WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/wcm.2690

RESEARCH ARTICLE

A novel software-defined networking approach for vertical handoff in heterogeneous wireless networks

Li Qiang[1], Jie Li[2*], Yusheng Ji[3] and Changcheng Huang[4]

[1] Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan
[2] Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan
[3] Information Systems Architecture Research Division, National Institute of Informatics, Tokyo, Japan
[4] Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada

ABSTRACT

We propose a novel vertical handoff scheme with the support of the software-defined networking technique for heterogeneous wireless networks. The proposed scheme solves two important issues in vertical handoff: *network selection* and *handoff timing*. In this paper, the network selection is formulated as a 0-1 integer programming problem, which maximizes the sum of channel capacities that handoff users can obtain from their new access points. After the network selection process is finished, a user will wait for a time period. Only if the new access point is consistently more appropriate than the current access point during this time period, will the user transfer its inter-network connection to the new access point. Our proposed scheme ensures that a user will transfer to the most appropriate access point at the most appropriate time. Comprehensive simulation has been conducted. It is shown that the proposed scheme reduces the number of vertical handoffs, maximizes the total throughput, and user served ratio significantly. Copyright © 2016 John Wiley & Sons, Ltd.

KEYWORDS
vertical handoff; software-defined networking; integer programming problem; total throughput; user served ratio

***Correspondence**
Jie Li, Faculty of Engineering, Information and Systems, University of Tsukuba, Japan.
E-mail: lijie@cs.tsukuba.ac.jp

1. INTRODUCTION

Heterogeneous wireless networks integrate a variety of wireless techniques to provide ubiquitous services [1]. In heterogeneous wireless networks, users may need to transfer their inter-network connections from one access point to another. The transferring operation among different kinds of access points is called vertical handoff [2]. There are two important issues [3] needed to be solved in vertical handoff: network selection and handoff timing. The network selection issue is to select an access point, to which the inter-network connection should be transferred. The handoff timing issue is to determine when the inter-network connection transferring should be implemented. The emergence of software-defined networking (SDN) technique [4] makes it possible to solve these two issues of vertical handoff in a novel perspective. SDN is a new networking paradigm, which provides a global centralized control of access points. In this paper, we make use of this feature of SDN and study the vertical handoff problem for heterogeneous wireless networks.

Consider the scenario shown in Figure 1. There is a general heterogenous wireless network environment that consists of Wi-Fi, LTE, WiMAX and 3G. A user walks to the company from home. When the user is at home, his (or her) smart phone connects to the Wi-Fi in the house. After the user goes out of home, his (or her) smart phone may connect to LTE, or WiMAX of the public library, or 3G. There are three available networks. This user needs to know which one should be selected (network selection issue) [5]. After a new access point is selected, this user also needs to know when the internetwork connection should be transferred to the new access point (handoff timing issue) [6]. Solutions to the network selection and handoff timing issues compose the vertical handoff scheme [7].

Because of lack of the global view, most of existing vertical handoff schemes failed to be global optimal. The emergence of SDN [8] technique provides a chance to break this limitation. An SDN controller has an abstracted centralized control of access points [9]. We make use of this feature of SDN, and propose a novel vertical handoff scheme named software-defined networking based

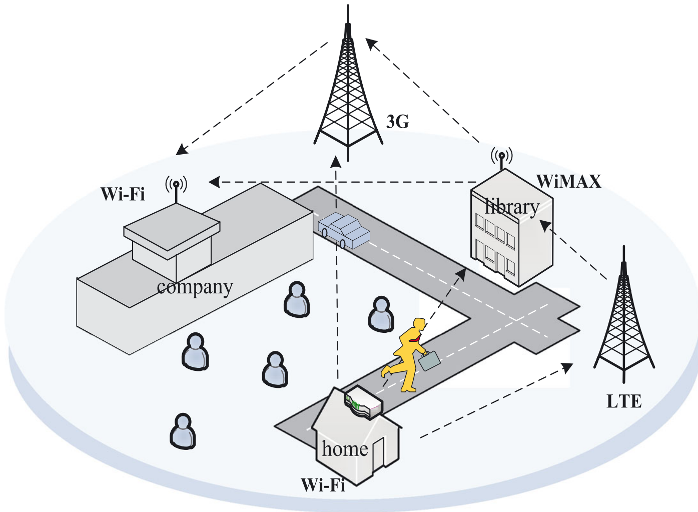


Figure 1. Illustration of vertical handoff.

vertical handoff (S-DNVH). In S-DNVH, users are divided into two classes: non-handoff users and handoff users. Non-handoff users will stay in the connections with their current access points. While handoff users will send their handoff requests to an SDN controller. The network selection is formulated as a 0-1 integer programming problem [10] by the SDN controller, with the objective of maximizing the sum of channel capacities that handoff users can obtain from their new access points. After the network selection process is finished, handoff users have to wait for a certain time period [11]. After the time period, the SDN controller evaluates the performances of current access points and new access points for users. If the new access points are consistently more appropriate than the current access points during the time period, users will transfer their inter-network connections to the new access points. The main contributions of this paper are summarized as follows:

- To the best of our knowledge, we are the *first* to apply the SDN technique in the study of vertical handoff problem for heterogeneous wireless networks. We investigate the architecture of SDN, and design a compatible vertical handoff procedure.
- We formulate the network selection issue as a 0-1 integer programming problem based on our previous work [12] and propose a network selection algorithm to solve it. In the proposed network selection algorithm, an SDN controller will allocate the most appropriate access point for each user.
- We propose a handoff timing algorithm to determine the time when the network selection results should be implemented. Based on limited information and simple calculation, the proposed handoff timing algorithm can predict the movement directions of users. Only if users are certain to move away from their current access points, will the network selection results be implemented.
- Through comprehensive experiments, we validate that our proposed scheme significantly reduces the number of vertical handoffs, maximizes the total throughput and user served ratio.

