IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 15, 2017, accepted January 26, 2018, date of publication February 5, 2018, date of current version June 29, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2802541

Det-LB: A Load Balancing Approach in 802.11 Wireless Networks for Industrial Soft Real-Time Applications

YUJUN CHENG, DONG YANG, (Member, IEEE), AND HUACHUN ZHOU
School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
Corresponding author: Dong Yang (dyang@bjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771040 and in part by the Electronic Information Technology Innovation and Cultivation under Grant Z171100001217004.

ABSTRACT Wireless control systems for industrial automation have been gaining popularity in recent years owing to the low cost of their components and their ease of deployment. Wireless local area network (WLAN) is a good candidate for industrial wireless applications, especially for soft real-time control systems. However, unbalanced traffic loads in areas that are covered by several access points (APs) could cause significant performance degradation, and the problem would be more serious in industrial contexts that have high requirements for timeliness and reliability. In this paper, we propose a deterministic load balancing algorithm (Det-LB) for WLAN, which is based on game theory, mainly intended for soft real-time control systems. The procedure of the proposed algorithm is described and analysed in detail. To validate the goodness of the proposed approach, we compare Det-LB with three other load balancing strategies (RLF, MMF, and DLBA) in a simulation. The results show that Det-LB can achieve better load balancing performance and improve network performance significantly. For instance, in a common scenario in which five APs are deployed, the RLF scheme has terrible performance and cannot be applied to industrial applications at all; Det-LB has 35% and 26% performance improvements over MMF in 500- and 250-byte payload size conditions, respectively, and has a 5% deadline miss performance boost and better packet loss performance compared with DLBA when the APs tend to be overloaded.

INDEX TERMS IEEE 802.11, load balancing, real-time networks, industrial wireless applications.

I. INTRODUCTION

In recent years, wireless communication has been widely adopted in the field of industrial automation [1], [2]. Due to the ease of deployment and the low cost of components, a large number of industrial applications have started to deploy wireless network control systems instead of wired control systems. As one of the most widely deployed network technologies, IEEE 802.11 wireless local area networks (WLAN) are a good candidate for industrial wireless applications, especially for soft real-time network control systems (NCS) [3], [4]. In practice, real-time control systems can be divided into two categories: hard and soft real-time systems. In hard real-time control systems, each of the tasks must be completed before a fixed deadline, and service within this span of time must be guaranteed. Missing a deadline can have catastrophic consequences. By contrast, although deadline misses degrade the quality of service (QoS) of the system in soft real-time control, slight deadline misses are tolerable as long as the QoS of the system does not fall beyond a threshold. The tolerance level depends on the specific requirements of underlying industrial applications. Thus, it is important to take the deadline constraint into consideration and keep it below the threshold when designing soft real-time control systems based on WLAN.

However, in a typical IEEE 802.11 network, if there is more than one AP covering the same area, the load of these APs is usually not balanced because the stations always associate with the AP that has the maximum received signal strength indicator (RSSI) in the default 802.11 standard mechanism (Fig. 1). The crowded and overloaded APs will experience severe QoS degradation, while the light load APs have plenty of idle communication resources. This problem may be more critical when adopting an 802.11 network in a soft-real time control system due to its relatively stringent QoS requirement

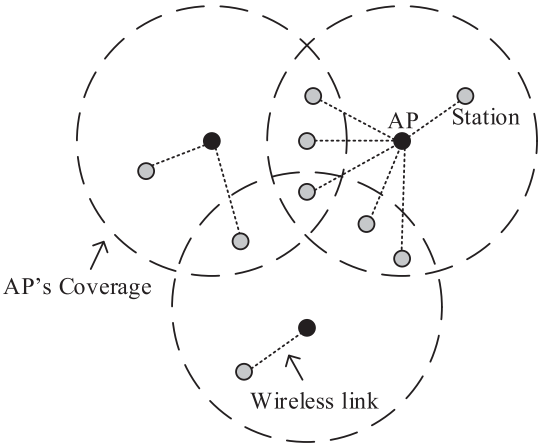


FIGURE 1. Stations in an overlapping area of multiple APs.

and will cause a considerable number of deadline misses beyond the threshold. Thus, an appropriate load balancing algorithm must be employed to prevent the network workload from being unbalanced in 802.11-based soft real-time control systems.

Load balancing is a technique that improves the distribution of workloads across multiple access points (APs) [5], [6]. It aims to optimize resource usage and avoid the overload of communication resources when different APs cover the same area or there is an overlapping area to which the stations can choose to connect. Based on the required performance parameters, the load balancing algorithm should decide to choose a more suitable AP for a designated station. Generally, a load balancing procedure is composed of two basic steps. First, the station that is overloaded or has downgraded performance should be removed from the currently connected AP. Second, the disconnected stations should associate with another appropriate AP, which has a lighter traffic load or better QoS performance. The two steps in the load balancing mechanism can be called a metric check and a load distribution procedure.

In this paper, we propose a deterministic load balancing algorithm (Det-LB) for WLAN, which is mainly intended for soft real-time control systems. The main contributions of this paper can be summarized as follows:

1) For the metric check procedure, most of the existing literature only takes signal strength, the number of associated stations or the bandwidth utilization into consideration, which may not satisfy the stringent QoS requirements very well in the deadline-aware NCS. Instead, this paper adopts the number of packet deadlines missed and packet loss ratio as two metrics to estimate whether a station should be removed from the current AP. A metric check algorithm is proposed based on the two real-time metrics.
2) In the load distribution procedure, a novel algorithm is presented to ensure the stations are able to select the optimal AP. The load distribution algorithm is based on an auction model in game theory, namely, a first-price sealed-bid auction game. To the best of our knowledge, this is the first study to model the load distribution process as an auction game.
3) To validate the performance of Det-LB, we compare Det-LB with three other general load balancing schemes (RLF, MMF and DLBA) with an NS-3 simulator. RLF is the default 802.11 association mechanism; MMF and DLBA are two representative works [7], [8]. We try our best to emulate industrial scenarios with particular simulation parameters, and the massive simulation results show better performance for Det-LB compared with the other three.

The rest of the paper is organized as follows. Section II reviews related work and motivates this research. Section III presents the proposed load balancing algorithm in detail. Section IV shows and analyses the simulation results of the proposed algorithm in detail. Finally, conclusions are given in Section V.

