= \alpha l\sum_{j=1}^{x}|E_j| - \beta l x + (\beta - \alpha)l\sum_{i=1}^{y}|V_i| \\ &\quad - \beta l y - \alpha l z, \\ &\geq (\alpha - \beta)l x + (\beta - \alpha)l \cdot 2y - \beta l y - \alpha l z, \\ &= (\alpha - \beta)l x + (\beta - 2\alpha)l y - \alpha l z, \\ &\geq (\alpha - \beta)l + (\beta - 2\alpha)l - \alpha l(D - 3), \\ &= \alpha l(2 - D). \end{aligned}$$

Therefore, the competitive ratio can be expressed by:

$$\lambda_C \leq \frac{\beta D + \alpha\delta}{\beta D + \alpha\delta + \alpha(2 - D)} = \frac{D + \gamma\delta}{D + \gamma(\delta - D + 2)}.$$

□

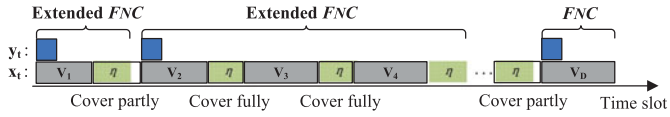


Fig. 4. An example of ECA solution.

5.3 Online Extra η Time-Slot Caching Algorithm

Our proposed EMF algorithm attempts to minimize the TCAM occupation cost by caching flow rules only when there are network traffic passing through switches. However, it would incur frequent remote processing at the controller under burst packet transmissions. In this subsection, we study to further reduce total cost by proposing the ECA (Extra Cache Algorithm (ECA) algorithm that specifies an expiration time for cached rules. As shown in Algorithm 3, we specify a parameter η as input. In each time slot, if we decide to conduct fetching action, the expiration time of fetched rules, which is denoted by `idle_timeout`, is set to ηl. We show how to set the value of η to achieve the closest performance with optimal solution by empirical analysis in next section.

5.3.1 General Cases of ECA

We let N_0 denote the number of empty periods whose length is less than η. By representing the total number of empty time slots covered by η with L, we have the following theorem.

Algorithm 3. Online Extra Cache Algorithm (ECA)

Input: η
1: **for** each time slot $t \in [1, T]$ **do**
2: **if** fetch action happens in slot t **then**
3: `idle_timeout` $\leftarrow \eta l$ for all entries to be installed
4: **end if**
5: t++
6: **end for**

Theorem 5. *The competitive ratios of ECA over* ***OPT-A***, ***OPT-B*** *and* ***OPT-C*** *are* $\zeta_A = \frac{D-N_0+\gamma(\delta+L)}{\delta}$, $\zeta_B = \frac{D-N_0+\gamma(\delta+L)}{1+\gamma(\delta+D-1)}$ *and* $\zeta_C = \frac{D-N_0+\gamma(\delta+L)}{D+\gamma(\delta-D+2)}$, *respectively.*

Proof. As shown in Fig. 4, compared with an original optimal solution where there is no extended FnC pattern included, in the ECA solution, if the length of an empty period is longer than η, the cached rules will be removed after the expiration time, and they will be refetched at the beginning of next valid period. The total TCAM occupation cost can be calculated by $\alpha l(\delta + L)$. Otherwise, rules are cached at the switch during the empty period, leading to remote processing cost $\beta l(D - N_0)$. Therefore, the total cost of ECA with η is

$$C_{ECA(\eta)} = \beta l(D - N_0) + \alpha l(\delta + L). \tag{6}$$

Following the similar analysis in Theorems 2, 3, and 4, we can easily obtain the competitive ratios of $\zeta_A = \frac{D-N_0+\gamma(\delta+L)}{\delta}$, $\zeta_B = \frac{D-N_0+\gamma(\delta+L)}{1+\gamma(\delta+D-1)}$ and $\zeta_C = \frac{D-N_0+\gamma(\delta+L)}{D+\gamma(\delta-D+2)}$ over OPT-A, OPT-B, and OPT-C, respectively. □