The rest of this paper is organized as follows. Section 2 is the related work. Section 3 is the system description and problem formulation. Sections 4 and 5 present the proposed scheme for network selection and handoff timing issues, respectively. Section 6 is the performance evaluation. Section 7 concludes this paper.

2. RELATED WORK

Here, we introduce the most typical related works in this section. Wang *et al.* [11] proposed an interesting policy-enabled vertical handoff scheme. The network selection is made by users. Periodically, users collect current network conditions as the inputs of their policy modules. The policy module makes trade-offs among network conditions, cost, power consumptions, and other criteria, then determines the “best" networks. In order to avoid the handoff instability problem, a random handoff timing mechanism is further proposed. After waiting for a stability period, users hand off to those networks, which have sustained high-performance during the stability period.

A fuzzy logic-based network selection scheme is proposed by J. Hou *et al.* [13], in which three input fuzzy variables that include the probability of a short interruption, the failure probability of handover to radio, and the size of unsent messages are considered. At first, these three fuzzy variables are fuzzified and converted into some input fuzzy sets by a singleton fuzzifier. Then the input fuzzy sets are mapped into output fuzzy sets by an algebraic product operation. Finally, the output fuzzy sets are defuzzified into a crisp decision point, which indicates a network selection result.

Lee *et al.* [14] carried out research from the aspect of optimization, and proposed an enhanced group handoff scheme. In enhanced group handoff scheme, each user evaluates its available access points on the remaining bandwidth. The network selection is formulated as a convex optimization problem. The objective of formulated problem is to minimize the handoff blocking probability. By using the Karush–Kuhn–Tucker condition, the formulated convex optimization problem is solved and a new access point can be determined. After the new access point is determined, a user transfers its inter-network connection to the new access point after an adjusted delay.

An interesting scheme proposed by Ciubotaru *et al.* [15] is called smooth adaptive soft handoff algorithm (SASHA). In SASHA, a user obtains a weighted sum of various performance parameters, and calculates the quality of service (QoS) values of its available access points. The user allocates its traffic according to the QoS values. As the user leaves an access point and gets closer to another, the QoS value of the leaving access point gets lower, and the QoS value of the approaching access point gets higher. As a result, traffic on the leaving access point is transferred to the approaching access point gradually.

3. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

3.1. Network architecture

For the upcoming problem formulation, we first exploit the special network architecture of SDN. Traditionally, each access point contains both a control plane and a data plane [8]. The control plane decides whether a traffic flow is admissible or not, and the route that the traffic flow should traverse. The data plane forwards the traffic flow according to the decision made by the control plane. In the SDN architecture, an SDN controller separates control planes from data planes of access points and provides a centralized control of these access points. The SDN controller communicates with access points via OpenFlow [16] and has a global view of the network environment. This feature of SDN gives us an opportunity to design a global optimal vertical handoff scheme.

Normally, an SDN controller can manage thousands of access points at the same time. When there are a large number of access points in an area, several SDN controllers can be deployed. Besides of this, in order to group manage the access points, multiple SDN controllers are also required so that each SDN controller can install a custom policy in its own domain. Stallings [8] pointed out the most common scenario is the numerous and nonoverlapping SDN domains scenario as shown in Figure 2. As we know, an SDN controller centrally controls the access points in its domain [17]. In fact, logical centralized control also exists in the multiple SDN controllers scenario. The Internet Engineering Task Force is currently working on the SDNi protocol [18], which helps the SDN controllers to exchange information and gain the global information. Thus, if an optimal vertical handoff scheme is proposed for the single SDN controller scenario, the proposed scheme also works for the multiple SDN controllers scenario. The extension work is just to add an “information exchange" stage at the beginning of proposed scheme. In this paper, we will consider the scenario with single SDN controller.

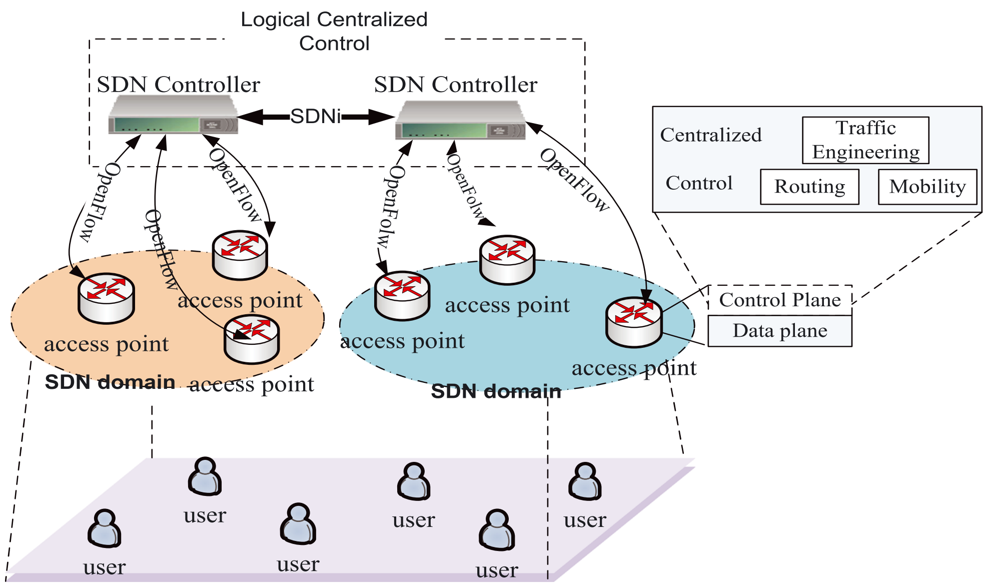


Figure 2. A network architecture of software-defined networking.