II. BACKGROUND

A. REVIEW OF STATE OF THE ART

The IEEE 802.11 standard does not define any mechanism for load balancing. Thus, there are a considerable number of works on this issue over the years. As a key part of the load balancing technique, load metrics should be specified in terms of the requirements of the network. Almost all existing IEEE 802.11 adaptors associate with the AP that has the strongest received signal strength. Based on the signal strength selection method, networks are prone to the problem [9] of uneven distribution of resources, which means some of the APs exceed or approach the load capacity while the load of the rest is relatively low. Balachandran *et al.* [10] propose a mechanism called Explicit Channel Switching. In this approach, the association procedure of stations is not merely based on signal strength, but it is also based on the workload of the corresponding APs. The algorithm provides a trade-off between signal strength and workload by switching stations from an overloaded AP with a stronger signal to a neighbouring lightly loaded AP where the signal to the access point may be weaker. However, timeliness and reliability are the most important features in industrial WLAN. These characteristics of the network cannot be reflected very well by the size of workload. Collotta *et al.* [8] use the number of deadline misses as an important load metric to reflect the network performance for industrial applications. There are two situations where a station should disconnect from the current AP: the signal strength is below the threshold, or the number of deadline missed packets is over a specified amount. Real-time requirements in industrial WLAN are fully taken into consideration in the algorithm. Collotta [11] describe a QoS concerned algorithm based on fuzzy logic. The load metric is provided by a specific fuzzy logic controller according to the combination of signal quality, deadline misses and packet loss.

On the other hand, Yen *et al.* [12] classify the existing load distribution schemes into two categories: wireless-station-based load distribution and network-based load distribution. In the wireless-station-based approach, wireless stations select an AP that maximizes their potential benefits all by themselves. Although this approach is easy to implement on both sides of stations and APs, it is hard to achieve network-wide load balance due to the lack of global administrators. On the other hand, network-based approaches adopt a centre network entity to control the load distribution of APs. Because the load distribution scheme in this paper adopts the network-based approach, the network-based approaches are quickly reviewed in the following paragraph.

Three basic techniques are usually adopted in network-based approaches to control the load of AP, including coverage adjustment, admission control, and association management. In the coverage adjustment technique [13], crowded APs reduce their transmission power of the beacon frame, so that new wireless stations are out of communication range and cannot associate with the crowded APs. Admission control mechanisms [14] allow an AP to reject the association of a new client if the load of the AP is beyond a certain threshold, and the mechanisms eliminate the overload risk of the AP. In the association management technique [15], a crowded AP can send a disassociation frame to an associated wireless stations according to some QoS requirements, and the targeted wireless stations reassociate with other lightly loaded APs. In addition, Bejerano *et al.* [7] model the load balancing problem as a max-min fair bandwidth allocation problem. The core idea of max-min fair bandwidth allocation is to achieve such a situation, that is, there is no way to give more bandwidth to any user without decreasing the allocation of a user with less or equal bandwidth. Xu *et al.* [16] propose a game theoretical approach, and the centralized and localized algorithms are designed for achieving the load balancing Nash equilibrium. In this way, the network can achieve an optimal load balancing state under the Nash equilibrium condition. However, these proposed load balancing algorithms hardly take timeliness and reliability into consideration when designed, and they may not be feasible for industrial WLAN.

B. AUCTION APPROACHES FOR WIRELESS LOAD BALANCING

With the fast development of wireless systems, the complexity of the wireless environment substantially increases due to the dynamic topology, heterogeneous architecture and growing network scale. Traditional static methods are sometimes inefficient or have difficulties addressing the problems in wireless systems, such as resource management. Therefore, researchers have begun to seek help from economics and business management approaches due to the similarity of their problem with, for instance, multiple participants transacting for resources under certain constraints. As one of the common models in economics, auctions are an interdisciplinary method used to solve these issues, especially for resource allocation in wireless systems [17], [18].

An auction is a process in which participants intend to buy or sell commodities. Generally, the basic roles in the auction process include bidder, seller, auctioneer and commodity [19]. In a typical auction process, bidders intend to buy a commodity from sellers with a specific set of rules hosted by an auctioneer. In the wireless context, the commodities are usually communication resources, such as bandwidth, licenses of spectrum, and time slots. The bidders and sellers are resource demanders and resource providers, respectively, and the auctioneer can be a centralized network coordinator to control the resource allocation procedure. More specifically, load balancing in wireless systems is also a resource allocation problem in a broad sense. Thus, it is feasible to adopt auction theory to model and analyse load balancing procedure in WLAN.

On the other hand, there are many kinds of auction types theoretically and practically. Here, we focus on sealed-bid auctions, which are often used to analyse wireless systems. In a sealed-bid auction, bidders put their bid in a sealed envelope and simultaneously give them to the auctioneer (seller), which means each bidder knows nothing about bidding strategies of others. A first-price sealed-bid auction is one of the most important sealed-bid auctions. In a first-price auction, the bidder who offers the highest bid will win the auction and get the commodity, and he must pay that price as payment. In a network-based load balancing scheme, the load information is sent to centralized network coordinator, and the APs do not know the load information of each other. In addition, it is reasonable to allocate the resource to the AP that intends to pay the highest price. Thus in Sec. III, we adopt a game theoretical first-price sealed-bid auction to model the load distribution process and analyse it in detail.

III. PROPOSED LOAD BALANCING ALGORITHM FOR INDUSTRIAL WLAN

The load balancing procedure is composed of the metrics check and the load distribution procedures. As we discussed in Sec. II, most of the existing load balancing mechanisms are based on load metrics such as signal strength, throughput, and the number of connected stations. Though these metrics may work well in a general IEEE 802.11 based wireless network, it can hardly meet the real-time requirements even in soft real-time industrial applications. Choosing appropriate load metrics to reflect the load information is the key to solving the load balancing problem in industrial WLAN. On the other hand, a good load distribution algorithm is also vital in the load balancing procedure. In this section, we present the design details of the metric check and load distribution algorithms of Det-LB.