5.3.2 Special Case of ECA

Suppose the length (denoted by variable X) of all the empty periods E_j $(j = 1, 2, \ldots, D-1)$ are exponentially distribution with mean value μ_e, i.e., $X \sim \exp(\frac{1}{\mu_e})$, the value of N_0 can be calculated by:

$$N_0 = (D-1) \cdot Pr(X \le \eta l) = (D-1)(1 - e^{\frac{-\eta l}{\mu_e}}). \tag{7}$$

And the total length of the completely covered empty periods by η can be written as

$$\begin{aligned} L_1 &= (D-1) \cdot E(X \le \eta l) = (D-1) \int_0^{\eta l} \left(X\left(\frac{1}{\mu_e} e^{\frac{X}{-\mu_e}}\right)\right) dX \\ &= \mu_e (D-1)[1 - \left(\frac{\eta l}{\mu_e} + 1\right) e^{\frac{-\eta l}{\mu_e}}]. \end{aligned} \tag{8}$$

On the other hand, the total length in the empty periods where partially cut by η shall be simply calculated as

$$L_2 = (D - 1 - N_0) \cdot (\eta l) = \eta l (D-1) e^{\frac{-\eta l}{\mu_e}}. \tag{9}$$

Therefore, the cost of ECA with parameter η, μ_e and l shall be

$$\begin{aligned} C_{ECA(\eta,\mu_e,l)} &= \alpha l(\delta + L_1 + L_2) + \beta l(D - N_0) \\ &= \alpha(\delta l + (D-1)\mu_e(1 - e^{\frac{-\eta l}{\mu_e}})) \\ &\quad + \beta l(1 + (1-D) e^{\frac{-\eta l}{\mu_e}}). \end{aligned} \tag{10}$$

Similarly, the competitive ratio of ECA under this special case can also be derived following the approach in the proof of Theorem 5. In performance evaluation, we conduct extensive simulations under various network settings to find out the best value of η leading to closet performance with optimal solution.

6 Evaluation

We conduct extensive simulations in this section to evaluate the performance of our proposed algorithms and the derived competitive ratios.

6.1 Simulation Settings

We adopt network traces [20], with a collection of TCP packets and ethernet frames captured in a wired hub using wireshark [28] tool. Each trace file is collected within around 50 seconds. We first process these trace files by accumulating the traffic volume of each time slot with length l. Both proposed algorithms (Greedy, EMF and ECA) and legacy algorithms (Proactive and Reactive) are implemented in our simulation. Note that, in the adopted legacy Proactive algorithm, as discussed in Section 2, rules are only fetched in the first time slot and cached all the remaining duration, resulting in a cost of $\beta l + T\alpha l$. In contrast, the Reactive algorithm triggers remote process at each time slot with a cost $(\beta + \alpha)\delta l$. To obtain the offline optimal solutions (denoted by Offline OPT), the commercial solver Gurobi optimizer [29] is used. In each suite of simulation, we always fix α=10 and the settings of other parameters are labeled below the figures.

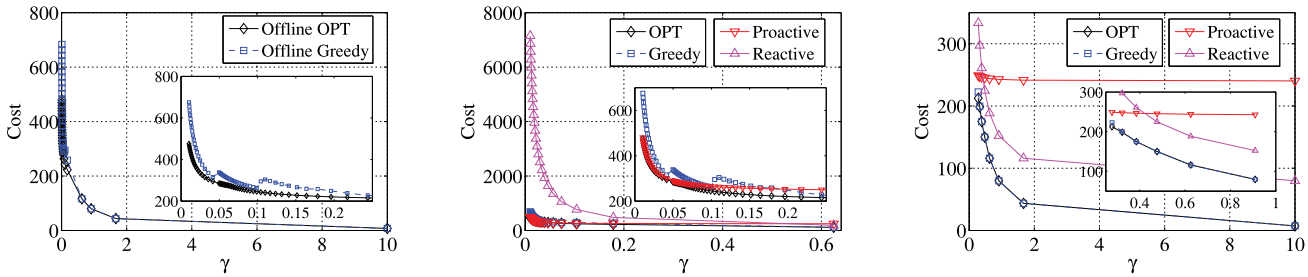


Fig. 5. Performance of offline Algorithms while varying γ.