3.2. Problem formulation

In this subsection, we formulate the vertical handoff problem for heterogenous wireless network environment by using the SDN technique [19]. Specifically, we consider a heterogenous wireless network environment that consists of m access points. Let $\mathcal{A}$ be the set of access points, $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$. These access points support different wireless technologies. With the support of the media-independent handover (MIH) standard [20], all of these access points can be centrally controlled by a single SDN controller. Access point a_i ($a_i \in \mathcal{A}$, $i = 1, 2, \ldots, m$) has l_i channels, a_i equally divides its frequency band among these channels. The bandwidth of each channel in access point a_i is denoted by b_i in MHz. If the access point a_i is connected by a user, this user will occupy one channel of a_i [14]. Hence, the access point a_i can serve at most l_i users simultaneously.

Consider that there are n users. Let $\mathcal{U}$ be the set of users, $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$. If user u_j ($u_j \in \mathcal{U}$, $j = 1, 2, \ldots, n$) is inside the coverage area of an access point, this access point is called an *available access point* of user u_j. In heterogenous wireless network environment, u_j may have several available access points. However, u_j can connect to at most one of its available access points at anytime. The connected available access point is called the *current access point* of user u_j. An adjacency matrix $\delta(t)$ is used to reflect the relationship between access points and users at time t as follows.

$$\delta(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \vdots \\ \boldsymbol{a_m} \end{array} \begin{array}{c} \begin{array}{cccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{array} \\ \begin{bmatrix} \delta_{11}(t) & \delta_{12}(t) & \ldots & \delta_{1n}(t) \\ \delta_{21}(t) & \delta_{22}(t) & \ldots & \delta_{2n}(t) \\ \vdots & & \ddots & \vdots \\ \delta_{m1}(t) & \delta_{m2}(t) & \ldots & \delta_{mn}(t) \end{bmatrix} \end{array}$$

where

$$\delta_{ij}(t) = \begin{cases} 1, & a_i \text{ is connected by } u_j \text{ at time } t, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Because most of the applications (e.g., music, video, and game) require higher download and lower upload rates, users are more concerned about the quality of downlink (from access point a_i to user u_j). All links refer to downlink unless otherwise specified from now on. For convenience, the notations used in this paper are summarized in Table I.

For each access point-user pair (a_i, u_j), assume that the received signal power of user u_j from available access point a_i at time t is $s_{ij}(t)$ in watts. Let $d_{ij}(t)$ denote the Euclidean distance between access point a_i and user u_j at time t. Let p_i denote the transmission power of access point a_i in watts. Then the value of $s_{ij}(t)$ can be calculated as follows [21].

$$s_{ij}(t) = p_i \cdot h_{ij} \cdot d_{ij}(t)^{-\alpha} \tag{2}$$

Table I. Notation summary.

m	Number of access points
$\mathcal{A}$	Set of access points $\{a_i\}$, $i = 1, 2, \ldots, m$
$\mathcal{A}'_j$	Set of available access points for user u_j, $\mathcal{A}'_j \subseteq \mathcal{A}$
$\mathcal{X}_j$	Set of access points which transmit signal by using the same frequency band as user u_j, $\mathcal{X}_j \subseteq \mathcal{A}$
l_i	Number of channels in access point a_i
b_i	Bandwidth of each channel in access point a_i in MHz
n	Number of users
$\mathcal{U}$	Set of users $\{u_j\}$, $j = 1, 2, \ldots, n$
$\delta(t)$	Adjacency matrix of $\mathcal{A}$ and $\mathcal{U}$ at time t, that is, $[\delta_{ij}(t)]$
γ_j	Basic bandwidth requirement of user u_j in Mbps
$s_{ij}(t)$	Received signal power of user u_j from access point a_i at time t in watts
$d_{ij}(t)$	Euclidean distance between access point a_i and user u_j at time t
p_i	Transmission power of access point a_i for each channel in watts
h_{ij}	Channel fading gain of channel (a_i, u_j)
α	Pass loss exponent
$n_{ij}(t)$	Additive white Gaussian noise power of channel (a_i, u_j) at time t in watts
$q_{ij}(t)$	Channel capacity of channel (a_i, u_j) at time t in Mbps
o_a	Control overhead of the authentication phase in Mbps
o_r	Control overhead of the reassociation phase in Mbps
$\mathcal{V}(t)$	Identifier vector of user set $\mathcal{U}$ at time t, that is, $(v_j(t))$, $j = 1, 2, \ldots, n$
$\mathcal{R}_j(t)$	Channel capacities provided for user u_j at time t, that is, $(r_{ij}(t))$, $i = 1, 2, \ldots, m$
$\mathcal{R}(t)$	Channel capacities provided for user set $\mathcal{U}$, that is, $\left[[\mathcal{R}_j(t)]^{\mathrm{T}}\right]$, $j = 1, 2, \ldots, n$
$\mathcal{F}_j(t)$	Network selection result of user u_j at time t, that is, $(f_{ij}(t))$, $i = 1, 2, \ldots, m$
$\mathcal{F}(t)$	Network selection results of user set $\mathcal{U}$, that is, $\left[[\mathcal{F}_j(t)]^{\mathrm{T}}\right]$, $j = 1, 2, \ldots, n$

where the channel fading gain h_{ij} follows an exponential distribution with rate μ ($h_{ij} \sim exp(\mu)$), and the pass loss exponent $\alpha > 2$ (varies depending on the channel conditions).

With the help of some techniques such as orthogonal frequency-division multiplexing (commonly applied to LTE, WiMAX and IEEE 802.11 b/g supported devices) [22] and code division multiple access (commonly applied to 3G devices), the interference among users that belong to the same access point can be controlled in a low level. In order to further weaken this kind of interference, we assume that the orthogonal codes are used in code division multiple access.

