Received September 24, 2017, accepted October 20, 2017, date of publication November 10, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2768402

An Identification Decision Tree Learning Model for Self-Management in Virtual Radio Access Network: IDTLM

SAI ZOU[1], (Student Member, IEEE), YULIANG TANG[1], (Member, IEEE), YANGLONG SUN[1], AND KE SU[2]

[1]Department of Communication Engineering, Xiamen University, Xiamen 361005, China
[2]Chongqing College of Electronic Engineering, Chongqing 401331, China

Corresponding author: Yuliang Tang (yuliang.tang@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 91638204, Grant 61731012, and Grant 61572174, in part by the Science and Technology Plan Project of Hunan Province under Grant 2016TP1020, and in part by the Chongqing Education Science Project of China under Grant kj1503011.

ABSTRACT Along with the blowout of new applications and the integration of the heterogeneous networks platforms in the future Internet of everything, the self-management of virtual radio access network is of significant importance. The urgent problem needed to be solved for the self-management in virtual radio access network is the match of application and virtual service (tailored service of virtual functions for the application). In this paper, an identification decision tree learning model (IDTLM) based on transfer learning has been proposed. First, we do research on the redundant problem of the traditional packet decision trees and reduce the dimensionality of features by a proximal gradient descent method and clustering the features by Lagrange's multiplier, so as to improve the online matching speed between applications and virtual service. And then, in consideration of the independent and non-identical distribution among online and trained data, and the possible change of virtualized network platform, a method of transfer learning is proposed to improve the quality of generalization for IDTLM. Finally, online test is done for IDTLM, and the result shows that the accuracy rate of trained applications can reach 99% and the accuracy rate can reach 96% if untrained applications are included. Meanwhile, theoretical analysis has been carried out for the transfer of IDTLM. The analysis shows that IDTLM system is of high recognizing speed and low false positive rate and it could adapt to the transfer of different scenarios.

INDEX TERMS Identification decision tree, virtual radio access network, machine learning, transfer learning, self-management.

I. INTRODUCTION

The future of internet is the internet of everything: person, data, things and procedures will be interrelated by mobile networks to construe an integral whole to provide an unprecedent opportunity and renovation for human beings [1]. Different network systems in the network of thing will be integrated by virtual technology to make a mobile platform with various applications to achieve the leaps and bounds of operation via [2], [3]. ITU-R defined 3 types of classic application scenarios in the June of 2015 [4] and IMT-2020 in China announced four scenarios of 5G in the May of 2015 [5]. Both of the two proved the previous points. Even so, it costs a lot to use manpower to deploy, optimize, repair and redeploy and it is delayed. Therefore, to reduce the cost of operation and maintenance and improve the optimized efficiency of network and operation quality, it is of great significance to do self-management of virtual radio access network [6], [7]. The urgent problem needed to be solved is to provide tailored functional service for the self-management of virtual radio access network, and meanwhile the operation via could be used to the maximum and the applications could gain the optimum QoS. Based on the 5G vision of being green, soft and fast, China Mobile proposed the concept of SDAI (Software Defined Air Interface), which is tailored for radio access system [8].

'Tailoring' is the logic mapping of application functions, which is the matching process between application logics and virtual service via [9]. That is to say, to preselect one group satisfying the application requirements from the function set of virtual network to do virtual service. The primary problem of 'tailoring' is the matching between application and virtual service to achieve the goal that the different requirement form applications could gain different service. This matching process is similar to the packet classification, which acts as the initial filter to the network in which network traffic is classified into flows based on a pre-defined set of rules [10]. Hence, the matching between applications and virtual service is a core concern for virtualization network services.

The traditional packet classification is to select the highest priority policy that matches all fields with a given packet header among a rule set whose rules consist of multiple fields; the packet is then processed according to the policy [11]. For high performance, these virtual switches must sustain rates of at least100 Gbps. So, it requires efficient packet classification algorithms. In [12], software-based packet has shown that wire speed can be achieved with optimization in the data path. However, Varvello *et al.* [13], pointed out that this method was not a true packet classification. In fact, the recent emergence of software defined network (SDN) and OpenFlow, with its large rule tables and long multi-dimensional tuples, have been imposing unforeseen challenges to current packet classifiers [14]. As a result, new directions to accelerate software-based switching are required to support the ever-increasing throughput demands.

The research on software-based packet classification algorithm can be divided into space division and non-partitioning division [15]. Space division is to divide original search space into many sub spaces, in which the network packet in each sub space has the unique subset to match the sub space. Thus the network packets can be divided as long as the sub space is determined. With the growing regular sets, the demand for the algorithm of space division is on increasing growth. Non-partitioning algorithm divides regular sets into many regular subsets based on certain heuristic pheromone, and every regular subset will be searched one by one or in parallel operation. Non-partitioning algorithm efficiently reduce the demand for storage space of current algorithm but increase the time for search. In [16], based on the previous two methods, Wooguil Pak and Young-June Cho proposed high performance and high scalable packet classification algorithm for supporting a large rule set and fast classification. Moreover, the performance can be easily improved. Meanwhile, the traditional packet classification usually only considers an IP packet that has five tuples of protocol types, source address, source port, destination address and destination port. The features of virtual radio access network are as follows:

1) Under the premise of the applications's QoS and the efficient usage of infrastructure, the virtual switch examines all signaling and partitions traffic into sub-streams. That means the relevance of virtual functions should be considered when the features of signaling are chosen. Meanwhile, it is needed to move switches and control software both of which are in change of open virtual access platforms. An Open Flow Switch consisting of one or more flow tables and a group table performs signaling lookups and forwarding via a secure Open Flow channel to an external controller. The switch communicates with the controller and the controller manages the switch via the Open Flow protocol. Based on the Open Flow protocol, the controller can add, update, and delete flow entries in flow tables, both reactively (in response to packets) and proactively. Each flow table in the switch contains a set of flow entries and each flow entry consists of match fields, counters, and a set of instructions to apply to the matching signaling.

2) The form of Packet Data Protocol (PDP) to define context is "$+CGDCONT = [< cid > [, < PDP\_type > [, < APN > [, < PDP\_addr > [, < d\_comp > [, < h\_comp > [, < pd_1 > [, < pd_n >]]]]]]]$ "; the form of service application quality is "$+CGQREQ = [< cid > [, < precedence > [, < delay > [, < reliability > [, < peak > [, < mean >]]]]]]$", which is different from the route packet form of core networks.

3) It develops an open-source architecture; therefore the application distribution density or infrastructure of virtualized network platform might change with time, which determines the transfer quality from offline training to online predicting. It is hard to get the training data of new emerged fields thus it is hard to train data comprehensively. Therefore, the change of infrastructure might result in the differences between the features of online predicting and those of offline training. Meanwhile, the applications of predicted data and trained data might have different distribution. All of the above might result in the expiration of trained data.

To overcome this problem, we do research in the radio access virtual environment. Radio access virtual system is the combination of a virtual data platform of central management and a group of heterogeneous radio access nodes. After finishing its function tasks, every heterogeneous radio node abstract redundant service ability into virtual network functions (VNFs) by virtualized technology. In consideration of the load of applications in the area, network management experts arrange each virtual element into virtual service to tailor for the application. When a virtual platform receives a signaling from a mobile terminal, it could automatically deploy relevant virtual service to serve it. And signaling and virtual service should be matched in this process. As it is time consuming and consumes lots of cache to calculate and recognize data online, it is necessary to establish an offline learning model to accelerate the speed of online recognization and reduce cache. Meanwhile the data distribution differs from area to area; thus the transfer quality of learning model should be studied. First we take the method of machine learning to calculate a huge amount of data offline by the features of VNFs. And based on the trained data, signaling identification model whose feature is VNFs abstracted from infrastructure is established to take up less memory than the traditional packet algorithm and to group packets more efficiently than the traditional one. Finally, transfer learning is used to solve the problem of the environment change online and offline(the dynamic change problem of virtual network function set and application distribution). Our contributions are as follows:

1) We proposed an identification decision tree learning model (IDTLM) in virtual radio access heterogeneous network. This model tries to solve the problem that there exist redundant features and rules in traditional decision trees. First the features of decision trees are chosen to reduce dimension disaster due to redundant features; and then the value of each feature is clustered to reduce the number of rule for packet division to improve search speed.

2) We improved packet classification decision-tree algorithm [17], [28], [40]. Firstly, the learning task is not to generate taxonomic model upon fully trained data. Meanwhile trained data and tested dada might not have the same distribution. IDTLM can use transfer learning to finish the recognizing task in new environment by the knowledge learnt from local environment.

3) We identify classifier properties that guarantee zero cost for additional fields represented by ranges or prefixes transparently to the implementation scheme.

4) We demonstrate the viability of our approach through simulations with real parameters which shows that the accuracy rate of trained applications can reach 99% and if untrained applications are included the accuracy rate can reach 96%. To test the transfer quality of the models, we expand the data set of ClassBench [50] to 100,000, each of which holds 100 fields. The results show that DTLM system is of high recognization speed, low erroneous speed and could adapt to the transfer of different scenarios.

The remainder of the paper is organized as follows. Section II is dedicated to the related work of radio access network virtualization and packet classification. Then, we give a system model in Section III. In sections IV, we present the technical details of solution method for model. Next, in Section V, experimental and numerical simulation are implemented. Then, we conclude our work in Section VI. Finally, we have carried on the proof to the related formula in Section VII.