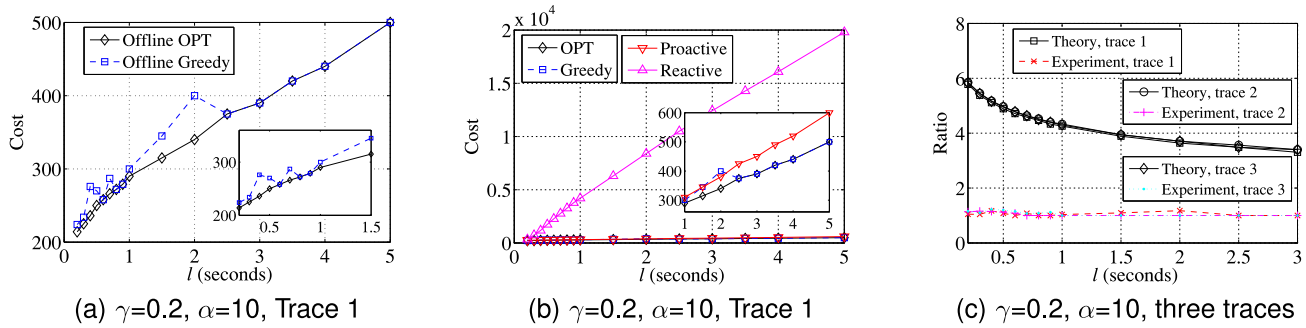


Fig. 6. Performance of offline Algorithms while varying l.

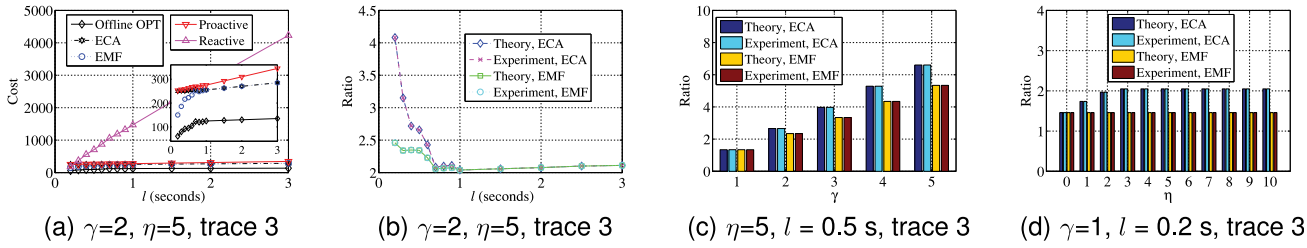


Fig. 7. Performance of EMF and ECA over OPT-A.

6.2 Evaluation of Offline Algorithm

As shown in Fig. 5, when $\alpha = 10$ and $l = 0.25$ s, total cost of all algorithms decreases as γ grows from 0.01 to 10. This is because under larger γ, remote fetching is preferred as it leads to low cost. It also can be seen from Fig. 5b that the Reactive strategy products much higher cost compared with other algorithms, because there are always remote fetching operations even if γ is small. On the other hand, the Proactive way is competitive to Greedy only when γ is very small, as shown in Fig. 5b, but it generates an even higher cost than the Reactive strategy, as it shows in Fig. 5c, once γ grows bigger than 0.4. In contrast, our proposed Greedy algorithm always performs close to the offline optimal cost. Specifically, from Fig. 5c we observe that the costs of all algorithms converge when γ is greater than 2 because they generate only fetch operations.