For the interference from other access points, let $g_{xj}(t)$ in watts be the interference caused by access point a_x ($a_x \in \mathcal{A}, x \neq i$) to user u_j at time t, where a_x transmits signal by using the same frequency band as user u_j. Let $d_{xj}(t)$ denote the Euclidean distance between access point a_x and user u_j at time t. Let p_x and h_{xj} denote the transmission power of access point a_x in watts and channel fading gain, respectively. Then the value of $g_{xj}(t)$ can be calculated as follows:

$$g_{xj}(t) = p_x \cdot h_{xj} \cdot d_{xj}(t)^{-\alpha} \tag{3}$$

Because the transmission powers of users are relative small, and the locations of users erratically change all the time, the interference from other users is usually negligible [23]. If the access point a_i is connected by a user,this user will occupy one channel of a_i. Hence, the access point a_i can serve at most l_i users simultaneously. Note that the bandwidth of each channel in access point a_i is b_i MHz. Let $n_{ij}(t)$ denote the additive white Gaussian noise [24] power of the channel (a_i, u_j) at time t in watts. Let $\mathcal{X}_j$ ($\mathcal{X}_j \subseteq \mathcal{A}$) be the set of access points that transmit signal by using the same frequency band as the user u_j. According to the Shannon equation [25], the channel capacity denoted by $q_{ij}(t)$ in Mbps is calculated as follows.

$$q_{ij}(t) = b_i \cdot \log_2 \left[1 + \frac{s_{ij}(t)}{\sum_{a_x \in \mathcal{X}_j} g_{xj}(t) + n_{ij}(t)} \right] \tag{4}$$

Assume that all of the transmitted data can be received correctly. Consider that a user may carry out multiple tasks at the same time, there are infinite data needed to be received by each user. Thus, we can make an assumption that the data receiving rates of users are equal to the channel capacities. Furthermore, in order to guarantee the quality of experience, user u_j has the basic bandwidth requirement denoted by γ_j in Mbps. Suppose that the current access point of user u_j is a_c. If the channel capacity of current access point cannot meet the corresponding basic bandwidth requirement ($q_{cj}(t) < \gamma_j$), user u_j will perform the vertical handoff. We call these users who need to perform handoff *handoff users*. If the channel capacity of current access point can satisfy the basic bandwidth requirement ($q_{cj}(t) \geq \gamma_j$), user u_j will stay in the connection with its current access point a_c. We call these users who do not need to performed handoff *non-handoff users*. A vector $\mathcal{V}(t) = (v_1(t), v_2(t), \ldots, v_n(t))$ is used to identify the kinds of users at time t. The value of $v_j(t)$ is given as follows, where $j = 1, 2, \ldots, n$.

$$v_j(t) = \begin{cases} 0, & \text{user } u_j \text{ is a handoff user at time } t, \\ 1, & \text{user } u_j \text{ is a non-handoff user at time } t \end{cases} \tag{5}$$

Let $r_{ij}(t)$ denote the channel capacity that access point a_i can provide for user u_j at time t, if a_i is selected as the *new access point*. The value of $r_{ij}(t)$ is given in Equation (6). If u_j is a handoff user ($v_j(t) = 0$), $r_{ij}(t)$ is equal to $q_{ij}(t)$. If u_j is a non-handoff user ($v_j(t) = 1$), $r_{ij}(t)$ is 0.

$$r_{ij}(t) = \begin{cases} q_{ij}(t), & v_j(t) = 0, \\ 0, & v_j(t) = 1 \end{cases} \tag{6}$$

Let $\mathcal{R}_j(t) = \left(r_{1j}(t), r_{2j}(t), \ldots, r_{mj}(t)\right)$. Based on $\mathcal{R}_j(t)$, the SDN controller selects a *new* access point for u_j. The network selection result of u_j at time t is denoted by $\mathcal{F}_j(t) = \left(f_{1j}(t), f_{2j}(t), \ldots, f_{mj}(t)\right)$. The value of $f_{ij}(t)$ is given as follows, where $i = 1, 2, \ldots, m$.

$$f_{ij}(t) = \begin{cases} 1, & a_i \text{ is selected by } u_j \text{ at time } t, \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

The channel capacity that handoff user u_j can obtain from the **new** access point is $\mathcal{F}_j(t) \cdot \left[\mathcal{R}_j(t)\right]^{\mathrm{T}}$ in Mbps, where $\left[\mathcal{R}_j(t)\right]^{\mathrm{T}}$ is the transposition of $\mathcal{R}_j(t)$.

$$\mathcal{F}_j(t) \cdot \left[\mathcal{R}_j(t)\right]^{\mathrm{T}} = \sum_{i=1}^{m} \left[f_{ij}(t) \cdot r_{ij}(t)\right] \tag{8}$$

The SDN controller has to ensure that if an access point a_i is selected as the new access point of user u_j, the channel capacity provided by a_i should satisfy the basic bandwidth requirement of u_j. That is, $\mathcal{F}_j(t) \cdot \left[\mathcal{R}_j(t)\right]^{\mathrm{T}}$ should subject to the following constrain.

$$\mathcal{F}_j(t) \cdot \left[\mathcal{R}_j(t)\right]^{\mathrm{T}} \geq \gamma_j \tag{9}$$

If there is no access point can satisfy the basic bandwidth requirement of user u_j, the vertical handoff of u_j is failed. Moreover, u_j will be discarded by its current access point.