II. RELEVENT BACKGROUND

A. BACKGROUND OF RADIO ACCESS NETWORK VIRTUALIZATION

Along with the development of information technology, needs of users and business format differ bigger and bigger, and thus traditional radio access network can not satisfy the needs. Researchers have tried to take the method of SDN and NFV to optimize the structure of radio access network. Multiple concurrent virtual network will be supported by abstract, refactoring, isolation mechanism in the shared network infrastructure [18]. Each virtual network can design and construct protocols upon its own resources as if it enjoys its own physical network. It could flexibly dispatch the resources in the whole network in consideration of the dynamic change of network environment so as to make the resources of the whole network under reasonable control. Virtual network could improve the utility rate of network resources, service quality and operation efficiency so as to reduce the cost of network operation and maintenance [19]. Network function virtualization first abstracts the traditional network hardware equipment and makes network function independent of hardware [20], [21]. Stanford University proposed SDN [22] to have the network under central control by separating the signalling and data in computer network. SDN makes network maintenances simpler and the network utility can be improved. Meanwhile, SDN could have separate network under unified control to shield the details of network bottom to realize the connection and integration among every virtual network function.

Li *et al.* [23] from Bell lab discussed the advantages of introducing software definition into radio access network and provided a general framework and the definitions of some function requirements. Kempf *et al.* [24] from Ericsson provided preliminary scheme to bring software defined network into radio core network. Based on that, Huawei proposed Mobile-Flow structure to redivide the functions of radio core network [25]. Meanwhile, Mobile-Flow structure completely separates data forwarding form control and add tunnel mechanism to switch equipment. The maintenance of tunneling could provide reliability assurance and mobile management. The flexibility of central control make it possible to realize the network integration and resource sharing of radio core network. Therefore, China Mobile proposed C-RAN structure which made use of centralized baseband pool to realize the central control of radio stations [26]. This structure is logically centralized and physically dispersed. It separates centralized base station equipment from radio frequency units and exchanges and connects through high radio frequency. Remote radio frequency radio provides network of high capacity and wild coverage which is of high broadwidth and low delay. Baseband pool consists of high quality processor which provides ability to process for every virtual base station. As a matter of fact, virtualization of mobile base station has been used as a relevant case for NFV in [27]. Zou *et al.* [9] proposed an open mapping framework which is independent of application and infrastructure. This framework employs reconfigurable computing from the application level (OSI Layer 7) to transport layer (OSI Layer 4), as well as adopts associated development tools that allow networking domain experts to easily customize the system,as shown in Fig. 1.

B. PACKET CLASSIFICATION

Packet classification is an essential function in Internet that provides value-added services. In [28], a compliance checking algorithms of the service work-flow has been proposed for dynamically gathered to perform functions in the Internet-of-Things based on the progressive work on Service Workflow Specification language. It optimizes the combination of Petri net nodes, reduces the overhead of time complexity, and accomplishes tasks within a polynomial time. However, it is functional in a fixed environment (Petri net nodes) and fixed work-flow rules (task). In [42], a novel open source monitoring architecture has been presented for

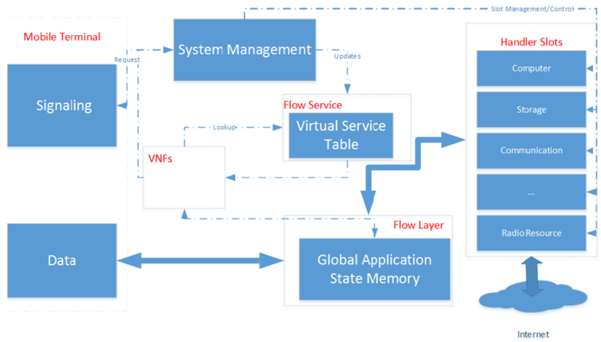


FIGURE 1. RSCMF platform architecture overview.

multi-gigabit (i.e., 10 Gbps+) network traffic streams. It is based on PF-RING Zero-Copy and tailored for deployment on commodity hardware for troubleshooting high-impact events that may arise from malicious actions such as DDoS attacks. The framework allows for further statistical, filtering and visualization modules, but isn't an interactive filtering mode of operation that don't enable network operators to zoom-in and examine IP ranges of interest in real-time. Owing to high flexibility in satisfying various industrial requirements for security and network systems, Wooguil-Pak and Young-June. Pak and Choi [16] present a software-based packet classification algorithm. The algorithm simultaneously supports high scalability and fast classification performance by merging partition decision trees in a search table. In [29], a reconfigurable platform and programming tools for high level network applications has been presented. It allows the rapid deployment of hardware-accelerated attack resilient interactive communication applications. As a configurable option, the platform can be provided with a unidirectional connection to the management interface. This would allow the sharing of the Core between Handlers and system management components for sending IP-based control messages over the private management interface.

Both hardware and software based solutions have been proposed to improve the performance of packet classification. Hardware proposals usually use parallel lookups to accelerate the search procedure, but hardware resources may limit the size of supported filter sets. In [13], with SDN large rule tables and long multi-dimensional tuples,an accelerating packet classification using the high parallelism and latency-hiding capabilities of graphic processing units (GPUs) has been presented. It has implement GPU-accelerated versions for both linear and tuple search. In [30], a scalable multi-match packet classification algorithm with TCAM and SRAM has been presented. Its scheme synthesizes TCAM compatible entries by using binary decision trees and employs SRAM for further comparisons. Each synthesized entry can be stored in one TCAM entry to significantly reduce TCAM consumption and fullfill low power consumption. The ternary content addressable memory is very expensive, has limited capacity, consumes large amounts of power, and generates tremendous amounts of heat. Liu *et al.* [11] proposed a packet classification scheme that used binary content addressable memory. It breaks a multi-dimensional lookup into a series of 1-D lookups, and converts the ternary matching problem into a binary string exact matching problem for each 1-D lookup. Qu and Prasanna [31] presented a two-dimensional pipelined architecture for packet classification on Field Programmable Gate Arrays (FPGA); this architecture achieves high throughput while supporting dynamic updates. In this architecture, modular Processing Elements (PEs) are arranged in a two-dimensional array. Each PE accesses its designated memory locally, and supports prefix match and exact match efficiently. Hardware-based solutions, such as packet classification accelerator chips or TCAM or FPGA, have been widely adopted for high-performance systems. Hence, the existing hardware platform can support the multi-domain signal classification of virtualization radio self-management system.

Software algorithms have better scalability; however, their data structures for storing filters significantly affect the performance. The schemes have been put forward by Finite State Automata (FSA) [32]–[35] or Regular Expression Matching (REM) [36], [37] to achieve fast matching speed in the scenarios where only limited 7 on-chip memory resources are available. Yu *et al.* [38] indicate that FSA have a limited size but can require expensive per-character processing (or high memory bandwidth requirements), whereas DFA offers limited per-character processing at the cost of a possibly large automaton. From an implementation perspective, REM can be classified into two forms: Memory-based solutions can be deployed on various platforms (general purpose processors, network processors, ASIC, FPGA, GPU) and have been recently extended to TCAM implementations; logic-based designs typically target FPGAs. They mainly solve the exact match. However, in the virtual wireless access network self-management need only fuzzy matching.

Chang and Chung [46] present a TCAM-based scheme for multi-match packet classification without single-match penalty. The scheme partitions a rule set based on range layering, which can be applied to achieve range encoding. However, it may incur extra memory. In [10], a range enhanced packet classification design on FPGA has been present based on OpenFlow1.0 (consisting of 12-tuple header fields). It uses specially designed codes to store the pre-computed results in memory, and uses a simple subrange comparison method to find the matching result in a sequential fashion to process the range fields.

Decision-tree-based packet classification algorithms show excellent search performance by exploiting the geometrical representation of rules in a classifier and searching for a geometric subspace to which each input packet belongs. However, decision tree algorithms involve complicated heuristics for determining the field and number of cuts. Moreover, fixed interval-based cutting not relating to the actual space that each rule covers is ineffective and results in a huge storage requirement. A packet classification algorithm using boundary cutting is proposed in [39]. It finds out the space that each rule covers and performs the cutting according to the space boundary. In [40], a rule organized optimal matching which splits the identification rules into several fields and elaborately organizes the matching order of the fields has been proposed. In [41], an asynchronous pipeline architecture based on a signature tree structure to combine the intermediate results returned from single-dimensional searches by hash tables has been proposed. Decision tree algorithms have been modified, supplement and optimization in [39]–[41]. However, they are for the same environment needs to be solved and their tranfer quality.

III. SYSTEM MODEL

We consider a open C-RAN as shown in Fig. 2, which consists of a radio access network and a cloud-based data center [9], [42]. Any company or person can provide and rent infrastructure for the customers based on open operation and business support system (OOBSS) [43]. OOBSS analyzes the components of Packet Data Protocol (PDP) for every application. As some functions needed by applications might be default, the binary digits of PDP character string of each request might be changeable. We use 0 to complement the default field; hence the multi parallel pipeline organization is used and every signalling is divided into substreams by binary digits [47], [48]. Set $B = \{b_1, b_2, \cdots\}$ as the required content set of PDP. b_j is the j^{th} field of PDP. Every field is a binary character string with w_i digits. Each time requests come, the field of PDP is checked and the frequency of each field is calculated. Details on the properties of PDP's each field are described in Table 1.

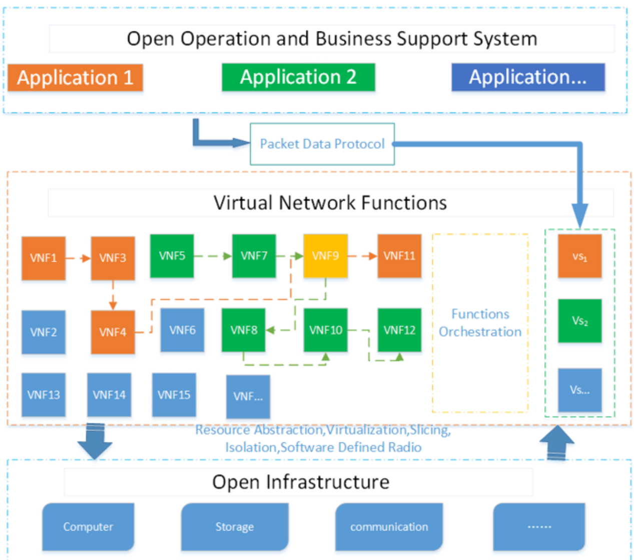


FIGURE 2. System framework.