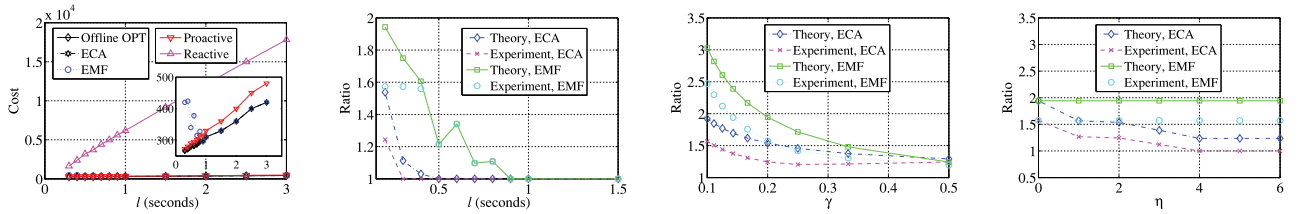
We then investigate the influence of time slot length on the total cost by changing the value of l from 0.1 to 5 s. As shown in Fig. 6a, total cost increases as the growth of l under optimal solutions. We observe that some fluctuation in the performance of our greedy algorithm. That is because in each iteration of while loop in Algorithm 1, we prefer tiny valid periods, which may have some time slots already contained. From Fig. 6b, it also can be seen that Greedy outperforms the legacy Proactive and Reactive strategies. The performance ratio of our algorithm and the optimal solution is shown in Fig. 6c, where the ratio is very close to 1, much better than the analytical upper bound.

6.3 Evaluation of Online EMF and ECA

Then, the online EMF and ECA algorithms are evaluated with real network traffic traces under three cases, respectively. Note that, the legacy Proactive and Reactive strategies are only shown in the group of simulations where l varies. Similar results are obtained by varying parameters γ and η and thus omitted.

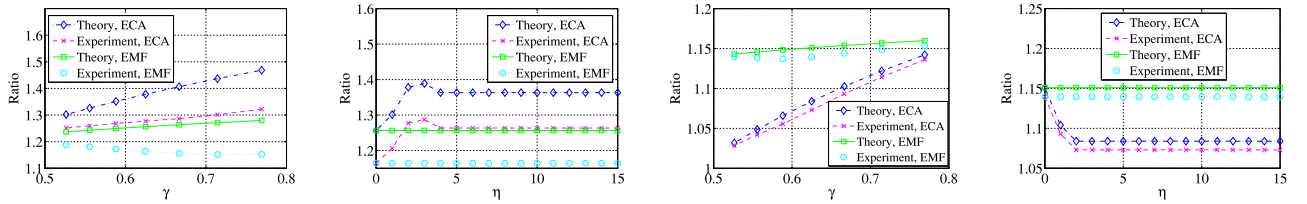
6.3.1 *Over OPT-A*

The total cost under different values of l is shown in Fig. 7a, where results of all algorithms show as increasing functions of l. On the other hand, the performance ratio of EMF and ECA with the optimal solutions decreases as the growth of l as shown in Fig. 7b. Furthermore, we observe that the performance ratio are always equal to the derived upper bound. That is because only SuF and NF patterns exist in the optimal solutions under this case, i.e., packets are always processed by the remote controller. Finally, we show the performance ratio under different value of η in Fig. 7d. We observe that ratio of ECA is same with EMF



(a) γ=0.2, η=2, trace 2 (b) γ=0.2, η=2, trace 2 (c) η=2, l = 0.2 s, trace 2 (d) γ=0.2, l = 0.2 s, trace 2

Fig. 8. Performance of EMF and ECA over OPT-B.



(a) η=2,l=0.2 s,trace 2 (b) γ=0.625,l=0.2 s,trace 2 (c) η=2,l=0.4 s,trace 2 (d) γ=0.625,l=0.4 s,trace 2

Fig. 9. Performance of EMF and ECA over OPT-C while varying γ and η.

when $\eta = 0$, and the ratios of EMF never change because of fixed γ and l. In Figs. 7b, 7c, and 7d, EMF outperforms ECA with closer performance to the optimal solution under most of settings. We attribute this phenomenon to the fact that rules are cached at switches for a longer time under ECA.