A. SYSTEM MODEL

In a network-based load balancing approach, a centre network entity called a network controller is used to manage the load balancing of APs. The network controller could be an access point simply or an independent entity directly connected to a wired backbone. Each AP sends its information to the controller, and the controller knows the basic global condition of the network. The hierarchical network architecture of the approach is shown in Fig. 2.

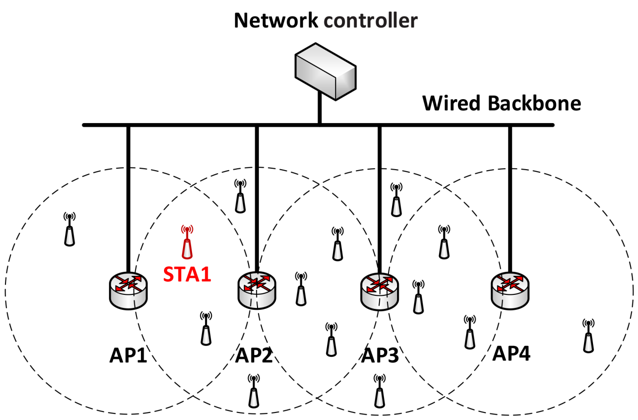


FIGURE 2. Typical hierarchical WLAN architecture.

As we mentioned before, the proposed load balancing algorithm is composed of a metric check and a load distribution process. Take the topology in Fig. 2 as an example, where there are more stations connected to AP2 and AP3 compared with the load situation of AP1 and AP4. Thus, the load of the network is relatively unbalanced if the traffic characteristics of each station are similar, and the network requires a specific load balancing algorithm. The metric check algorithm starts when a station (STA1) is aware of the potential unbalanced load situation. It checks the designated metrics and judges whether the current node is crowded or overloaded. If the metric is beyond a certain threshold, STA1 should decide to leave the current AP (AP2) and send a disassociation request to the controller; otherwise it stays still. If STA1 decides to leave, then the controller executes the load distribution algorithm and distributes the station to another AP in the overlapping area that has a lighter load, such as AP1. After the load distribution algorithm finishes, STA1 can get better performance as well as the stations that still associate with AP2. The detailed design of the two algorithms are described in Sec. III-B and Sec. III-C.

B. METRIC CHECK ALGORITHM

When deploying WLAN in an industrial scenario, the timeliness and reliability are two critical factors that should be considered. Thus, it is necessary to evaluate QoS performances in terms of the deadline miss ratio (here we define deadline miss as the packet delay being beyond the delay threshold predetermined by the underlying industrial applications, yet the packet is not lost) and packet loss for each station in a fixed interval, which are enough to show the deterministic feature in soft real-time applications. In addition, signal strength also needs to be considered when designing to ensure network performance.

In a metric check algorithm, a station should decide whether to keep associating or disassociate with an AP. The decision is based on whether the network performance, which is reflected by the load metrics, meets the specified QoS requirements. When the stations communicate with the APs, these basic load metrics are recorded locally. The proposed metric check algorithm begins to work at this time.

Let A_s denote the set of stations associated with AP s; then for a station i in A_s, we consider a fixed time interval t_i. The transmission packet number in this interval is denoted by n_i, and the set of transmission packets in this interval is denoted by N_i. The deadline miss ratio D_i, packet loss rate L_i and minimal signal strength S_i are denoted as follows:

$$D_i = \frac{N_d^i}{n_i}, i \in A_s \tag{1}$$

$$L_i = \frac{N_{loss}^i}{n_i}, i \in A_s \tag{2}$$

$$S_i = \min_{k \in N_i}\{s_k^i\}, i \in A_s \tag{3}$$

where N_d^i is the number of deadline miss packets in interval t_i, N_{loss}^i is the packet loss in t_i, and s_k^i is the received signal strength of packet k. Based on the three basic load metrics above, station i has the capability to assess the performance and decide whether to stay or disassociate with the current AP in interval t_i. First, we take the signal strength into consideration, if

$$S_i \leq S_{thresh}, \tag{4}$$

it proves that the signal strength of station i is weak or that station i is out of the communication range of the current AP; it needs to disassociate with the current AP, where S_{thresh} is the predetermined signal strength threshold. Otherwise, the signal strength is sufficient to maintain normal communication, and then station i should take D_i and L_i into account. The probability of disassociation for station i is shown as following:

$$P_{D_i} = \begin{cases} \alpha_{D_i}, & D_i \geq D_{thresh} \\ 0, & D_i < D_{thresh} \end{cases} \quad i \in A_s, \tag{5}$$

$$P_{L_i} = \begin{cases} \alpha_{L_i}, & L_i \geq L_{thresh} \\ 0, & L_i < L_{thresh} \end{cases} \quad i \in A_s, \tag{6}$$

where P_{D_i} and P_{L_i} are the disassociation probabilities fordeadline miss and packet loss respectively. D_{thresh} is the delay tolerance threshold, and L_{thresh} is the packet loss rate tolerance threshold. α_{D_i} and α_{L_i} are the predetermined possibilities to disassociate with the current AP, and the values of the possibilities are between zero and one. When the average delay D_i and packet loss rate L_i are below the threshold, the corresponding disassociation probabilities P_{D_i} and P_{L_i} are zero, which means station i remains associated with the AP. However, when D_i and L_i are beyond the threshold, the disassociation probabilities are α_{D_i} and α_{L_i} respectively. The reason why we adopt α_{D_i} and α_{L_i} which are between zero and one rather than absolute one is to avoid burst transmission over a short period of time changes to the association state of station i, and the values of the two can be adjusted according to the different real-time requirements in various industrial applications. Once station i decides to disassociate with the current AP, a disassociation frame will be sent to the network controller through the AP. We define P as the disassociation probability of station i, and the proposed metric check algorithm is shown in Algorithm 1.