Note that if user u_j is a non-handoff user ($v_j(t) = 1$), it does not have any new access point ($\sum_{i=1}^{m} f_{ij}(t) = 0$). Thus, we can get a constraint for non-handoff users that is $\sum_{i=1}^{m} f_{ij}(t) + v_j(t) = 1$. If user u_j is a handoff user ($v_j(t) = 0$), when the vertical handoff of u_j is failed, u_j will not have any new access point neither ($\sum_{i=1}^{m} f_{ij}(t) = 0$). In this case, we can obtain a relationship that is $\sum_{i=1}^{m} f_{ij}(t) + v_j(t) = 0$. If user u_j is a handoff user ($v_j(t) = 0$) and the vertical handoff of u_j is successful, the SDN controller will allocate a new access point to u_j ($\sum_{i=1}^{m} f_{ij}(t) = 1$). In this case, we can obtain a relationship, that is, $\sum_{i=1}^{m} f_{ij}(t) + v_j(t) = 1$. In summary, the network selection result of user u_j should satisfy the following constraint.

$$\sum_{i=1}^{m} f_{ij}(t) + v_j(t) \leq 1 \tag{10}$$

Given a set of users $\mathcal{U}$, their network selection results are denoted by $\mathcal{F}(t) = [[\mathcal{F}_1(t)]^{\mathrm{T}}, [\mathcal{F}_2(t)]^{\mathrm{T}}, \ldots, [\mathcal{F}_n(t)]^{\mathrm{T}}]$. Note that access point a_i has l_i channels, because each user will occupy one channel, a_i can serve at most l_i users simultaneously. There are $\sum_{j=1}^{n} \left[v_j(t) \cdot \delta_{ij}(t)\right]$ non-handoff users are connecting to access point a_i at time t. Therefore, the number of handoff users assigned to a_i ($\sum_{j=1}^{n} f_{ij}(t)$) should satisfy the following constraint.

$$\sum_{j=1}^{n} \left[v_j(t) \cdot \delta_{ij}(t)\right] + \sum_{j=1}^{n} f_{ij}(t) \leq l_i \tag{11}$$

We express $\mathbb{T}\,(\mathcal{F}(t))$ as the sum of the channel capacities that handoff users can obtain from their new access points. $\mathbb{T}\,(\mathcal{F}(t))$ is then calculated as follows.

$$\mathbb{T}\,(\mathcal{F}(t)) = \sum_{j=1}^{n} \left\{\mathcal{F}_j(t) \cdot \left[\mathcal{R}_j(t)\right]^{\mathrm{T}}\right\} \tag{12}$$

The goal in the paper is to maximize the sum of channel capacities that handoff users can obtain from their new access points $\mathbb{T}\,(\mathcal{F}(t))$. Non-handoff users will stay in the connections with their current access points. The optimization problem of maximizing $\mathbb{T}\,(\mathcal{F}(t))$ is theoretically formulated as follows.

$$\text{Maximize } \mathbb{T}\,(\mathcal{F}(t)) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[f_{ij}(t) \cdot r_{ij}(t)\right] \tag{13}$$

subject to

$$\sum_{i=1}^{m} f_{ij}(t) + v_j(t) \leq 1, j = 1, 2, \ldots, n, \tag{14a}$$

$$\sum_{i=1}^{m} \left[f_{ij}(t) \cdot r_{ij}(t)\right] \geq \gamma_j, j = 1, 2, \ldots, n, \tag{14b}$$

$$\sum_{j=1}^{n} \left[v_j(t) \cdot \delta_{ij}(t)\right] + \sum_{j=1}^{n} f_{ij}(t) \leq l_i, i = 1, 2, \ldots, m \tag{14c}$$

The first constraint Equation (14a) indicates that each user has at most one new access point. The second constraint Equation (14b) guarantees that the channel capacity provided by the new access point can satisfy the basic bandwidth requirement of a handoff user. The last constraint Equation (14c) ensures that the number of users connected to an access point is smaller than the number of channels in this access point.

The vertical handoff problem is formulated as a 0-1 integer programming problem. Although the formulated problem is an NP-hard problem, the computation complexity is acceptable in its particular application context. The latest OPENFLOW version 1.4 supported access points provide 40 GbE services [26], and the computing capability of SDN controller is considered to be infinite [27]. Compared with the powerful SDN devices, most of vertical handoff scenarios involve limited number of users. Thus, we think that the network selection process is completed in a very short time interval which can be neglected. Moreover, if we eliminate non-handoff users from consideration, the computation complexity can be further reduced.

4. NETWORK SELECTION ALGORITHM

Based on previous formulations, we study the network selection issue of the vertical handoff in this section. We assume that users are selfish, they will select access points in the always-best-connected (ABC) way [28] if allowed. That is, users always choose the access points which have the best performance as their new access points. As an example shown in Figure 3, there are three network access points (i.e., a_1, a_2, a_3). These access points support different wireless technologies. In the coverage area of three access points, there are four users, that is, u_1, u_2, u_3, u_4). At first, user u_1 was connecting to the access point a_1. When the channel capacity of (a_1, u_1) cannot meet its basic bandwidth requirement, u_1 performed the vertical handoff. Suppose that the performance of a_2 is better than a_3, thus u_1 will choose a_2 as its new access point. At the same time, other users (u_2, u_3 and u_4) may also choose a_2 as their new access points. Because the resources are limited, a_2 will be exhausted. While there is no user connects to a_3. System resources are unreasonably utilized.

In order to efficiently utilize the system resources and construct a good access point associating strategy for each user, we propose a network selection algorithm for vertical handoff. An SDN controller selects access points for users in three phases: initialization, request matrix construction,

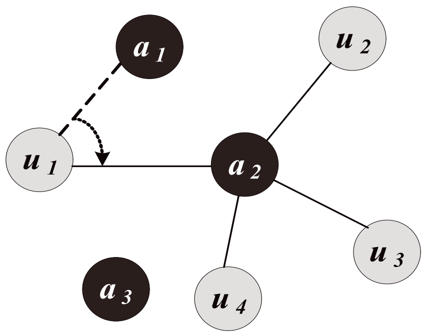


Figure 3. An example of the network selection issue.