6.3.2 Over OPT-B

In this case, packets tend to be processed at local switch because γ becomes small. In Figs. 8a and 8b, we have similar observations with the results under OPT-A case, except ECA outperforms EMF. This can be attributed to that there are mainly CIE patterns and few NF patterns under OPT-B case, and the extra caching durations of ECA cover many empty periods. Accordingly, the cost of ECA is smaller than EMF, particularly when γ and η become large. Finally, both algorithms converge to the optimal solution under larger value of l. In respect of performance ratio, Fig. 8c shows ratio is decreasing function of γ for both ECA and EMF. Because larger γ leads to more short FCN patterns in optimal solution, which makes ECA and EMF close to optimal solution. Therefore, their ratios decrease and approach to 1 gradually. As we mentioned above, Fig. 8d shows the benefit of a larger η in ECA, because more empty periods are covered and much fetching cost can be saved under OPT-B case.

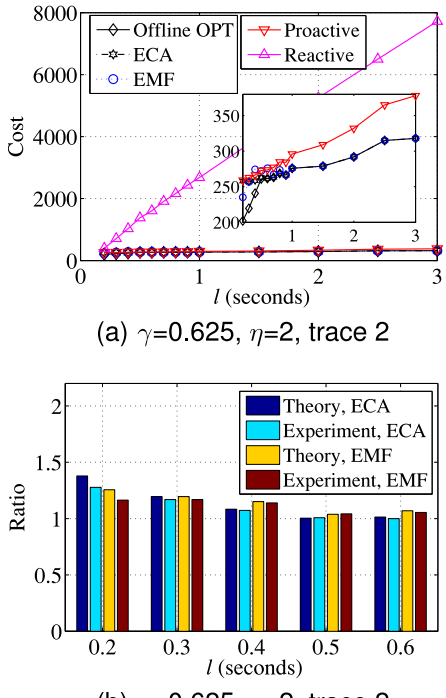
6.3.3 Over OPT-C

The performance of ECA and EMF is investigated in Figs. 9 and 10. We observe that ECA and EMF show distinct performance under different settings. For example, in Fig. 10a, the cost of ECA is larger than EMF when $l = 0.2$, but it becomes opposite when l is greater than 0.3. We have similar observations in other figures. Interestingly, in Fig. 9b, we observe that the curve of ECA first increases, and then decreases when η is greater than 3, finally converging to 1.28 after $\eta = 4$. This is because when η is small, it only covers few empty periods with short length, and the advantage of extra caching is not obvious. As η becomes larger, the number of covered empty periods grows, leading to reduced total cost. When all empty periods are covered by large η, the performance of ECA becomes stable. In Fig. 9d, the ratios of ECA keep decreasing and then converging because short empty periods become fewer when l increases to 0.4.

Additionally, in Figs. 7a, 8a and 10a, we can always observe that the Reactive approach creates extremely high total cost and the proposed EMF and ECA perform better than both Proactive and Reactive strategies. This also validates the efficiency of the proposed online algorithms.