Algorithm 1 Metric Check Algorithm

Input: D_i,L_i,S_i,D_{thresh},L_{thresh},S_{thresh}
Output: P

if $S_i < S_{thresh}$ **then**
  $P \leftarrow 1$
  return P
else
  $P \leftarrow 0$
end if
if $D_i \geq D_{thresh}$ **then**
  if $L_i \geq L_{thresh}$ **then**
    $P \leftarrow (\alpha_{D_i} + \alpha_{L_i} - \alpha_{D_i}\alpha_{L_i})$
    return P
  else
    $P \leftarrow \alpha_{D_i}$
    return P
  end if
else
  if $L_i \leq L_{thresh}$ **then**
    $P \leftarrow \alpha_{L_i}$
    return P
  else
    $P \leftarrow 0$
    return P
  end if
end if

C. LOAD DISTRIBUTION ALGORITHM

After a station is ready to disassociate with the AP that cannot meet its requirements, the controller should choose a better AP that has a lighter load and greater real-time performance for the station. To select the optimal AP for the station, we model the problem as a first-price sealed-bid auction based on game theory. The roles of different network entities in the auction model are as follows:

- Seller: the overloaded AP that cannot meet real-time requirements of some stations.
- Bidder: one of the surrounding APs that is in the communication range of the unsatisfied stations.
- Commodity: one of the stations that has already sent a disassociation frame to the overloaded AP.

The auction starts when station i decides to choose a better AP and sends a disassociation frame to the current AP. Let there be k lighter load APs ($M = \{AP_1, AP_2, ..., AP_k\}$); the bid of each bidder is b_k and v_k is the value of the commodity to each bidder. It is obvious that the value of v_k depends on the load size and real-time performance of each bidder. If a bidder has more spare communication resources, the commodity has a higher value to the bidder, and the bidder may intend to bid higher; eventually the bidder has more chances to win the auction and own the commodity. On the other hand, if a bidder is crowded or even overloaded, the commodity has less value to the bidder, and the bidder may intend to bid a lower price, and eventually the bidder has less chance to win the auction. Thus, the commodity value v_k can be defined as follows:

$$v_k = \lambda X_k \tag{7}$$

where λ is a correlation coefficient between 0 and 1, and X_k is a parameter that reflects the current performance of the bidders; the simplest setting method is to set the parameter to the load idleness ratio ρ_k of AP_k. According to the derivation from the first-price sealed-bid auction game theoretic model, it is an equilibrium bidding strategy for each bidder to bid as shown in Equ. 8.

$$b_k(v_k) = \frac{k-1}{k} v_k. \tag{8}$$

In addition, we set a threshold for the load idleness ratio ρ_{min} of AP_k to avoid overcrowded APs. If the load idleness ratio of AP_k is below ρ_{min}, AP_k will refuse the association request from the unsatisfied stations. Thus, the bid function for each bidder is

$$b_k(v_k) = \begin{cases} b_k(v_k) = \frac{k-1}{k} v_k, & \rho_{min} \leq X_k \leq 1 \\ 0, & 0 \leq X_k < \rho_{min} \end{cases} \tag{9}$$

Since all bidders give their bids ($B = \{b_1, b_2, ..., b_k\}$), the bidder who bids highest will win the auction and get the commodity according to the regulations in the first-price sealed-bid auction, and then the load distribution algorithm is finished. The detailed load distribution algorithm is shown in Algorithm 2.

Algorithm 2 Load Distribution Algorithm

Input: $X_1, X_2, ..., X_k, \rho_{min}$
Output: Auction Winner AP_a

for each $i \in [1, k]$ **do**
  if $X_i < \rho_{min}$ **then**
    $b_i \leftarrow 0$
  else
    $b_i \leftarrow \frac{k-1}{k} \lambda X_i$
  end if
end for
$a = \arg\max_{i \in [1,k]} \{b_1, b_2, ..., b_i\}$
return a

Another possible problem is the ping-pong effect between adjacent APs. Considering two adjacent APs that have relatively high load yet can just meet the predetermined deadline miss requirement, if a new station, whose load will make any of the two APs overload, connects to the network, a ping-pong effect will appear. The station will disconnect from the

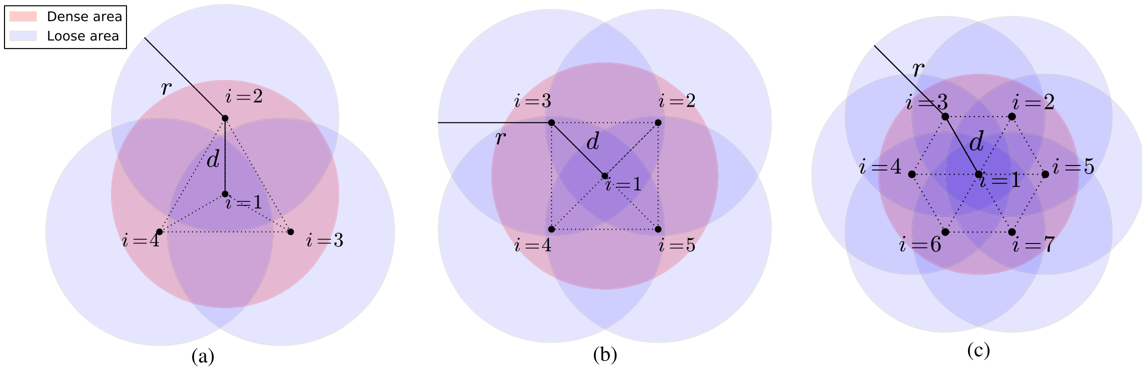


FIGURE 3. Example simulation topologies. (a) Simulation topology with 4 APs. (b) Simulation topology with 5 APs. (c) Simulation topology with 7 APs.

current AP and associate with the other repeatedly. To avoid this effect, stations are limited in their ability to connect to the same AP repeatedly if the load condition of the AP stays unchanged.

IV. PERFORMANCE EVALUATION

To demonstrate the advantage of the proposed load balancing algorithm, we present our simulation results and analysis in this section.

A. SYSTEM SETUP

The proposed scheme is implemented in a ns-3 simulator with WiFi modules. The simulation process is based on a multi-cell WLAN scenario, where one AP is located on the central dense area, and most stations are distributed in this area, while multiple other APs surround the central AP. The test topology examples are given in Fig. 3, and i stands for the index of APs, r stands for AP's coverage and d stands for the distance between the central AP and surrounding APs. In the simulation, we adopt a log-distance path loss model (Equ.10) described in [20]. The path loss exponent is set to 3.3, which is for the indoor environment. In addition, d is the distance between an AP and an STA.

$$PL(d) = 40 + 3.3 \times 10 \times \log d. \tag{10}$$

We can also derive the transmission range d_{max} of each AP from Equ.11

$$P_t - PL(d_{max}) = R_{thresh} \tag{11}$$

where P_t is the transmission power, $PL(d_{max})$ is the path loss when the distance of the AP and a station is at d_{max}, and R_{thresh} is the reception energy detection threshold. P_t and R_{thresh} are set to 16dbm and -96dbm according to [21] during the simulations respectively, and we can figure out the communication range of each AP (d_{max}) to be 151 m. In addition, we set the distance between the central AP and the surrounding APs to 100 m. The other main parameters of the simulated system are listed in Tab.1 in detail. To avoid co-channel interference among neighbouring APs, we assume different orthogonal channels are set. The implementation

TABLE 1. Main simulation parameters.