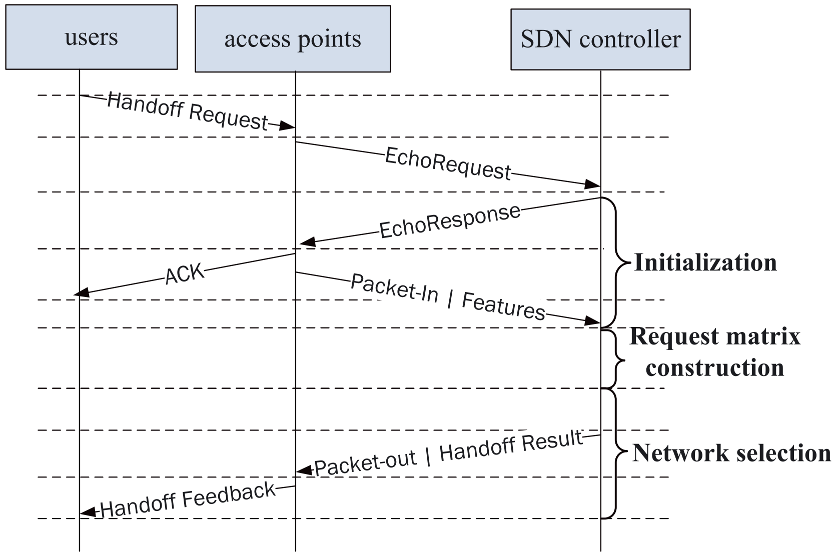


Figure 4. The procedure of network selection.

and network selection. The network selection procedure is illustrated as Figure 4 shows in which we consider the OPENFLOW 1.3.3.

4.1. Algorithm details

Phase 1. Initialization

At the beginning of each time slot, users evaluate their current access points and determine whether to perform vertical handoffs or not. If a user needs vertical handoff (handoff user), it will send handoff request frames to its available access points. The values of request frames are equal to the channel capacities. If a user does not need vertical handoff (non-handoff user), it also sends request frames to its available access points. In this case, the request frames are just like Hello messages, and their values are 0. After receiving the handoff requests, access points send EchoRequest frames to the SDN controller to initialize the network selection. Then the SDN controller will reply the EchoResponse Frames to these access points.

Phase 2. Request matrix construction

Access points put the features of users (i.e., the channel capacities) in the Data fields of Packet-In messages, and send the Packet-In messages to the SDN controller. These Packet-In messages will be converged in the SDN controller. According to these received messages, the SDN controller constructs a request matrix $\mathcal{R}(t) = \left[[\mathcal{R}_1(t)]^{\mathrm{T}},\ [\mathcal{R}_2(t)]^{\mathrm{T}},\ \ldots,\ [\mathcal{R}_n(t)]^{\mathrm{T}}\right]$. At first, the request matrix $\mathcal{R}(t)$ is incomplete because users may not be able to reach all access points. There are some elements, whose values are unknown. These elements are defined as the unassigned elements.

Definition 4.1 (*Unassigned element*). *The unassigned element is the element of a request matrix, whose value is unknown becaue of the corresponding access point and user cannot communicate directly.*

Because the SDN controller has a global view, it can complete the request matrix $\mathcal{R}(t)$ after a simple calculation in which the unassigned elements will be set to be 0.

Theorem 4.1. *The calculation rule of unassigned element maintains the consistency of vertical handoff request matrix and will not affect the network selection results.*

Proof. Each column of the request matrix $\mathcal{R}(t)$ corresponds to a vector $\mathcal{R}_j(t)$, where $j = 1, 2, \ldots, n$. For non-handoff users, all elements in $\mathcal{R}_j(t)$ are 0, and the value of corresponding column in $\mathcal{R}(t)$ will be set to be 0. During the network selection process, there is no new access point will be assigned to these users. For handoff users, the unassigned elements in $\mathcal{R}(t)$ that correspond to their unavailable access points are set to be 0. During the network selection process, some other access points with positive evaluations will be selected.

Phase 3. Network selection

After constructing the vertical handoff request matrix $\mathcal{R}(t)$, the SDN controller selects new access points for handoff users. The network selection is formulated as a 0-1 programming problem (Equation (13)). The SDN controller calculates the network selection results by solving this 0-1 programing problem. There are many tools that can be used by the SDN controller to solve the linear problems, such as LINGO and CVX on MATLAB. The network selection results are presented as a $m \times n$ matrix $\mathcal{F}(t)$. According to $\mathcal{F}(t)$, the SDN controller sends out Packet-Out messages. If the element $f_{ij}(t)$ is 1, that means a_i is the new access point of user u_j. Therefore, the SDN controller puts this result in the data filed of Packet-Out message, and sends the Packet-Out message [29] to access point a_i. Then, access point a_i sends a feedback frame to user u_j to notify this result.

4.2. An example

In this section, we will use an example to explain the network selection algorithm in detail. Specifically, we consider a scenario shown in Figure 5. There are three access points (i.e., a_1, a_2, a_3). These access points support different wireless technologies. An SDN controller centralized controls these access points. In the coverage area of three access points, there are five users (i.e., u_1, u_2, u_3, u_4, u_5). The basic bandwidth requirements of these five users are 2 Mbps (i.e., $\gamma_1 = 2$ Mbps), 3 Mbps (i.e., $\gamma_2 = 3$ Mbps), 4 Mbps (i.e., $\gamma_3 = 4$ Mbps), 4.5 Mbps (i.e., $\gamma_4 = 4.5$ Mbps) and 4 Mbps (i.e., $\gamma_5 = 4$ Mbps), respectively. Suppose that the channel capacity of (a_1, u_1) is 4 Mbps, and u_1 is a non-handoff user. Meanwhile, other four users (u_2, u_3, u_4, and u_5) are handoff users.