6.4 Evaluation of Special Case of ECA

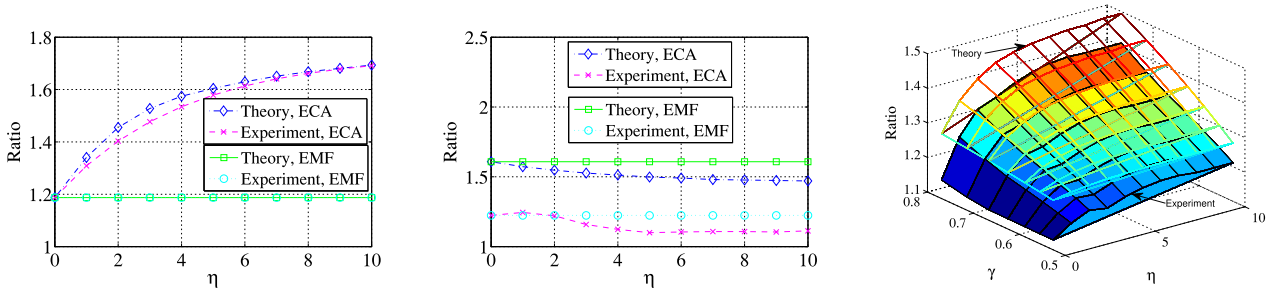
Finally, we study the performance of ECA when lengths of empty periods are exponential distribution. We consider randomly generated network traffic with $\mu_e = 1.0$ and $T = 100$.



(a) γ=0.625, η=2, trace 2

(b) γ=0.625, η=2, trace 2

Fig. 10. Performance of EMF and ECA over OPT-C while varying l.



(a) ECA over OPT-A under special case, μ_e=1.0, γ=1, l=0.25.

(b) ECA over OPT-B under special case, μ_e=1.0, γ=0.2, l=0.25.

(c) ECA over OPT-C under special case, μ_e=1.0, l=0.5.

Fig. 11. Performance of ECA over OPT-A, OPT-B and OPT-C under special case.

As shown in Fig. 11a, ratio of ECA algorithm increases as η grows from 0 to 10, which shows the same performance with Fig. 7d. In Fig. 11b, we have similar observation with Fig. 9b because of the same reasons. In Fig. 11c, we set $l = 0.5$, and performance ratios are always below the theoretical bound as γ and η changes within [0.5, 0.8] and [0, 10], respectively. The simulation results also suggest that η shall be set to small values no matter how γ changes.

7 Conclusion

In this paper, we studied traffic flow provisioning problem by formulating it as a minimum weighted flow provisioning problem with objective of minimizing the total cost of TCAM occupation and remote packet processing. An efficient heuristic algorithm is proposed to solve this problem when network traffic is given. We further propose two online algorithms to approximate the optimal solution when network traffic information is unknown in advance. Finally, extensive simulations were conducted to validate the performance of theoretical analysis of the proposed algorithms, using the real traffic traces.

References

[1] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "Openflow: Enabling innovation in campus networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, 2008.

[2] M. Casado, M. Freedman, J. Pettit, J. Luo, N. Gude, N. McKeown, and S. Shenker, "Rethinking enterprise network control," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1270–1283, Aug. 2009.

[3] N. Kang, Z. Liu, J. Rexford, and D. Walker, "Optimizing the one big switch abstraction in software-defined networks," *Proc. ACM 9th ACM Conf. Emerging Netw. Experiments Technol.*, 2013, pp. 13–24.

[4] T. Benson, A. Anand, A. Akella, and M. Zhang, "Microte: Fine grained traffic engineering for data centers," in *Proc. 7th Conf. Emerging Netw. Experiments Technol.*, 2011, pp. 1–12.

[5] S. Agarwal, M. Kodialam, and T. Lakshman, "Traffic engineering in software defined networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, Apr. 2013, pp. 2211–2219.

[6] Y. Kanizo, D. Hay, and I. Keslassy, "Palette: Distributing tables in software-defined networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2013, pp. 545–549.

[7] A. Ishimori, F. Farias, E. Cerqueira, and A. Abelém, "Control of multiple packet schedulers for improving QoS on openflow/SDN networking," in *Proc. 2nd Eur. Workshop Softw. Defined Netw.*, 2013, pp. 81–86.

[8] H. Huang, P. Li, S. Guo, and B. Ye, "The joint optimization of rules allocation and traffic engineering in software defined network," in *Proc. IEEE 22nd Int. Symp. Quality Service*, May 2014, pp. 141–146.

[9] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, "Nox: towards an operating system for networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 3, pp. 105–110, 2008.