Transmission power	16dbm
Reception energy threshold	-96dbm
CCA threshold	-99dbm
Noise figure	7db
Channel bandwidth	20MHz
Path loss model	Log-distance
Path loss exponent	3.3
AP communication range	151m
Distance between central and peripheral APs	100m
UDP traffic period	10ms,20ms
packet payload size	250bytes,500bytes
D_{thresh} in Det-LB	5%
L_{thresh} in Det-LB	5%
R_{thresh} in Det-LB	-96dbm
ρ_{min} in Det-LB	0.1
λ (in Equ. (7))	0.8

of all the stations and APs is based on the IEEE 802.11b protocol, and the transmission rate of all the stations and APs is a step function of the SINR of the link following the IEEE 802.11b protocol.

Because the proposed load balancing algorithm is mainly intended for industrial soft real-time NCS, we adjust the wireless environment in the simulation, and make it similar to the wireless conditions in industrial scenarios as much as possible. In fact, high external interference exists in harsh and unstable industrial environments, which brings high bit error rates (BER = 10^{-2} - 10^{-6}) [22] in device communication, and the packet loss ratio is up to 10% when there is no harsh industrial interference (e.g., A working portal crane [23]) nearby. Thus, we set an interference source that introduces packet loss ratio at approximately 1%. In addition, periodic network traffic is another important feature for soft real-time NCS. To emulate the soft real-time NCS network traffic, stations are installed with two kinds of UDP applications that generate periodic traffic every 10 ms and 20 ms.

We also implement three existing load balancing algorithms as a comparison, namely, the RSSI-largest-first (RLF) scheme, the max-min fairness (MMF) method [7] and the dynamic load balancing approach (DLBA) proposed in [8]. The RLF scheme is the default association scheme in the

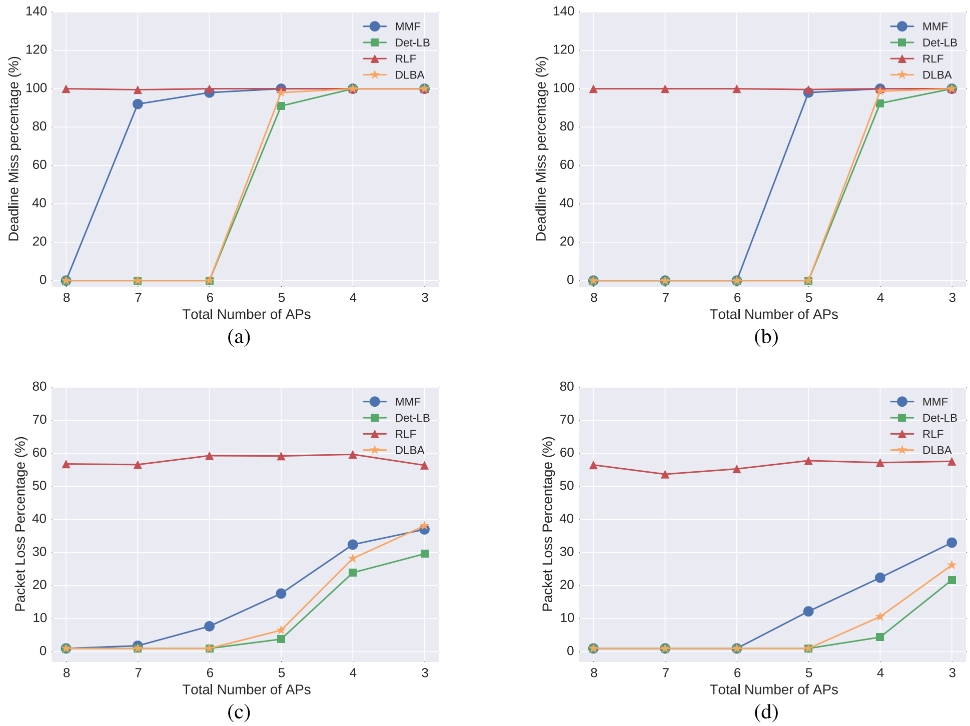


FIGURE 4. Simulation results when varying the number of AP. (a) Deadline miss ratio with 500B payload size. (b) Deadline miss ratio with 250B payload size. (c) Packet loss ratio with 500B payload size. (d) Packet loss ratio with 250B payload size.

IEEE 802.11 protocol, and stations associate with APs only according to the RSSI value. The basic idea of MMF is that a station selects APs not only based on RSSI values but also based on the number of stations already associated with an AP. It aims to achieve max-min fairness network-wide. DLBA is a distributed load balancing algorithm for industrial WLAN.

In the simulation, we mainly focus on the performance improvement that the proposed load balancing algorithm can bring for industrial WLAN. Thus, we choose the ratio of transmission deadline misses, which is set to 10 ms and 20 ms depending on the applications, and packet loss ratio to test the network performance, as the two metrics are the most significant in industrial scenario. To ensure the reliability of the simulation results, each simulation is conducted five times, and the results are the average of the five. The results are obtained from the worst 20% performance stations to make the results more distinct as the worst case is the most significant in industrial context.

B. SIMULATION RESULTS

To validate the characteristics of different algorithms more precisely, we implement a variety of simulation scenarios, including varying the number of APs with fixed number of stations and varying the number of stations with fixed number of APs.

1) VARYING THE NUMBER OF APS

In this scenario, 60 stations are deployed in the simulation, and 90% of the stations are located in the central dense area. The stations send periodic data to APs. The period of half the stations are set to 10 ms, and others are set to 20 ms. The number of APs varies from 3 to 8, and the packet payload size is fixed at 250 bytes or 500 bytes.