Meanwhile, any company or person can rent out their redundant infrastructure through open infrastructure system(OIS) if their needs are satisfied [44]. OIS abstract lease

TABLE 1. Properties of PDP's each field.

name	description	attribute
cid	Context identifier of PDP. A number parameter, used to stipulate context definition of specific PDP(1-32). This parameter is local for Terminate and Mobile Terminate.	
PDP_type	Protocol type of grouped data. A parameter of character string, used to stipulate character string parameter of grouped data protocol.	'IP' is internet protocol; 'ppp' is point to point Protocol.
APN	Access name. A parameter of character string, used to choose GGSN or exterior grouped data network as logical name.	If this value is empty or neglected, preselected value is required.
PDP_addr	A parameter of character strings, used to identify MT in the address space of PDP.	If this value is empty or neglected, TE should provide a value in the start process of PDP; if this fails, a dynamic address is required.
d_comp	A number parameter to control dada compression of PDP	'0',closed; '1',open; retention value
h_comp	A number parameter to control PDP compression	'0',closed; '1',open; retention value
precedence	A number parameter to stipulate priority types.	'0', preselected; '1', high priority; '2', normal priority; '3',low priority
delay	A number parameter to stipulate delay	'0',<0.5; '1',<5; '2',<50; '3',try to type.
reliability	A number parameter to stipulate the types of reliability	'0','1','2','3','4','5'
peak	A number parameter to stipulate peak throughput	'0','...','9'
mean	A number parameter to stipulate the types of average throughput	'0','...','31
pd	Character string parameter form 0 to N, such as designated quality service	

these resources into the VNFs in C-RAN by virtualization, slicing, isolation, software defined radio, and so on. Meanwhile OIS performs centralized signal processing and provides coordinated radio resource and interference management for each access infrastructure. A practical front haul is always capacity and delay constrained, which can significantly reduce the spectrum efficiency gain achieved by C-RAN [47]. OIS not only can abstract infrastructure into VNFs, but also can collect relevant hardware resource conditions to make hardware resources under central management and monitoring to make virtual network functions independent of any hardware platforms [9]. Set $A = \{a_1, a_2, \cdots\}$ as the set of VNFs. a_i is a specific VNF; the size of A is determined by the number of functions abstracted by infrastructure.

By analyzing the occurrence of local applications and the use of relevant infrastructure, network management experts orchestrate VNFs into virtual service (VS). A VS consists of a or many VNF a_i, which is $vs_j = \{\overline{A}|\overline{A} \subseteq A\}$. Each VNF is a binary character strings of w_i digits. Network self-management system allocates tailored service for application according to the request from users. Each VNF corresponds to a field type in PDP. Based on bit vector strategy [10], multi-match packet classification can also be used by multi-function devices that perform single-match packet classification for each function [45], [46]. Entry invalidation scheme appends a valid bit to each entry in ternary content addressable memory (TCAM), where the valid bit of a matching rule is disabled to avoid being matched again. The maths description of the match between applications and virtual service is shown in table 2.

TABLE 2. Maths description of system framework.

PDP request	a_1	a_2	$\cdots$	a_n	VS
B_1	0011	0000	$\cdots$	$\cdots$	vs_1
B_2	*11	0010	$\cdots$	$\cdots$	vs_1
B_3	01*	*	$\cdots$	$\cdots$	vs_2
B_4	10*	*	$\cdots$	$\cdots$	vs_3
B_5	11*	*	$\cdots$	$\cdots$	vs_4
B_m	$\cdots$	$\cdots$	$\cdots$	$\cdots$	vs_m

IV. SOLVING STRATEGIES

As virtual radio access platform is open-source architecture, there are two difficulties: (1) trained samples for learning and tested samples can not satisfy the independent identically distributed condition. (2) the features of trained samples for learning might not be the same as those of tested samples(network platform might change). Based on given data X, we directly train data to establish identifying model $P(Y|X)$, as shown in offline learning strategy of 'A' chapter. In 'B' chapter, the methods to identify applications online are provided. As the distribution of original area and target area might not be the same and joint probability $P(Y, X)$ might not be considered, sometimes good results may not be gained. The knowledge transmission between original areas and target areas will be studies to improve the quality of algorithm. Meanwhile, the features of original areas and target areas might not be the same, this may result in unreliability of classifiers gained by original feature training. We proposed a feature exchange strategy, as shown in 'c' chapter. The whole process is shown in Figure 3.

A. OFFLINE LEARNING STRATEGY

To improve classifying speed of radio data packets and reduce the use of TCAM, IDTLM is set up in given trained data $D = \{(X_1, Y_1), \cdots, (X_m, Y_m)\}, X_i \subseteq A, X_i = \{A_i, \cdots, A_j\} = \{x_1, \cdots, x_m\}, Y_i \subseteq B, Y_i = \{b_i, \cdots, b_j\} = \{y_1, \cdots, y_m\}$. Then IDTLM is used to classify in virtual radio access online platforms. The establishing process of IDTLM is as follows:

Step 1: To avoid the phenomenon that when the VNFs are excessive, dimension dewater emerges, and to reduce the difficulties of learning, set the identifying feature set of IDTLM as A', $A' \subseteq X \subseteq A$, as shown in part 'Feature Selection' of this section.

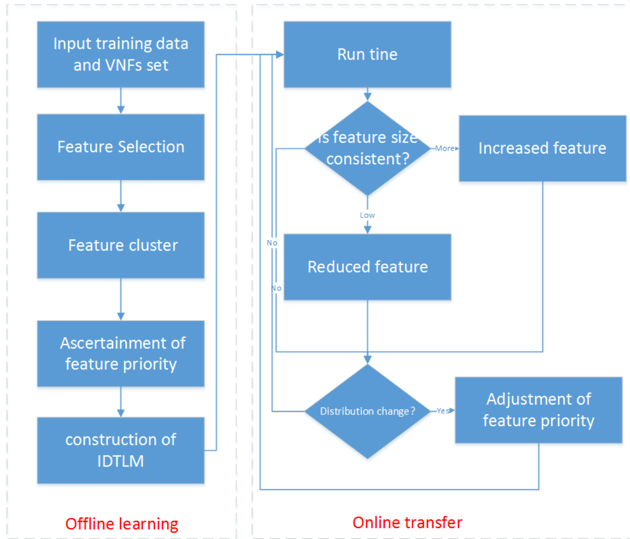


FIGURE 3. Solution process.

Step 2: To reduce the number of rules of online feature identification, each feature x_i' in identifying feature set A' is clustered, as shown in part 'Feature Cluster' of this section.

Step 3: To reduce the contrastive number of data packets in each field, the priority level of identifying features is confirmed, as shown in part 'Ascertainment of Feature Priority' of this section.

Step 4: IDTLM is Constructed according to the priority level of identifying features, as shown in part 'Construction of IDTLM' of this section.

1) FEATURE SELECTION

In consideration of linear regression model, square error is loss function, and thus the optimized objective of trained samples for the feature selection is

$$\min_{\omega} \sum_{i=1}^{m} (y_i - \omega x_i)^2 \tag{1}$$

ω is the coefficient of features. When the features of samples are too many and the number of samples are comparatively few, formula (1) is easy to be overfitting. L_1 norm is used to regularize, and formula (1) is changed into

$$\min_{\omega} \sum_{i=1}^{m} (y_i - \omega x_i)^2 + \lambda \|\omega\|_1 \tag{2}$$

Set $f(x) = \sum_{i=1}^{m} (y_i - \omega x_i)^2$, then formula(2) is changed into

$$\min_{x} f(x) + \lambda \|\omega\|_1 \tag{3}$$

If $f(x)$ is derivable function and $\nabla f(x)$ satisfys the condition of Lipschitz continuity, then there exists a constant L ($L > 0$). For any x', it can be written as

$$\|\nabla f(x') - \nabla f(x)\|_2^2 \leq L\|x' - x\|_2^2 \tag{4}$$

∇ is differential operator. Second order Tailor expansion of $f(x)$ is done to neighbouring x_k, available to

$$f(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{L}{2}\|x - x_k\|^2 \quad (5)$$

$\langle \nabla *, * \rangle$ is inner product. Obviously, in x_{k+1} lowest value could be gained by

$$x_{k+1} = x_k - \frac{\nabla f(x_k)}{L} \quad (6)$$

According to Proximal gradient descent method (solving process is in APPENDIX A), x_{k+1} could be gained by

$$x_{k+1} = arg \min_x \frac{L}{2}\|x - (x_k - \frac{\nabla f(x_k)}{L})\|_2^2 + \lambda \|x\|_1 \quad (7)$$

Set $z = x_k - \frac{\nabla f(x_k)}{L}$, feature a_i is the i^{th} component of X. Suppose feature a_i is independent, $a_i a_j$ $(i \neq j)$ does not exist. Expand the components of formula (7), closed-form solution could be gained by

$$a_i = \begin{cases} z^i - \lambda/L & \lambda/L < z^i \\ 0 & z^i \leq \lambda/L \\ z^i + \lambda/L & z^i < -\lambda/L \end{cases} \quad (8)$$

Solution process is in APPENDIX B. Take non 0 feature a_i as the identifying feature set A' of IDTLM.

2) FEATURE CLUSTER

Divide samples by their similar degree, so that the similar degree of the same elements is higher than that of different elements. The aim is to maximize the homogeneity of the same element and to maximize the heterogeneity of different elements. The main ground for this is that the samples in the same data cluster should be alike while those in different dada clusters should be differential. Every feature a_i in the feature set A' from formula (8) is clustered.