[10] S. Shin and G. Gu, "Attacking software-defined networks: A first feasibility study," in *Proc. 2nd ACM SIGCOMM Workshop Hot Topics Softw. Defined Netw.*, 2013, pp. 165–166.

[11] K. Giotis, C. Argyropoulos, G. Androulidakis, D. Kalogeras, and V. Maglaris, "Combining openflow and sflow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments," *Comput. Netw.*, vol. 62, pp. 122–136, 2014.

[12] U. C. Kozat, G. Liang, and K. Kokten, "On diagnosis of forwarding plane via static forwarding rules in software defined networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2014, pp. 1716–1724.

[13] S. Sharma, D. Staessens, D. Colle, M. Pickavet, and P. Demeester, "Fast failure recovery for in-band openflow networks," in *Proc. 9th Int. Conf. Des. Reliable Commun. Netw.*, 2013, pp. 52–59.

[14] N. Handigol, B. Heller, V. Jeyakumar, D. Mazieres, and N. McKeown, "I know what your packet did last hop: Using packet histories to troubleshoot networks," in *Proc. 11th USENIX Symp. Netw. Syst. Des. Implementation*, 2014.

[15] M. Borokhovich, L. Schiff, and S. Schmid, "Provable data plane connectivity with local fast failover: Introducing openflow graph algorithms," in *Proc. 3rd Workshop Hot Topics Softw. Defined Netw.*, 2014, pp. 121–126.

[16] K. Pagiamtzis and A. Sheikholeslami, "Content-addressable memory (CAM) circuits and architectures: A tutorial and survey," *IEEE J. Solid-State Circuits*, vol. 41, no. 3, pp. 712–727, Mar. 2006.

[17] Y. Sun and M. S. Kim, "Tree-based minimization of tcam entries for packet classification," in *Proc. 7th IEEE Consumer Commun. Netw. Conf.*, 2010, pp. 1–5.

[18] N. Katta, J. Rexford, and D. Walker, "Infinite cacheflow in software-defined networks," Dept. Comput. Sci., Princeton University, Princeton, NJ, USA, Tech. Rep. TR-966-13, 2013.

[19] M. F. Bari, A. R. Roy, S. R. Chowdhury, Q. Zhang, M. F. Zhani, R. Ahmed, and R. Boutaba, "Dynamic controller provisioning in software defined networks," in *Proc. 9th Int. Conf. Netw. Service Manag.*, 2013, pp. 18–25.

[20] S. Yang, J. Kurose, and B. N. Levine, "Disambiguation of residential wired and wireless access in a forensic setting," in *IEEE Int. Conf. Comput. Commun.*, 2013, pp. 360–364.

[21] E. Spitznagel, D. Taylor, and J. Turner, "Packet classification using extended tcams," in *Proc. 11th IEEE Int. Conf. Netw. Protocols*, 2003, pp. 120–131.

[22] Openflow switch specification [Online]. Available: www.opennetworking.org/sdn-resources/openflow, 2015.

[23] M. Yu, J. Rexford, M. J. Freedman, and J. Wang, "Scalable flow-based networking with DIFANE," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 351–362, 2010.

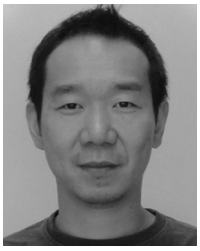
[24] S. Sharma, D. Staessens, D. Colle, M. Pickavet, and P. Demeester, "Openflow: Meeting carrier-grade recovery requirements," *Comput. Commun.*, vol. 36, no. 6, pp. 656–665, 2013.

[25] F. Dürr, "Towards cloud-assisted software-defined networking," Institute of Parallel and Distributed Systems, Universität Stuttgart, Stuttgart, Germany, Tech. Rep. 2012/04, 2012.

[26] X. N. Nguyen, D. Saucez, C. Barakat, and T. Thierry, "Optimizing rules placement in openflow networks: Trading routing for better efficiency," in *Proc. ACM SIGCOMM Workshop Hot Topics Softw. Defined Netw.*, 2014, pp. 127–132.