Fig. 4 shows the deadline miss ratio and packet loss ratio of different load balancing algorithms. We observe that the objective values associated with each algorithm increase as the number of APs decreases. This is intuitively expected because each AP is more loaded if the number of APs decreases, as the number of stations does not change. This causes worse network performance, and the deadline miss ratio and packet loss ratio will increase. Comparing Fig. 4(a) and Fig. 4(b), we observe the deadline miss ratio of each algorithm decreases when adopting a 250 byte payload size instead of 500 bytes. The reason is that the traffic load is lighter when adopting 250 bytes payload size as the traffic is periodic. A similar conclusion about packet loss can be drawn when comparing Fig. 4(c) and Fig. 4(d).

Fig. 4 further shows that the Det-LB algorithm has better deadline miss and packet loss performance than the other three algorithms, regardless of the payload size and the number of APs. We take Fig. 4(a) as an example. In the RLF scheme, stations tend to associate with APs that have the largest signal strength, and most of the stations in dense areas

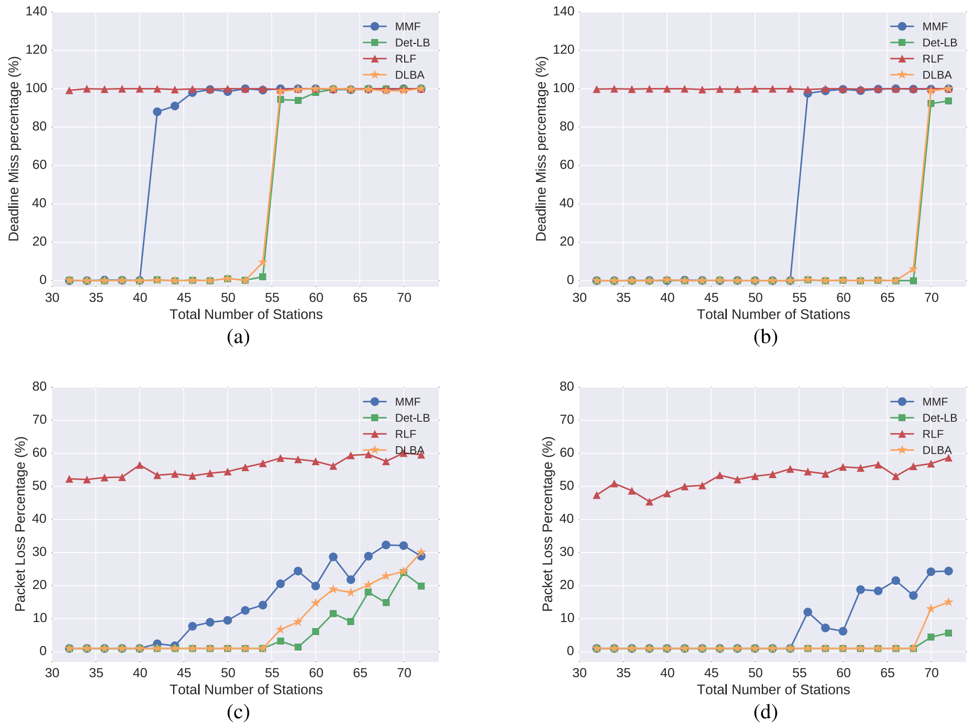
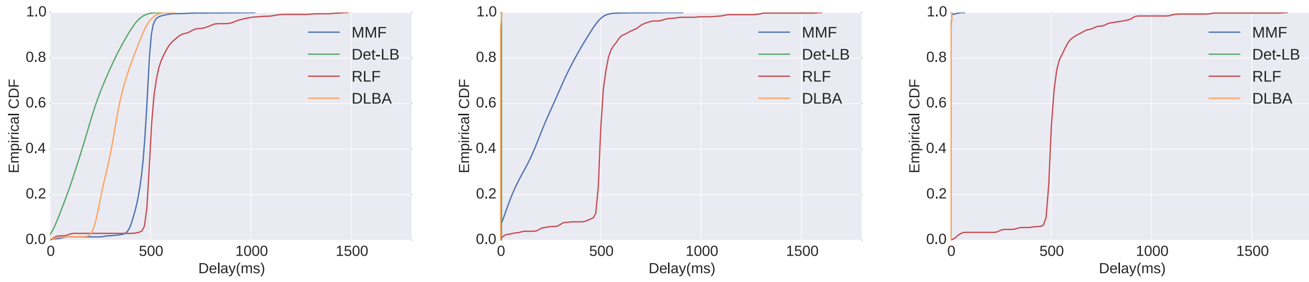


FIGURE 5. Simulation results when varying the number of stations. (a) Deadline miss ratio with 500B payload size. (b) Deadline miss ratio with 250B payload size. (c) Packet loss ratio with 500B payload size. (d) Packet loss ratio with 250B payload size.