Set k as the type number of applications already known, cluster center is $V = \{v_1, v_2, \cdots, v_k\}$. $x^{a_i} = \{x_1^{a_i}, x_2^{a_i}, \cdots, x_m^{a_i}\}$ is the a_i feather in sample set D, then $x_b^{a_i} v_c$ means sample $x_b^{a_i}$ belongs to group v_c. Set value as 0 or 1, group optimized model of feature a_i in useful feature set A' is

$$obj \quad min \sum_{b=1}^{m} \sum_{c=1}^{k} \left([x_b^{a_i} v_c]^f \times dist_{b,c} \right)$$
$$s.t. \sum_{c=1}^{k} x_b^{a_i} v_c = 1$$
$$x_b^{a_i} v_c = \begin{cases} 1 & if \ x_b^{a_i} \in v_c \\ 0 & other \end{cases} \quad (9)$$

$f > 1$ is a constant which can control the fuzzy degree of cluster results. $dist_{b,c}$ is the Minkowski distance of the b^{th} sample and c^{th} group center.

$$dist_{b,c} = (|x_b^{a_i} - v_c|^d)^{(1/d)} \quad (10)$$

d is the measure of Minkowski distance. When $d = 1$, Minkowski distance is Manhattan distance. When $d = 2$, Minkowski distance is Euclidean distance. When $d \Rightarrow \propto$, Minkowski distance is Chebyshev distance. According to Lagrange's multiplier (solving process is in APPENDIX C), The iterative formula of feature cluster could be gained by

$$v_c = \frac{\sum_{b=1}^{m} (x_b^{a_i} v_c \times x_b^{a_i})}{\sum_{b=1}^{m} x_b^{a_i} v_c}, \quad 1 \leq c \leq k \quad (11)$$

$$x_b^{a_i} v_c = \left[\sum_{e=1}^{k} \left(\frac{dist_{b,c}}{dist_{e,c}} \right)^{\frac{1}{f-1}} \right]^{-d}, \quad 1 \leq b \leq m, \ 1 \leq c \leq k \quad (12)$$

3) ASCERTAINMENT OF FEATURE PRIORITY

Before the construction of the identifying trees of business, the priority level should first be ascertained. Generally speaking, the more samples the branch nodes of decisions trees contain, the higher the purity and the higher the priority level. The information entropy of feature a_i is

$$Ent(D) = -\sum_{i=1}^{k} p_{(a_i)} \times \log_2 p_{(a_i)} \quad (13)$$

$p_{(a_i)}$ is the probability for feature a_i to occupy data set. It can be seen form formula (11) that feature a_i has $V = \{v_1, v_2, \cdots, v_k\}$ values. If feature a_i is used to divide data set, then V branch nodes will be generated. The v_i branch node sample is written as D^v. Give branch node v_i weight $|D^v|/|D|$, then information gain of feature a_i in sample set D is

$$Gain(D, a_i) = Ent(D) - \sum_{v_1}^{V} \left[\frac{D^v}{D} \times Ent(D^v) \right] \quad (14)$$

The higher the information gain, the higher the purity of the features. Thus information gain is in favor of features with more available number. To reduce the bad effects from this preference, gain ratio is defined as:

$$Gainratio(D, a_i) = \frac{Gain(D, a_i)}{Iv(a_i)} \quad (15)$$

$$Iv(a_i) = -\sum_{v_1}^{V} \left[\frac{D^v}{D} \times \log_2 \frac{D^v}{D} \right] \quad (16)$$

4) CONSTRUCTION OF IDTLM

It can be seen form formula (15) that information gain rate is in favor of features with less number. Thus when IDTLM is constructed, the feature a_i in identifying feature set A' with the highest information and higher than the average is chosen as root node. Then the v_i branch will be identified. If the recognization of application could be improved, the feature a_i' from the v_i branch node in sample D^v with the highest information gain and higher than the average will be chosen as the v_i branch node of IDTLM; otherwise the v_i branch is identified as leaf node. And so on, the process continues until all the identifying feature set A' is used. The construction process of IDTLM is shown in Algorithm 1. $(*, *)$ is a two-tuples.

Algorithm 1 Construction of IDTLM

Input: $D, A' = (a_i, v_c)$
Output: $root$

$root = \max\limits_{A'} Gainratio(D, a_i)$
$node = root$
treenodegenerate (D, A')
if $D.Y \in vs_*$ **then**
  $node = vs_*$
  return
end if
if $A' = \emptyset$ or $D.X.a_i$ is same **then**
  $node = max(vs_*)$
  return
end if
for $a_i \to v_{j=1}, j \leq c, j++$ **do**
  if $D_v = \emptyset$ **then**
    $node = (a_i, max(D_v))$
    return
  end if
  if $\frac{max(d_v.Y)}{D_v} \leq \sum_{k=1}^{(a_i,v_c)} \frac{D_{(a_i,v_k)}.Y}{D_v}$ **then**
    $node = node \to v_j$
    $node = \max\limits_{A'-a_i} Gainratio(D_v, A' - a_i)$
    treenodegenerate $(D_{(a_i,v_k)}, A' - a_i)$
  else
    $node = vs(max(D_v.Y))$
    return
  end if
end for

Obviously, the generative process of IDTLM is a recursive process. There are four situations in which the overall process might be recursive: (1) Samples of current nodes belong to the same type without need for division. (2) Current features are empty or the features of all the samples are the same thus it is impossible to divide. (3)Sample set of current nodes is empty thus it is impossible to divide. (4)The increase of child nodes could not bring the improve of Generalization of IDTLM thus the division is stopped. We identify current nodes as leaf nodes and set its type as the one with most samples.

B. ONLINE IDENTIFYING STRATEGY

The online identifying process of IDTLM is a process of traversal tree, as shown in Algorithm 2. From the root node of the tree, the node value of IDTLM is compared with the value of the corresponding field. If the value of the field belongs to the sub-domain of IDTLM nodes, the sub-domain of IDTLM will be recursed until the corresponding virtual service is found.

C. ONLINE TRANSFER STRATEGY

To improve the adaptability of IDTLM, transfer learning is carried out aiming at the distribution change of the applications in new scenarios. After the process of time T, first it will be identified whether the feature of new scenario changes and if it changes, the transfer strategy of feature change will be used; and then it will be identified whether the distribution of the applications changes and if it happens, the priority level of corresponding features will be regulated.

Algorithm 2 Online Identifying Strategy

Input: $root, B = \{b_1, b_2, \cdots\}$
Output: vs_k

SearchIDTLMtree()
$node = root$
if $b_i = node.name$ **then**
  for $node \to v_{j=1}, j \leq e, j++$ **do**
    if $node \to v_j \notin VS$ **then**
      if $b_i.value \in node \to v_j.value$ **then**
        SearchIDTLMtree$(node \to v_j, B - b_i)$
      end if
    else
      return (vs_k)
    end if
  end for
end if

Algorithm 3 Reduced Feature Strategy

Input: $\overrightarrow{A}, Ngainratio, root$
Output: $root$

Reduce()
if $a' \in \overrightarrow{A}$ **then**
  $Ngainratio.a' =' 0'$
end if

1) TRANSFER STRATEGY OF FEATURE CHANGE

First, it will be checked whether the virtual network functions change. If they change, set $\overrightarrow{A}$ as reduced feature, and set $\overleftarrow{A}$ as increased feature set. And then if the network functions of virtual network in transfer scenarios decrease, the new information gain rate in $\overrightarrow{A}$ will be 0, as shown in Algorithm 3.

And then if the number of VNFs in transfer scenarios is on the increase, A feature selection in '1) Feature Selection' will be used to select the features in set $A + \overleftarrow{A} - \overrightarrow{A} - A'$. To reduce the repetitive calculation of feature a' which is not of great importance of application recognition, set the probability of reselecting features a' as

$$p(a') = \begin{cases} 0 & \varphi \geq \psi \\ \dfrac{\psi - \varphi}{\psi} & \varphi < \psi \end{cases} \tag{17}$$

ψ is the threshold value. φ is the count which is the number of a' rejected. At last, the new features are compared with the parent nodes of leaf nodes in IDTLM. If the genericity of IDTLM could be improved, insert the features; otherwise $\varphi = \varphi + 1$. The process of the increase of IDTLM features is shown as Algorithm 4.

Algorithm 4 Add Feature Strategy

Input: $A + \overleftarrow{A} - \overrightarrow{A} - A'$, $Ngainratio$,$root$
Output: $root$

```
Add()
if a'.v_c ∈ VS then
    treenodegenerate(a', D_N)
end if
```

2) TRANSFER STRATEGY OF DISTRIBUTION CHANGE FOR EACH APPLICATION

First, it will be checked from root nodes whether the maximum information gain of child nodes is bigger than the new information gain rate itself; if not, the whole IDTLM will be checked. Or it shows that the distribution of each application is under change, and Algorithm 5 will be used to solve the problem. And then the situation will be researched that the information gain rate of child nodes is bigger than that of parent nodes. Algorithm 5 shows that first child nodes will replace parent nodes; and then the information gain rate of parent nodes will be compared with that of child nodes. If that of parent nodes is bigger then parent nodes will replace child nodes and meanwhile the child nodes of child nodes will be given to the child nodes of parent nodes which have the highest information gain rate otherwise the traversal tree will be through the whole IDTLM.

Finally, if the distribution of applications changes, pruning process will be taken from the parent nodes of leaf child nodes [49]. Set $\mid T \mid$ as the number of leaf child nodes of decision-making tree IDTLM. So, $\mid T \mid$ is the complexity degree of model. t_i is the number of a specific leaf node, t_i has N_{t_i} sample nodes, $H_{t_i}(T)$ is the empirical entropy of leaf node t_i, $\beta \geq 0$ is parameter and the loss function of pruning is defined as:

$$C_\beta(T) = \sum_{i=1}^{|T|} N_{t_i} \times H_{t_i}(T) + \beta \mid T \mid \tag{18}$$

Empirical entropy is

$$H_{t_i}(T) = -\sum_{i=1}^{|T|} \frac{N_{t_i}}{N_t} \log_2 \frac{N_{t_i}}{N_t} \tag{19}$$

Parameter $\beta \geq 0$ controls the influence between the complexity degree of control model IDTLM and the fitting degree of trained data. The bigger the β, the simpler the model. When $\beta = 0$, the complexity degree will not be considered. Pruning is to chose the minimum model of loss function when β is certain. The proving process is shown in APPENDIX D.