[27] R. Cohen, L. Lewin-Eytan, J. S. Naor, and D. Raz, "On the effect of forwarding table size on SDN network utilization," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2014, pp. 1734–1742.
[28] Wireshark [Online]. Available: www.wireshark.org, 2015.
[29] G. Optimization, "Gurobi optimizer reference manual," 2013.

Huawei Huang received the master's degree in computer science from the China University of Geoscience (Wuhan) in 2013. He is currently working toward the PhD degree in the School of Computer Science and Engineering, The University of Aizu, Japan. His research interests is mainly in the area of software-defined networking and wireless networks. He is a student member of the IEEE.



Song Guo (M'02-SM'11) received the PhD degree in computer science from the University of Ottawa, Canada, in 2005. He is currently a full professor in the School of Computer Science and Engineering, the University of Aizu, Japan. His research interests are mainly in the areas of protocol design and performance analysis for wireless networks and distributed systems. He has published more than 250 papers in refereed journals and conferences in these areas and received three IEEE/ACM best paper awards. He currently serves as an associate editor of the *IEEE Transactions on Parallel and Distributed Systems*, an associate editor of the *IEEE Transactions on Emerging Topics in Computing for the track of Computational Networks*, and on editorial boards of many others. He has also been in organizing and technical committees of numerous international conferences. He is a senior member of the IEEE and the ACM.

Peng Li received the BS degree from Huazhong University of Science and Technology, China, in 2007, and the MS and PhD degrees from the University of Aizu, Japan, in 2009 and 2012, respectively. He is currently an associate professor in the University of Aizu, Japan. His research interests include networking modeling, cross-layer optimization, network coding, cooperative communications, cloud computing, smart grid, performance evaluation of wireless and mobile networks for reliable, energy-efficient, and cost-effective communications. He is a member of the IEEE.

Weifa Liang (M'99-SM'01) received the the BSc degree from Wuhan University, China, in 1984, the ME degree from the University of Science and Technology of China in 1989, and the PhD degree from the Australian National University in 1998, all in computer science. He is currently an associate professor in the Research School of Computer Science, Australian National University. His research interests include design and analysis of energy-efficient routing protocols for wireless ad hoc and sensor networks, cloud computing, graph databases, design and analysis of parallel and distributed algorithms, approximation algorithms, combinatorial optimization, and graph theory. He is a senior member of the IEEE.

Albert Y. Zomaya is the chair professor of high-performance computing & networking and australian research council professorial fellow in the School of Information Technologies, Sydney University. He is also the director in the Centre for Distributed and High Performance Computing, which was established in late 2009. He published more than 500 scientific papers and articles and is author, co-author or editor of more than 20 books. He served as the editor in chief of the *IEEE Transactions on Computers* (2011-2014) and currently serves as an editor in chief of *Springers Scalable Computing*. He also serves as an associate editor for 22 leading journals and is the founding editor of the *Wiley Book Series on Parallel and Distributed Computing*. He was the chair the IEEE Technical Committee on Parallel Processing (1999-2003) and currently serves on its executive committee. He is the Vice-Chair, IEEE Task Force on Computational Intelligence for Cloud Computing and serves on the advisory board of the IEEE Technical Committee on Scalable Computing and the steering committee of the IEEE Technical Area in Green Computing. He has delivered more than 150 keynote addresses, invited seminars, and media briefings and has been actively involved, in a variety of capacities, in the organization of more than 600 conferences. He received the IEEE Technical Committee on Parallel Processing Outstanding Service Award in 2011, the IEEE Technical Committee on Scalable Computing Medal for Excellence in Scalable Computing in 2011, and the IEEE Computer Society Technical Achievement Award in 2014. His research interests are in the areas of parallel and distributed computing and complex systems. He is a chartered engineer, a fellow of the AAAS, IEEE, and IET (United Kingdom).