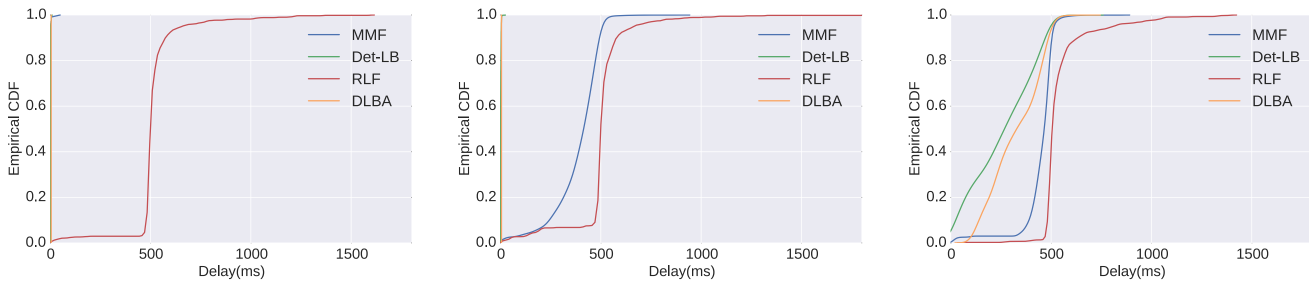
will associate with the central AP. The load is highly unevenly distributed in this situation. The central AP has a heavy traffic load while the surrounding APs are nearly idle. Thus, the network performance with the RLF load balancing algorithm is always terrible as the number of APs changes from 8 to 3. The MMF scheme uses the number of currently associated stations as the load reference; however, it does not consider the different applications installed on stations. Although the number of stations may be balanced, stations that are given different periodic applications (10 ms and 20 ms) might still be distributed unevenly. The worst case is if half the APs manage stations that are using 20 ms periodic applications, while the other half manage stations which are using 10 ms periodic applications. Thus, the deadline miss percentage increases to 92% in the condition of 7 APs, and it is worse when the number of APs decreases. In Det-LB and DLBA, the stations mainly consider the designated metrics (deadline miss and packet loss) to choose APs, and an AP that meets the requirement can always be found by stations only if the whole network is overcrowded. Therefore, the deadline miss percentage remains zero until the number of APs decreases to 5. However, DLBA works in a distributed way and does not take the ping-pong effect into consideration, which leads to worse performance when the network load is relatively high. In summary, Fig. 4(a) shows that the RLF scheme totally cannot meet the industrial requirement, the MMF scheme is able to support 60 stations with at least 8 APs, Det-LB and DLBA can support 60 stations with at least 6 APs under the given circumstance yet DLBA has approximately 5% worse performance when the network load is relatively high. In Fig. 4(b), MMF and Det-LB/DLBA can support 60 stations with at least 6 APs and 5 APs, while RLF still has almost 100% deadline miss percentage. Fig. 4(c) and Fig. 4(d) illustrate the packet loss ratio of three load balancing schemes. The results show that the packet loss performance of Det-LB is better than the other three in both the 500 and 250 byte payload size scenarios.

To indicate the transmission delay in detail, Fig. 6 shows the empirical cumulative distribution functions of transmission delay in 5 AP (Fig. 6(a)), 7 AP (Fig. 6(b)) and 8 AP(Fig. 6(c)) scenarios respectively. In Fig. 6(c), the transmission delay of all packets in MMF, Det-LB and DLBA are almost below 1 ms, while the delay in RLF is mainly distributed at approximately 500 ms, and the maximum delay is up to 1700 ms due to the congested load. In Fig. 6(b), packet delay in the RLF scheme, Det-LB and DLBA is nearly unchanged. However, the delay in MMF increases rapidly, and it is mainly distributed between 1 ms and 500 ms, because some of the APs are getting crowded or even overloaded, which leads to sharp declines in delay performance. In Fig. 6(a), the four load balancing strategies have varying degrees of delay distribution. Although all of them have bad performance on deadline miss number, Det-LB still has the best delay performance while RLF is still the worst.



(a) (b) (c)

FIGURE 6. ECDF of transmission delay when varying the number of APs. (a) 5 APs. (b) 7 APs. (c) 8 APs.



(a) (b) (c)

FIGURE 7. ECDF of transmission delay when varying the number of stations. (a) 34 stations. (b) 48 stations. (c) 60 stations.

2) VARYING THE NUMBER OF STATIONS

In this scenario, 5 APs are deployed in the simulation, and the number of station is increased gradually from 32 to 72 with a step length of 2, which can reflect the variation of objective values more accurately compared with the scenarios in Sec. IV-B.1. Other simulation parameters, including the traffic period and packet payload size, are similar to the parameters set in Sec. IV-B.1.

Fig. 5 shows the deadline miss ratio and the packet loss ratio of different load balancing algorithms. The results are similar to the results in Fig. 4, and Det-LB still has better deadline miss and packet loss performance than the other three algorithms. Fig. 5(a) shows in the 500 byte payload size scenario; the RLF scheme is at almost 100% deadline miss and cannot meet the requirements of industrial applications, while the MMF algorithm, DLBA and Det-LB can support up to 40, 52, 54 stations, respectively. Thus, the deadline miss performance of Det-LB is improved by 35% over the performance of MMF. In Fig. 5(b), we observe the RLF scheme still has a terrible deadline miss number, while up to 54 and 68 stations can be managed by APs in MMF and Det-LB/DLBA, respectively; the deadline miss performance of Det-LB has a boost of 26% compared with MMF in the 250 byte payload size scenario. Fig. 5(c) illustrates the packet loss ratio of the three algorithms in the 500 byte payload size scenario. The results show the packet loss ratio of RLF is approximately 50%, and the packet loss ratio of MMF, DLBA, Det-LB increases gradually as the number of stations increases. The results further show the point where the packet loss of Det-LB and MMF starts increasing is the same as the point where the deadline misses of Det-LB and MMF start increasing in Fig. 5(a) and Fig. 5(b). Thus, we conclude that Det-LB has a performance boost of 35% and 26% compared with MMF in the 500 and 250 byte payload size scenarios, respectively. When the APs tend to be overloaded, the Det-LB has around a 5% deadline miss performance superiority and better packet loss performance than DLBA.

Fig. 7 shows the empirical cumulative distribution functions of transmission delay in 34 AP (Fig. 7(a)), 48 AP (Fig. 7(b)) and 60 AP (Fig. 7(c)) scenarios. The results are similar to the results shown in Fig. 6, and we can still observe Det-LB has the best packet transmission delay performance in all conditions.

V. CONCLUSION

In this paper, we propose a deterministic load balancing algorithm (Det-LB) for WLAN, which is mainly intended for soft real-time control systems. To achieve network load balancing, two sub-algorithms are designed for the metric check and load distribution procedure. In the metric check algorithm, we choose the deadline miss ratio and packet loss ratio as the vital metrics for a station to decide whether to disassociate with the current AP or stay. On the other hand, in the load distribution procedure, we model the problem as a first-price sealed-bid auction in game theory. An optimal load distribution scheme is given after detailed analysis. To validate the performance of the proposed algorithm, we compare the real-time performance with different load balancing strategies in a simulation, including the RLF and the MMF load balancing algorithms. The simulations are conducted with various scenarios, and the results illustrate that Det-LB has better performance than the RLF and MMF schemes. In a scenario where 5 APs are deployed, Det-LB has 35% and 26% performance improvements over MMF in the 500 byte and 250 byte payload size conditions, respectively, and has a 5% deadline miss performance boost and better packet loss performance than DLBA when the APs tend to be overloaded.