3) A CASE OF TRANSFER STRATEGY

The decision-making tree constructed by IDTLM is shown in Fig 4. After process of time T, the information gain rate before and after transfer is shown in table 3. The results of the transfer strategy of distribution change for each application is shown Fig 5.

Algorithm 5 Transfer Strategy of Distribution Change for Each Application

Input: $A' = (a_i, v_c)$, $Ngainratio$, $root$
Output: $root$

```
looktree()
node = root
if node.Ngainratio ≥ max(node → v_*.Ngainratio) then
    for node → v_{i=1}; i ≤ c; i++ do
        if node → v_i ∉ VS then
            looktree(node → v_i)
        else
            return
        end if
    end for
else
    distribution-change(node,node → max(Ngainratio))
    temp1 = node
    temp2 = node → max(Ngainratio)
    node = temp2
    for temp2 → v_{j=1}, j ≤ c,j++ do
        if temp1.Ngainratio ≥ temp2 → v_j.Ngainratio then
            temp3 = temp2 → v_j
            temp2 → v_j = temp1
            temp1 → v_j = temp3
        else
            temp2 = temp2 → v_j
            if temp2 ∈ VS then
                return
            else
                distribution-change(node,temp2)
            end if
        end if
    end for
end if
```

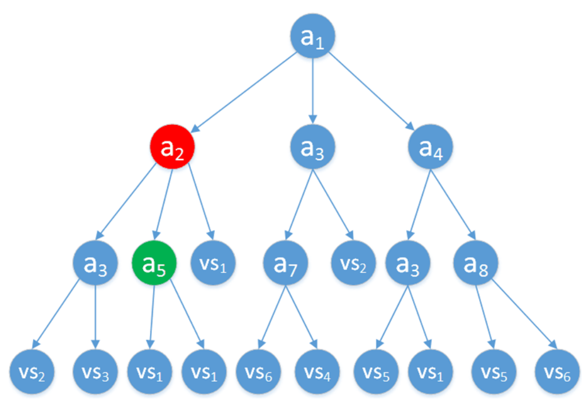


FIGURE 4. Model chart of decision-making trees constructed by IDTLM.

V. SIMULATION AND DISCUSSION

A. SYSTEM TEST

1) TEST ENVIRONMENT

To test the quality of IDTLM, test verification environment is established as shown in Fig 6. Wireless router TL-WR841N

TABLE 3. Comparison of information gain rate of each feature before and after transfer, after the process time T.

Name of feature	Gainratio	Ngainratio
A_1	4.088	3.56
A_2	3.78	3.78
A_3	2.68	2.68
A_4	0.54	0.54
A_5	3.6	3.6
...	...	...

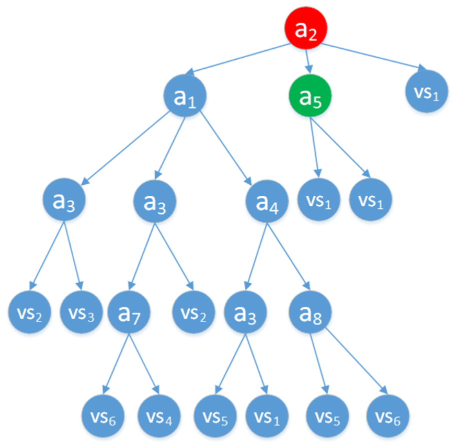


FIGURE 5. Model chart of decision-making tree constructed by IDTLM after the use of change and transfer strategy for applications.

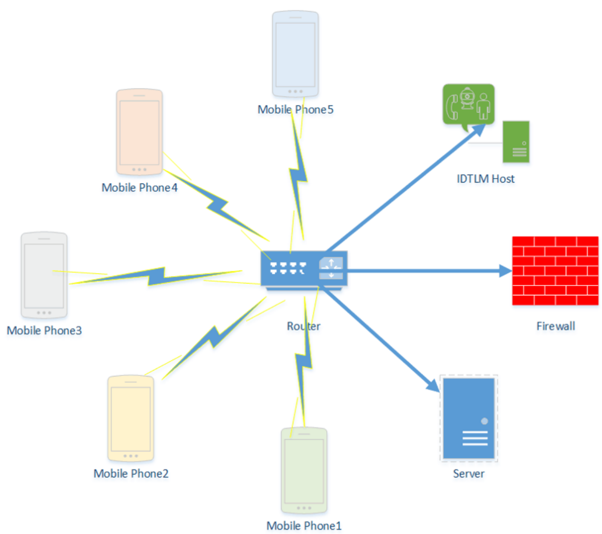


FIGURE 6. Environment of IDTLM quality test.

300M is used to connect with the internet whose broad-width is 100Mbps and it could be segregated from the outer net. Test terminate is C8500 mobile phone and PC of Huwei. Test environment is used to simulate mobile communication environment, and network card of the main engine in IDTLM system is set as promiscuous mode to obtain signal requests. We first test the information request from a single application and then test the information request form various applications. For the convenience of evaluating the quality

TABLE 4. Recognition result of single application.

Application	HTTP	QQ	Bittorrent	pplive	wechat
Number of single	982	1266	317	466	389
False positive rate	0	0	0	0	1.028%

TABLE 5. Composition information of mixed data files for known applications.

Application	QQ	pplive	wechat	Bittorrent	HTTP	fetion
Number of application	1266	466	389	317	982	232
Size of application	372.037	440.458	378.8	49.849	667.596	48.688

of the system, we define false positive rate. False positive rate is the rate when the data flow of network applications is wrongly recognized as another network application. False positive rate is

$$U_F = \frac{P_n}{P_{all}} \tag{20}$$

P_n is the number of false plosive request and P_{all} is the number of all the requests.

2) TEST OF SINGLE APPLICATION

The application data flow of 'HTTP', 'QQ', 'Bittorrent', 'pplive' and 'wechat' are chosen for test. Every application corresponds to one mobile terminal in table 5. That terminal sends single application to mobile computer on its own. IDTLM system then recognizes applications whose results are shown in table 4.

After IDTLM system test, the false positive rate of 'HTTP', 'QQ', 'Bittorrent', and 'pplive' is 0. Among the 389 packets sent by 'WeChat', four video data sent by 'WeCha't is wrongly recognized as stream media application. The false positive rate of 'WeChat' is 1.028%. The total false positive rate of IDTLM system for single application is 0.117%.

3) TEST OF MIXED APPLICATIONS

Mobile terminals in table 5 send applications to mobile computers and we take these data for research. IDTLM system test shows that there are 3652 data packets and there is 1957.43KB data flow. It can be seen from the analysis that 1.91kb data packets are wrongly recognized as other applications. False positive rate is 0.1095%. The test result is shown in Fig 7.

4) MIXED TESTS OF NEW APPLICATIONS

And then several groups of known application data of mobile communication('Fetion' and 'QQ' etc.) are combined with 'BitTorrent' data and data of untrained applications. All the data include 2052 known applications packets(613.16KB flow), 489 unknown applications packets(476.15KBflow) and the total amount of packets is 2541(1089.31KB flow).

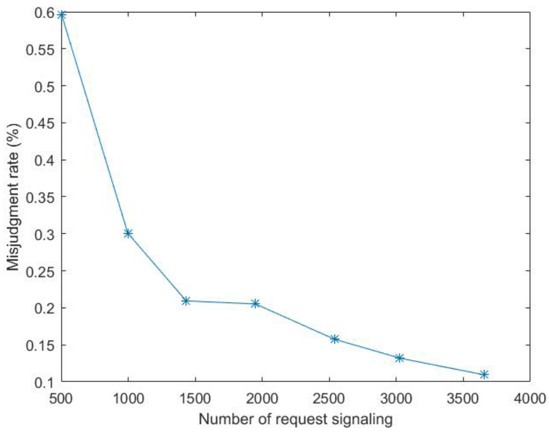


FIGURE 7. Test results of known mixed applications.

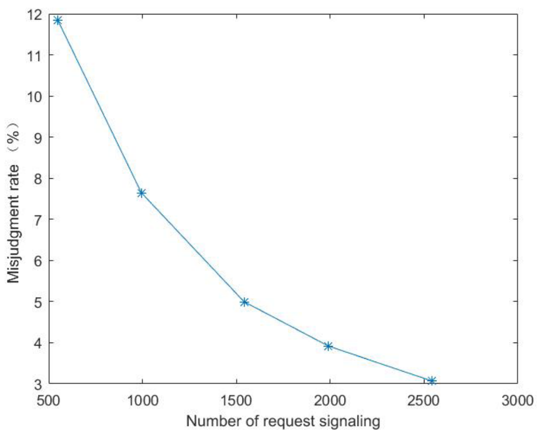


FIGURE 8. Test results of new mixed applications.

After the test of IDTLM system, there are 2031 known business packets and 602.82KB flow. Of the trained applications two data packets are recognized as other businesses and of the untrained applications there are 76 data packets are wrongly recognized as other businesses. The false positive rate of this test is 3.0696%,which is shown in Fig 8.

B. ANALYSIS OF TRANSFER QUALITY

To test the transfer quality of model, the test is carried out from the aspects of time complexity, storage space, recognition rate in both static and dynamic environment. Our data come from Classbench [50]. As there are 7 data in Classbench and each datum has 50 K rules in 6 fields, we expand the data on this basis to 100,000 and each datum has 100 fields.