Phase 1. Initialization

Because u_1 is a non-handoff user, it will send request frames $r_{11}(t)$ and $r_{21}(t)$ to the available access points a_1 and a_2, respectively. The values of $r_{11}(t)$ and $r_{21}(t)$ are 0. Meanwhile, other four users u_2, u_3, u_4, and u_5 are handoff users; they also have to send out request frames. Take u_2, for instance, a_1, a_2 and a_3 are available to u_2. Let $r_{12}(t)$ denote the handoff request frame sent from u_2 to a_1. In our example, $r_{12}(t)$ is 7, which means if u_2 selects a_1 as its new access point, the channel capacity provided by a_1 is 7 Mbps. Similarly, other users send the vertical handoff request frames to their available access points.

Phase 2. Request matrix construction

After receiving the Packet-In messages, the SDN controller constructs a request matrix $\mathcal{R}(t)$ as shown in Equation (15). At first, there are four unassigned elements (i.e., $r_{14}(t)$, $r_{15}(t)$, $r_{23}(t)$ and $r_{31}(t)$) in the matrix $\mathcal{R}(t)$. Take the element $r_{14}(t)$, for instance, because the access point a_1 is unavailable to user u_4, u_4 will not send a request frame to a_1. Therefore, the SDN controller cannot determine the value of $r_{14}(t)$ in the beginning.

$$\mathcal{R}(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} 0 & 7 & 5 & r_{14}(t) & r_{15}(t) \\ 0 & 1 & r_{23}(t) & 2 & 6 \\ r_{31}(t) & 5 & 3 & 4 & 2 \end{bmatrix} \end{array} \tag{15}$$

Based on this primary request matrix, the SDN controller calculates the values of unassigned elements. All of the unassigned elements are set to be 0, and the completed request matrix $\mathcal{R}(t)$ is as follows.

$$\mathcal{R}(t) = \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} 0 & 7 & 5 & \mathbf{0} & \mathbf{0} \\ 0 & 1 & \mathbf{0} & 2 & 6 \\ \mathbf{0} & 5 & 3 & 4 & 2 \end{bmatrix} \end{array} \tag{16}$$

Phase 3. Network selection

Based on the request matrix $\mathcal{R}(t)$, the SDN controller formulates the network selection as a 0-1 programming problem (Equation (13)). The solution of this 0-1 programming problem is the network selection result, which is presented as a matrix $\mathcal{F}(t)$. Assume that the number of channels in access points a_1, a_2, and a_3 are 2, 3, 4, respectively. After some calculations, the SDN controller can obtain a 3×5 matrix $\mathcal{F}(t)$ as follows.

$$\begin{aligned} \mathcal{F}(t) &= \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} f_{11}(t) & f_{12}(t) & f_{13}(t) & f_{14}(t) & f_{15}(t) \\ f_{21}(t) & f_{22}(t) & f_{23}(t) & f_{24}(t) & f_{25}(t) \\ f_{31}(t) & f_{32}(t) & f_{33}(t) & f_{34}(t) & f_{35}(t) \end{bmatrix} \end{array} \\ &= \begin{array}{c} \\ \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \boldsymbol{a_3} \end{array} \begin{array}{c} \begin{array}{ccccc} \boldsymbol{u_1} & \boldsymbol{u_2} & \boldsymbol{u_3} & \boldsymbol{u_4} & \boldsymbol{u_5} \end{array} \\ \begin{bmatrix} 0 & 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 0 & 0 \end{bmatrix} \end{array} \end{aligned} \tag{17}$$

As a non-handoff user, u_1 will not obtain any new access point. User u_1 stays in the connect with its current access

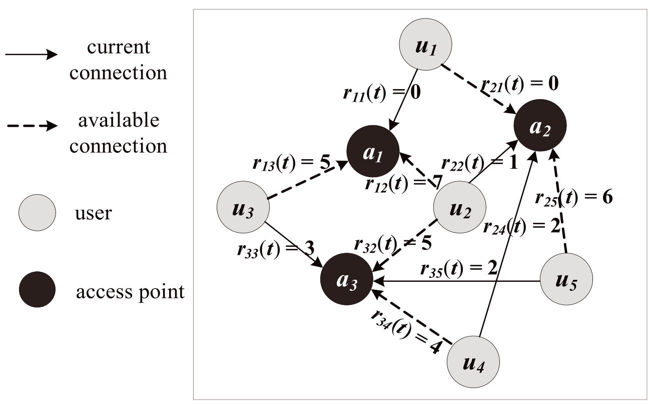


Figure 5. An example of the proposed network selection algorithm.

point a_1. Consequently, access point a_1 can only serve one more user. In order to maximize the sum of channel capacities that handoff users can obtain from their new access points, the SDN controller allocates access point a_1 to user u_3 instead of user u_2. For user u_4, because the channel capacities provided by its available access points a_2 and a_3 cannot satisfy the basic bandwidth requirements, the vertical handoff of u_4 is failed. User u_4 does not have any new access point and will be discarded by its current access point a_2. According to the network selection result $\mathcal{F}(t)$, the SDN controller sends out Packet-Out messages. The value of $f_{13}(t)$ is 1 means the new access point of user u_3 is a_1. That is, u_3 should transfer its inter-network connection from the current access point a_3 to a_1. Therefore, a_1 will send a feedback frame to u_3 to notify this selection result. Similarly, a_2 sends a feedback frame to u_5. a_3 sends a feedback frame to u_2.

5. HANDOFF TIMING ALGORITHM

Because users are always moving around in wireless environment, the network selection results should not be implemented immediately. For example as shown in Figure 6, user u_1 is moving back and forth. When user u_1 leaves the access point a_1 and closes to the access point a_2, its inter-network connection will be transferred from a_1 to a_2. When u_1 leaves a_2 and backs to a_1, the inter-network connection will be transferred from a_2 to a_1 again. The inter-network connection of user u_1 is switched between access points a_1 and a_2 times and times again [30]. This common example reveals that inappropriate handoff timing will incur numerous unnecessary handoffs.