REFERENCES

[1] M. Magno, D. Boyle, D. Brunelli, B. O'Flynn, E. Popovici, and L. Benini, "Extended wireless monitoring through intelligent hybrid energy supply," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 1871–1881, Apr. 2014.

[2] O. Kreibich, J. Neuzil, and R. Smid, "Quality-based multiple-sensor fusion in an industrial wireless sensor network for MCM," *IEEE Trans. Ind. Electron.*, vol. 61, no. 9, pp. 4903–4911, Sep. 2014.

[3] G. Tian, S. Camtepe, and Y.-C. Tian, "A deadline-constrained 802.11 MAC protocol with QoS differentiation for soft real-time control," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 544–554, Apr. 2016.

[4] F. Tramarin, S. Vitturi, M. Luvisotto, and A. Zanella, "On the use of IEEE 802.11n for industrial communications," *IEEE Trans. Ind. Inf.*, vol. 12, no. 5, pp. 1877–1886, Oct. 2016.

[5] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless. Commun.*, vol. 21, no. 2, pp. 18–25, Apr. 2014.

[6] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Trans. Netw.*, vol. 23, no. 3, pp. 836–850, Jun. 2014.

[7] Y. Bejerano, S. J. Han, and L. Li, "Fairness and load balancing in wireless LANs using association control," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 560–573, Jun. 2007.

[8] M. Collotta, G. Pau, V. M. Salerno, and G. Scatà, "A distributed load balancing approach for industrial IEEE 802.11 wireless networks," in *Proc. IEEE Int. Conf. Emerg. Technol. Factory Autom.*, Sep. 2012, pp. 1–7.

[9] F. Barać, M. Gidlund, and T. Zhang, "Ubiquitous, yet deceptive: Hardware-based channel metrics on interfered WSN links," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, pp. 1766–1778, May 2015.

[10] A. Balachandran, P. Bahl, and G. M. Voelker, "Hot-spot congestion relief in public-area wireless networks," in *Proc. 4th IEEE Workshop Mobile Comput. Syst. Appl.*, Jun. 2002, pp. 70–80.

[11] M. Collotta, "Flba: A fuzzy algorithm for load balancing in IEEE 802.11 networks," *J. Netw. Comput. Appl.*, vol. 53, pp. 183–192, Jul. 2015.

[12] L. H. Yen, T. T. Yeh, and K. H. Chi, "Load balancing in IEEE 802.11 networks," *IEEE Internet Comput.*, vol. 13, no. 1, pp. 56–64, Jan. 2009.

[13] Y. Li, H. Ji, X. Li, and D. Cai, "A novel load balancing algorithm in IEEE 802.11 wireless lans with cell breathing," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2009, pp. 1–4.

[14] T. Inaba, S. Sakamoto, T. Oda, L. Barolli, and M. Ikeda, "A QoS-aware admission control system for WLAN using fuzzy logic," in *Proc. Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Mar. 2016, pp. 499–505.

[15] S. Kim, S. Choi, and B. G. Lee, "A joint algorithm for base station operation and user association in heterogeneous networks," *IEEE Commun. Lett.*, vol. 17, no. 8, pp. 1552–1555, Aug. 2013.

[16] W. Xu, C. Hua, and A. Huang, "A game theoretical approach for load balancing user association in 802.11 wireless networks," in *Proc. Global Commun. Conf. (GLOBECOM)*, Miami, FL, USA, Dec. 2010, pp. 1–5.

[17] Y. Zhang, C. Lee, D. Niyato, and P. Wang, "Auction approaches for resource allocation in wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1020–1041, 3rd Quart., 2013.

[18] C. A. Gizelis and D. D. Vergados, "A survey of pricing schemes in wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 1, pp. 126–145, 1st Quart., 2011.

[19] S. Parsons, J. A. Rodriguez-Aguilar, and M. Klein, "Auctions and bidding: A guide for computer scientists," *ACM Comput. Surv.*, vol. 43, no. 2, pp. 1–59, 2011.

[20] T. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[21] T. Henderson and G. Pei. (2009). *Validation of NS-3 802.11b PHY Model*. [Online]. Available: https://www.nsnam.org/~pei/80211b.pdf

[22] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches," *IEEE Trans. Ind. Electron.*, vol. 56, no. 10, pp. 4258–4265, Oct. 2009.

[23] A. Willig, M. Kubisch, C. Hoene, and A. Wolisz, "Measurements of a wireless link in an industrial environment using an IEEE 802.11-compliant physical layer," *IEEE Trans. Ind. Electron.*, vol. 49, no. 6, pp. 1265–1282, Dec. 2002.

YUJUN CHENG received the B.S. degree in communication engineering from Beijing Jiaotong University, Beijing, China, in 2014, where he is currently pursuing the Ph.D. degree in communications and information science. His specific areas of research interest mainly focus on wireless medium access control protocols, industrial wireless networks, and real-time computing.

DONG YANG (M'11) received the B.S. degree from Central South University, Changsha, China, in 2003, and the Ph.D. degree in communications and information science from Beijing Jiaotong University, Beijing, China, in 2009. From 2009 to 2010, he was a Post-Doctoral Research Associate with Jönköping University, Jönköping, Sweden. In 2010, he joined the School of Electronic and Information Engineering, Beijing Jiaotong University. He is currently a Professor, His research interests are network technologies, including Internet architecture, industrial Internet, and wireless sensor networks.

HUACHUN ZHOU received the B.S. degree from the Peoples Police Officer University of China in 1986, and the M.S. degree in telecommunication automation and Ph.D. degree in telecommunications and information system from Beijing Jiaotong University (BJTU) in 1989 and 2008, respectively.

In 1994, he joined the Institute of Automation Systems, BJTU, where he is currently a Lecturer. From 1999 to 2009, he was a Senior Engineer with the School of Electronic and Information Engineering, BJTU, and with the Network Management Research Center, BJTU. Since 2009, he has been a Professor with the National Engineering Laboratory for Next Generation Internet Interconnection Devices, BJTU. He has authored over 40 peer-reviewed papers and he holds 17 patents. His main research interests are in the area of mobility management, mobile and secure computing, routing protocols, network management, and satellite network.