1) OPERATING TIME AND TRANSFER QUALITY

Fig 9 shows the operating time of CPU to run 10000 data when IDTLM system is transferred from training platform to testing platform. 'Normal' is the test result when test data and trained data barely change. This figure also shows the operating time of CPU running 10000 data when there is 10%, 30% and 50% change in the application distribution

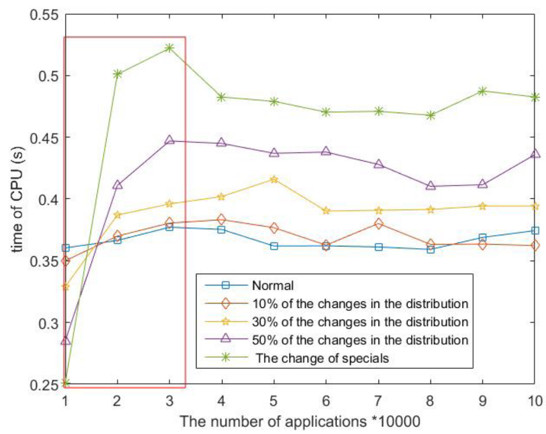


FIGURE 9. With the change of the number of applications, the CPU run time which the application do match the virtual service.

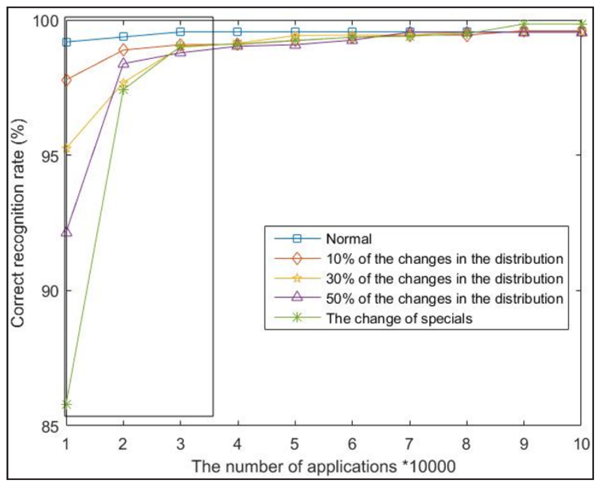


FIGURE 10. With the change of the number of applications, the correct recognition rate which the application do match the virtual service.

of test data and trained data. 'The change of specials' is the operating time of CPU running 10000 data when system test model changes (features changes). It can be seen from Fig 9 that the change of specials is the shortest for the first 10000 data. Meanwhile it can be seen that the bigger the probability of the change of application distribution for test data and trained data, the shorter the operating time of CPU. Whether the system models of application distribution change, after the operation of 30000 data, the operating time of CPU to recognize applications tend to balance.

2) RECOGNITION RATE AND TRANSFER QUALITY

Figure 10 shows the recognition rate $1 - U_F$ of system to recognize 10000 data when IDTLM system is transfered from train platform to test platform. Of the first 10000 test data, the recognition of the change of specials is the lowest. The bigger the probability of the change is application distribution of test data and train data, the lower the recognition rate. Whether the system model or the application distribution changes, after

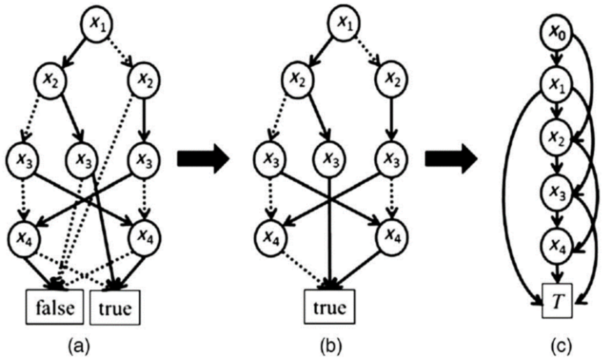


FIGURE 11. Example of OBDDs to RCOBDDs transformation: (a) initial tree; (b) removing false node; and (c) result of transformation [28].

the operation of 30000 data, the recognition rate for every test case tend to be the recognition rate of Normal test case.

C. DISCUSSION

1) DISCUSSION OF TIME COMPLEXITY

The search time complexity of IDTLM is the actual compared times of request signals. The number of storage space is the biggest branch node number. Set $p_{(vs_i)}$ as the probability of vs_i and $L_{(vs_i)}$ is the length of IDTLM coding. The search time complexity of IDTLM is

$$O(IDTLM) = \sum_{vs_i}^{\Omega} p_{(vs_i)} \times L_{(vs_i)} \quad (21)$$

In [28], a node optimized decision algorithm of RCOBDDS was proposed. N is the number of features. $P(x)$ is hit rate. $\sum_{j=1}^{N} p_{(vs_i)} \times (N-j)$ is the comparative times when the application in Fig 11 does not match with x1, x2 or x3. Its time complexity is

$$\begin{aligned} &O(RCOBDDS) \\ &\quad = P(x) \times \sum_{vs_i}^{\Omega} p_{(vs_i)} \times L_{(vs_i)} + (1 - P(x)) \\ &\quad\quad \times \left[\sum_{vs_i}^{\Omega} p_{(vs_i)} \times L_{(vs_i)} + \sum_{j=1}^{N} p_{(vs_i)} \times (N-j) \right] \quad (22) \end{aligned}$$

In [40], a rule optimized decision tree algorithm of ROOM was proposed. M is the biggest degree of nodes. The time complexity is

$$O(ROOM) = \begin{cases} (N-1)^M & \textit{it is worst} \\ N \times M & \textit{it is best} \end{cases} \quad (23)$$

It could be shown from formula (22) to (23), the time complexity of IDTLM is superior to RCOBDDS [8] and ROOM [40]. The proving process is in APPENDIX E and APPENDIX F.

2) DISCUSSION OF TEST RESULTS

It can be seen from table 4 that there is false positive rate in 'WeChat'. That's because 'WeChat' use encryption method to transmit data. The data content sent by 'WeChat' include various types of applications: video, voice, sound and image etc. It can be seen from Fig 7 and 8 that the longer the use of IDTLM system, the lower the false positive rate. That's because IDTLM system adopts transfer learning. It could automatically adapt to the change of application distribution and system recognizing models. It can be seen from figure 8 that during the operation of IDTLM, although it keeps modificating system models, the time to recognize 10 000 data is almost the same every time. That's because the modification of the system is in the background not affecting the decision of foreground. It can be seen from Fig 9 that the bigger the change of application distribution, the shorter for the IDTLM to recognize and in the later stage, recognition time tend to balance. That's mainly because the change of application distribution leads to the change of rules to identify features. In the later period, IDTLM system automatically adjusts the matching order of features to make time tend to be normal. It can be seen from Fig 9 and 10 that the bigger the change of application distribution, the shorter the time for IDTLM to recognize and the higher the false positive rate. That means IDTLM system could adapt to the transfer of different scenarios.

VI. CONCLUSION

This paper does research on applications and the match problem between the applications and tailored virtual service. It first uses Proximal gradient descent method to reduce dimensionality and uses Lagrangeąŕs multiplier to group the features, which solve the problem that there are redunant rules in traditional decision-making trees of packet classification and improve the matching speed of decision-making trees of packet classification and reduce cache. Then a new transfer learning method is proposed to improve the generalization quality of decision-making trees of packet classification. At last, simulated experiments show that the accuracy for trained applications could reach 99% and the accuracy rate could reach 96 % if untrained applications are included. Meanwhile, DTLM system is of high recognition speed, low false positive rate and could adapt to the transfer of different scenarios.

This paper mainly does research on virtualized radio access environment. In real world, the VNFs for the relevant hardwares in radio access network can not satisfy the requirements of automatic organization and management. Therefore in the following research, we will further research the functions of radio access network and the choreography technology of virtuallized network functions.

APPENDIX A

Set $f(x) = P$, $\omega = Q$, and $P + Q = S$ in(3). $P + Q = S$ convex relaxation to (3), the augmented Lagrangian function is

$$L(P,Q,\mu)=\mu\left(\|P\|+\lambda\|Q\|^{1}\right)+\frac{\|S-P-Q\|_2^2}{2} \tag{24}$$

$g(P,Q,\mu)$ is differentiable function. $f(P,Q)$ is Lipschitz continuity function.

$$g(P,Q,\mu)=\mu\left(\|P\|+\lambda\|Q\|^{1}\right) \tag{25}$$
$$f(P,Q)=\frac{\|S-P-Q\|_2^2}{2} \tag{26}$$
$$L(P,Q,\mu)=g(P,Q,\mu)+f(P,Q) \tag{27}$$

When $g(P,Q,\mu)$ is constant, and to the formal asymptotic approximations $f(P,Q)$ with the second order Taylor expansion, the formal asymptotic approximations of (27) is

$$\begin{aligned}&L(P,Q,\mu,Y^P,Y^Q)\\&\quad=g(P,Q,\mu)+f(Y^P,Y^Q)\\&\qquad+<\nabla f(Y^P,Y^Q),(P-Y^P,Q-Y^Q)>\\&\qquad+\frac{L\|P-Y^P,Q-Y^Q\|_2^2}{2}\end{aligned} \tag{28}$$

$f(Y^P,Y^Q)$ is $f(P,Q)$ near a known point. For $f(P,Q)$ partial derivatives, available to

$$\frac{\partial f}{\partial P}=-(S-P-Q)=P-(S-Q) \tag{29}$$
$$\frac{\partial f}{\partial Q}=-(S-P-Q)=Q-(S-P) \tag{30}$$
$$\Rightarrow\frac{\partial f}{\partial P}|_{\substack{P=Y^P\\Q=Y^Q}}=Y^P+Y^Q-S=Y^P-(S-Y^Q) \tag{31}$$
$$\frac{\partial f}{\partial Q}|_{\substack{P=Y^P\\Q=Y^Q}}=Y^P+Y^Q-S=Y^Q-(S-Y^P) \tag{32}$$

Further, (28) can be simplified as:

$$\begin{aligned}&L(P,Q,\mu,Y^P,Y^Q)\\&\quad=g(P,Q,\mu)+f(Y^P,Y^Q)\\&\qquad+<\nabla f(Y^P,Y^Q),(P-Y^P,Q-Y^Q)>\\&\qquad+\frac{L\|P-Y^P\|_2^2}{2}+\frac{L\|Q-Y^Q\|_2^2}{2}\end{aligned} \tag{33}$$

Set Q is constant, and $Q=Q^K$, $Y^P=Y_k^P$, $Y^Q=Y_k^Q$, $\mu=\mu^k$. Remove the constant, the (33) available to

$$\begin{aligned}&P^{k+1}\\&=arg\min_P L(P,Q,\mu,Y^P,Y^Q)\\&=arg\min_P \mu^k\|P\|+<Y_k^P-(S-Y_k^Q),P>+\frac{L\|P-Y^P\|_2^2}{2}\\&=arg\min_P \mu^k\|P\|+\frac{L\|P-[Y_k^P+(S-Y_k^Q-Y_k^P)/L]\|_2^2}{2}\end{aligned} \tag{34}$$

The related elements substituted into (34), for gain to (7).

APPENDIX B

If x_{k+1} is extreme value point, then $x_{k+1}\in\partial(7)$, to gain

$$\begin{aligned}&\Leftrightarrow 0\in\partial\left[\frac{L}{2}\|x-(\frac{\nabla f(x_k)}{L})\|_2^2+\lambda\|x\|_1\right]\\&\Leftrightarrow 0\in\frac{2\lambda}{L}\partial\|x\|_1+x-(\frac{\nabla f(x_k)}{L})\\&\Leftrightarrow x-(\frac{\nabla f(x_k)}{L})\in\frac{2\lambda}{L}\partial\|x\|_1\\&\Leftrightarrow x-z\in\frac{2\lambda}{L}\partial\|x\|_1\end{aligned} \tag{35}$$

$\partial(7)$ is derivative. At "0" only for the absolute value function that can guide, namely:

$$\partial|x|_1=\begin{cases}1 & x_i>0\\-1 & x_i<0\\ [-1,1] & x_i=0\end{cases} \tag{36}$$

And for $\lambda\|x\|_1$ level 1 norm, available (8).

APPENDIX C

With the Lagrange's multiplier, (9) can be converted to:

$$\begin{aligned}F(*,\lambda)=&\sum_{b=1}^{m}\sum_{c=1}^{k}(x_b^{a_i}v_c)^f+\lambda_1(\sum_{c=1}^{k}x_1^{a_i}v_c-1)\\&+\cdots+\lambda_b(\sum_{c=1}^{k}x_b^{a_i}v_c-1)\\&+\cdots+\lambda_m(\sum_{c=1}^{k}x_m^{a_i}v_c-1)\end{aligned} \tag{37}$$

The partial derivatives (37), can be:

$$\frac{\partial F(*,\lambda)}{\partial x_b^{a_i}v_c}=f\times dist_{b,c}\times(x_b^{a_i}v_c)^{f-1}+\lambda_b=0 \tag{38}$$
$$\Longrightarrow (x_b^{a_i}v_c)^{f-1}=\frac{-\lambda_b}{f\times dist_{b,c}} \tag{39}$$
$$\begin{aligned}\Rightarrow x_b^{a_i}v_c&=\left(\frac{-\lambda_b}{f\times dist_{b,c}}\right)^{\frac{1}{f-1}}\\&=\left(\frac{-\lambda_b}{f}\right)^{\frac{1}{f-1}}\times\left(\frac{1}{dist_{b,c}}\right)^{\frac{1}{f-1}}\end{aligned} \tag{40}$$

Let (40) be substituted into (9) type of constraint, available:

$$\begin{aligned}1&=\sum_{c=1}^{k}x_b^{a_i}v_c\\&=\sum_{c=1}^{k}\left(\frac{-\lambda_b}{f}\right)^{\frac{1}{f-1}}\times\left(\frac{1}{dist_{b,c}}\right)^{\frac{1}{f-1}}\\&=\left(\frac{-\lambda_b}{f}\right)^{\frac{1}{f-1}}\times\sum_{c=1}^{k}\left(\frac{1}{dist_{b,c}}\right)^{\frac{1}{f-1}}\\&=\left(\frac{-\lambda_b}{f}\right)^{\frac{1}{f-1}}\times\sum_{c=1}^{k}\left[\frac{1}{(|x_b^{a_i}-v_c|^d)^{\frac{1}{d}}}\right]^{\frac{1}{f-1}}\end{aligned} \tag{41}$$

$$\Rightarrow \left(\frac{-\lambda_b}{f}\right)^{\frac{1}{f-1}} = \frac{1}{\sum_{c=1}^{k}\left[\frac{1}{\left(|x_b^{a_i}-v_c|^d\right)^{\frac{1}{d}}}\right]^{\frac{1}{f-1}}} = \frac{1}{\sum_{e=1}^{k}\left[\frac{1}{\left(|x_b^{a_i}-v_e|^d\right)^{\frac{1}{d}}}\right]^{\frac{1}{f-1}}} \tag{42}$$

Let (42) be substituted into (40). The iterative formula of $x_b^{a_i} v_c$ is

$$\begin{aligned} x_b^{a_i} v_c &= \left(\frac{-\lambda_b}{f}\right)^{\frac{1}{f-1}} \times \left(\frac{1}{dist_{b,c}}\right)^{\frac{1}{f-1}} \\ &= \frac{1}{\sum_{e=1}^{k}\left[\frac{1}{\left(|x_b^{a_i}-v_e|^d\right)^{\frac{1}{d}}}\right]^{\frac{1}{f-1}}} \\ &\quad \times \left[\frac{1}{\left(|x_b^{a_i}-v_c|^d\right)^{\frac{1}{d}}}\right]^{\frac{1}{f-1}} \\ &= \left[\sum_{e=1}^{k}\left(\frac{dist_{b,c}}{dist_{e,c}}\right)^{\frac{1}{f-1}}\right]^{-d} \end{aligned} \tag{43}$$

$$\frac{\partial F(*,\lambda)}{\partial v_c} = -(x_b^{a_i} v_c)^{f-1} \times dist_{b,c} = 0 \tag{44}$$

$$\Rightarrow v_c = \frac{\sum_{b=1}^{m}(x_b^{a_i} v_c \times x_b^{a_i})}{\sum_{b=1}^{m}(x_b^{a_i} v_c} \tag{45}$$

APPENDIX D

When β sufficiently small, there is:

$$C_\beta(T) < C_\beta(t_i) \tag{46}$$

When β increases to a certain value, there is:

$$C_\beta(T) \geq C_\beta(t_i) \tag{47}$$

$$\Rightarrow \beta = \frac{\sum_{i=1}^{|T|} N_{t_i} \times H_{t_i}(T) - C_\beta(T)}{|T|-1} \tag{48}$$

APPENDIX E

$$\begin{aligned} (22) &= \sum_{vs_i}^{\Omega} p_{(vs_i)} \times L_{(vs_i)} + (1-p(x)) \sum_{j=1}^{N} p_{(vs_i)} \times (N-j) \\ &\geq \sum_{vs_i}^{\Omega} p_{(vs_i)} \times L_{(vs_i)} = O(IDTLM) \end{aligned} \tag{49}$$

APPENDIX F

$\because N' \leq N$

$\therefore$

$$(21) = \begin{cases} (N'-1)^M \\ N'M \end{cases} \leq (23) \tag{50}$$

REFERENCES

[1] Q. W. Xu. (Apr. 2013). *Cisco's Viewpoint: Operational and Commercial*. [Online]. Available: http://www.cisco.com/c/dam/global/zh_cn/solutions/sp/pdf/sp_customer_new_letter_4.pdf,

[2] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 84–91, Apr. 2016.

[3] K. Liang, L. Zhao, X. Chu, and H.-H. Chen, "An integrated architecture for software defined and virtualized radio access networks with fog computing," *IEEE Netw.*, vol. 31, no. 1, pp. 80–87, Jan./Feb. 2017.

[4] *Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, document ITU-R WP5D and ITU-RM, USA, 2015. [Online]. Available: http://www.itu.int/rec/R-REC-M.2083

[5] "5G concept of white paper," Promotion Group, Beijing, China, White Paper IMT-2020 (5G) and IMT-2020, 2015. [Online]. Available: http://www.imt-2020.org.cn/en/documents/3

[6] C. Ramirez-Perez, and V. Ramos, "SDN meets SDR in self-organizing networks: Fitting the pieces of network management," *IEEE Commun. Mag.*, vol. 54, no. 1, pp. 48–57, Jan. 2016.

[7] K. Tsagkaris, G. Poulios, P. Demestichas, A. Tall, Z. Altman, and C. Destré, "An open framework for programmable, self-managed radio access networks," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 154–161, Jul. 2015.

[8] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[9] S. Zou *et al.*, "RSCMF : A mapping framework for resource scalability collaboration in radio access networks with SDR and virtualization," *Electron. Lett.*, to be published.

[10] Y.-K. Chang and C.-S. Hsueh, "Range-enhanced packet classification design on FPGA," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 2, pp. 214–224, Apr./Jun. 2016.

[11] A. X. Liu, C. R. Meiners, and E. Torng, "Packet classification using binary content addressable memory," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1295–1307, Jun. 2016.

[12] L. Rizzo, "Netmap: A novel framework for fast packet I/O," in *Proc. USENIX*, Boston, MA, USA, Jun. 2012, p. 9.

[13] M. Varvello, R. Laufer, F. X. Zhang, and T. V. Lakshman, "Multilayer packet classification with graphics processing units," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2728–2741, Oct. 2016.

[14] T.-S. Chen, D.-Y. Lee, T.-T. Liu, and A.-Y. Wu, "Dynamic reconfigurable ternary content addressable memory for openflow-compliant low-power packet processing," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 10, pp. 1661–1672, Oct. 2016.

[15] Y.-X. Qi and J. Li, "Theoretical analysis and algorithm design of high-performance packet classification algorithms," *Chin. J. Comput.*, vol. 36, no. 2, pp. 408–421, 2013.

[16] W. Pak and Y.-J. Choi, "High performance and high scalable packet classification algorithm for network security systems," *IEEE Trans. Depend. Sec. Comput.*, vol. 14, no. 1, pp. 37–49, Jan./Feb. 2017.

[17] Y.-C. Cheng and P.-C. Wang, "Packet classification using dynamically generated decision trees," *IEEE Trans. Comput.*, vol. 64, no. 2, pp. 582–587, Feb. 2015.

[18] N. Chowdhury and R. Boutaba, "A survey of network virtualization," *Comput. Netw.*, vol. 54, no. 5, pp. 862–876, 2010.

[19] N. M. M. K. Chowdhury and R. Boutaba, "Network virtualization: State of the art and research challenges," *IEEE Commun. Mag.*, vol. 47, no. 7, pp. 20–26, Jul. 2009.

[20] (2014). *Network Functions Virtualization: An Introduction, Benefits, Enablers, Challenges and Call for Action*, [Online]. Available: https://portal.etsi.org/Portals/0/TBpages/NFV/Docs/NFV_White_Paper3.pdf

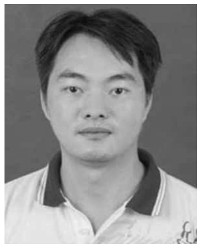
[21] C. Liang and F. R. Yu, "Wireless virtualization for next generation mobile cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 61–69, Feb. 2015.

[22] N. McKeown *et al.*, "OpenFlow: Enabling innovation in campus networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, Apr. 2008.

[23] L. E. Li, Z. M. Mao, and J. Rexford, "Toward software-defined cellular networks," in *Proc. Eur. Workshop Softw. Defined Netw. (EWSDN)*, 2012, pp. 7–12.

[24] J. Kempf, B. Johansson, S. Pettersson, H. Lüning, and T. Nilsson, "Moving the mobile evolved packet core to the cloud," in *Proc. IEEE WiMob*, Oct. 2012, pp. 784–791.

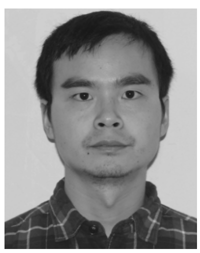
[25] K. Pentikousis, Y. Wang, and W. Hu, "Mobileflow: Toward software-defined mobile networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 44–53, Jul. 2013.
[26] M. Hadzialic, B. Dosenovic, M. Dzaferagic, and J. Musovic, "Cloud-RAN: Innovative radio access network architecture," in *Proc. 55th Int. Symp. ELMAR*, 2013, pp. 115–120.
[27] (2013). *Network Functions Virtualization: Use Cases*, [Online]. Available: http://www.etsi.org/deliver/etsi_gs/nfv/001_099/001/01.01.01_60/gs_nfv001v010101p.pdf
[28] W. Viriyasitavat, L. D. Xu, and W. Viriyasitavat, "A New approach for compliance checking in service workflows," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1452–1460, May 2014.
[29] S. Muhlbach and A. Koch, "A reconfigurable platform and programming tools for high-level network applications demonstrated as a hardware honeypot," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 10, pp. 1919–1932, Oct. 2014.
[30] Y.-C. Cheng and P.-C. Wang, "Scalable multi-match packet classification using TCAM and SRAM," *IEEE Trans. Comput.*, vol. 65, no. 7, pp. 2257–2269, Jul. 2016.
[31] Y. R. Qu and V. K. Prasanna, "High-performance and dynamically updatable packet classification engine on FPGA," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 1, pp. 197–209, Jan. 2016.
[32] X. Wang, B. Liu, J. Jiang, Y. Xu, Y. Wang, and X. Wang, "Kangaroo: Accelerating string matching by running multiple collaborative finite state machines," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 10, pp. 1784–1796, Oct. 2014.
[33] Y. Xu, J. Jiang, R. Wei, Y. Song, and H. J. Chao "TFA: A tunable finite automaton for pattern matching in network intrusion detection systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 10, pp. 1810–1822, Oct. 2014.
[34] M. Avalle, F. Risso, and R. Sisto, "Scalable algorithms for NFA multi-striding and NFA-based deep packet inspection on GPUs," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1704–1717, Jun. 2016.
[35] M. Leogrande, F. Risso, and L. Ciminiera, "Modeling complex packet filters with finite state automata," *IEEE/ACM Trans. Netw.*, vol. 23, no. 1, pp. 42–55, Feb. 2015.
[36] C. R. Meiners, J. Patel, E. Norige, A. X. Liu, and E. Torng, "Fast regular expression matching using small TCAM," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 94–109, Feb. 2014.
[37] J. Patel, A. X. Liu, and E. Torng, "Bypassing space explosion in high-speed regular expression matching," *IEEE/ACM Trans. Netw.*, vol. 22, no. 6, pp. 1701–1714, Dec. 2014.
[38] X. D. Yu, B. Lin, and M. Becchi, "Revisiting state blow-up: Automatically building augmented-FA while preserving functional equivalence," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 10, pp. 1822–1833, Oct. 2014.
[39] H. Lim, N. Lee, G. Jin, J. Lee, Y. Choi, and C. Yim, "Boundary cutting for packet classification," *IEEE/ACM Trans. Netw.*, vol. 22, no. 2, pp. 443–456, Apr. 2014.
[40] H. Li and C. Hu, "MP-ROOM: Optimal matching on multiple PDUs for fine-grained traffic identification," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 10, pp. 1881–1893, Oct. 2014.
[41] Y. Xu, Z. Liu, Z. Zhang, and H. J. Chao, "High-throughput and memory-efficient multimatch packet classification based on distributed and pipelined hash tables," *IEEE/ACM Trans. Netw.*, vol. 22, no. 3, pp. 982–995, Jun. 2014.
[42] M. Kallitsis, S. A. Stoev, S. Bhattacharya, and G. Michailidis, "AMON: An open source architecture for online monitoring, statistical analysis, and forensics of multi-gigabit streams," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 6, pp. 1834–1848, Jun. 2016.
[43] Y. Fu, H. Xiong, X. Lu, J. Yang, and C. Chen, "Service usage classification with encrypted Internet traffic in mobile messaging apps," *IEEE Trans. Mobile Comput.*, vol. 15, no. 11, pp. 2851–2864, Nov. 2016.
[44] B. Niu, Y. Zhou, H. Shah-Mansouri, and V. W. S. Wong, "A dynamic resource sharing mechanism for cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8325–8338, Dec. 2016.
[45] P.-C. Wang, "Scalable packet classification for datacenter networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 1, pp. 124–137, Jan. 2014.
[46] D.-Y. Chang and P.-C. Wang, "TCAM-based multi-match packet classification using multidimensional rule layering," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1125–1138, Apr. 2016.
[47] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.
[48] K. Kogan, S. I. Nikolenko, O. Rottenstreich, W. Culhane, and P. Eugster, "Exploiting order independence for scalable and expressive packet classification," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1251–1264, Apr. 2016.
[49] Z. H. Zhou, *Machine Learning*. Beijing, China: Tsinghua Univ. Press, 2016, pp. 80–83.
[50] ClassBench. (2004). *A Packet Classification Benchmark*. [Online]. Available: http://www.arl.wustl.edu/classbench/



SAI ZOU received the B.S. degree in computer science from Hengyang Normal University, Hengyang, China, in 2004, and the M.S. degree in software engineering from Hunan University, Changsha, China, in 2007. He is currently pursuing the Ph.D. degree in communications and information system with the School of Information Science and Engineering, Xiamen University, Xiamen, China. He was a Professor with the Chongqing College of Electronic Engineering, Chongqing, China. He is also a Visiting Student with the Division of CSIRO Computational Informatics, CSIRO, Australia. He has published over 20 papers in refereed mainstream journals and reputed international conferences and has been granted over six patents in his research areas. His current research interests are wireless communication systems and sensor network.

YULIANG TANG received the B.S. degree from the Chongqing University of Posts and Telecommunications, China, in 1987, the M.S. degree from the Beijing University of Posts and Telecommunications, China, in 1996, and the Ph.D. degree in communication engineering from Xiamen University, China, in 2009. He is currently a Professor with the Department of Communication Engineering, Xiamen University. He has published over 90 papers in refereed mainstream journals and reputed international conferences and has been granted over six patents in his research areas. His current research interests are wireless communication systems and vehicular ad hoc network.

YANGLONG SUN received the B.S. and M.S. degrees from Zhengzhou University, China, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree in communication engineering with Xiamen University, China. His research is focus on vehicular ad hoc network, mobile 5G networks, and intelligent transportation system.



KE SU received the B.S. degree in human resource management from the Chongqing University of Technology, Chongqing, China, in 2005, and the M.S. degree in technology economy and management from Chongqing University, Chongqing, in 2012. He is currently pursuing the Ph.D. degree in regional economics with Chongqing Technology and Business University, Chongqing. He was an Associate Professor with the Chongqing College of Electronic Engineering, Chongqing. He is currently a Visiting Scholar with the Renmin University of China, Beijing, China, with the project of young and middle-aged backbone teachers of Ministry of Education. He has published over ten papers and a monograph in refereed mainstream journals and reputed international conferences. His current research interests are human resource management and regional economic development.