Besides that, the ping-pong effect will also lead to some unnecessary handoffs. A user handoff from its current access point to the new access point in order to obtain higher channel capacities. However, the SDN controller optimizes the network selection results from the global perspective which inevitably sacrifices the individual interests. If the performance gain between the current access point and the new access points is less than the handoff overhead, the handoff is not a wise choice in this case.

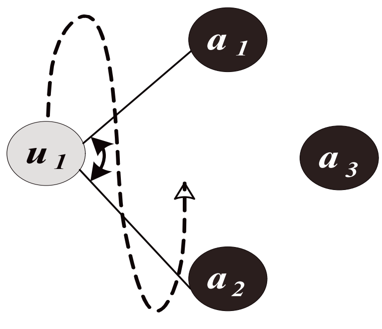


Figure 6. An example of the handoff timing issue.

5.1. Discussions

In our proposed scheme, users will wait for a stability period τ [11] after their network selection processes are finished. During the stability period, if the new access points are consistently more appropriate than their current access points, users will handoff to their new access points. Before we start introducing our proposed handoff timing algorithm, there are three things needed to be explained specially.

The first thing is about the meaning of *appropriate*. In this paper, we use"appropriate" instead of “better". The reason is even the new access point is better than the current one, if the current access point can satisfy the basic bandwidth requirement of a user, this user should not perform the vertical handoff.

The second thing is about the handoff overhead. When we are judging whether a new access point is appropriate or not, we need to take both the performance gain and the handoff overhead into account. According to the IEEE 802.11 standard handoff procedure [31], users will experience the authentication phase and the reassociation phase before transferring to their selected new access points. Anshul *et al.* [32] studied various scenarios and observed that the average authentication delay and the average reassociation delay are 1.3 ms and 2.3 ms, respectively. During the authentication phase, two authentication frames will be exchanged between the handoff user and the new access point. Normally, an Authentication frame contains 24 bytes header, at least 6 bytes body and 4 bytes frame check sequence (FCS) [33]. Therefore, the control frames will cost o_a Mbps bandwidth during the authentication phase as the following equation shows.

$$o_a = \frac{2 \cdot (3.4e-5)\ \text{Mb}}{1.3e-3\ \text{s}} \doteq 0.0052\ \text{Mbps} \qquad (18)$$

During the reassociation phase, the handoff user will send out an reassociation-request frame and the new access point will reply an reassociation-response frame. An reassociation-request frame consists of 24 bytes header, at least 10 bytes body and 4 bytes FCS. While an reassociation-response frame has 24 bytes header, at least 6 bytes body and 4 bytes FCS. Therefore, the control frames will cost o_r Mbps bandwidth during the Reassociation phase as the following equation shows.

$$o_r = \frac{(3.8e-5+3.4e-5)\ \text{Mb}}{2.3e-3\ \text{s}} \doteq 0.0031\ \text{Mbps} \qquad (19)$$

The last thing is about the length of a *stability period*. For a handoff user u_j, its available access points set is denoted by $\mathcal{A}'_j$, and the size of $\mathcal{A}'_j$ is $|\mathcal{A}'_j|$. Let τ_x be the length of stability period in the x th round, where $x = 1, 2, \ldots$. For an available access point a_i ($a_i \in \mathcal{A}'_j$), its channel capacity at time t is denoted by $q_{ij}(t)$. After waiting for a stability period τ_x, the channel capacity of a_i becomes $q_{ij}(t+\tau_x)$. Thus, the change rate of channel capacity during the stability period τ_x can be represented by $\frac{q_{ij}(t+\tau_x)}{q_{ij}(t)}$. We make use of the average change rate of channel capacities for all available access points to adjust the length of stability period. If u_j is suggested to wait for another stability period, the value of τ_{x+1} can be calculated as follows.

$$\tau_{x+1} = \frac{\sum_{a_i \in \mathcal{A}'_j} \dfrac{q_{ij}(t+\tau_x)}{q_{ij}(t)}}{\left|\mathcal{A}'_j\right|} \tau_x \tag{20}$$

Equation (20) indicates that if the average change rate of channel capacities during τ_x is bigger than 1 (i.e., $\tau_{x+1} > \tau_x$), that is the connection quality of user u_j is getting better, u_j will wait for a longer stability period in the next round. Otherwise, u_j will wait for a shorter stability period in the next round. H.J. Wang *et al*. [11] pointed out that the stability period is proportional to the handoff latency. Moreover, S. Sharma *et al*. [34] proved by experiment that the handoff latency is about 0.1 s. Thus, we accordingly set the initial value of stability period τ_1 to be 0.1 s in this paper.

5.2. Algorithm details

In this section, we take a handoff user u_i for instance to explain our proposed handoff timing algorithm. For the user u_1, suppose that its current access point is a_1, and its new access point which has been determined by the SDN controller is a_2 as shown in Figure 7. Following the previous definitions, the channel capacities of a_1 and a_2 at time t were $q_{11}(t)$ and $q_{21}(t)$ respectively. After waiting for a stability period τ, the channel capacities of a_1 and a_2 become $q_{11}(t+\tau)$ and $q_{21}(t+\tau)$ respectively. Based on the channel capacities, the Authentication overhead o_a and the Reassociation overhead o_r, the SDN controller makes a judgement and notifies u_i whether it should transfer the inter-network connection from a_1 to a_2 or not.

The network performance can be regarded as stable when considering the user mobility, which is a widely accepted assumption [35], [36]. Therefore, we are working under the assumptions as shown in Figure 7. The coverage area of a_1 is the circular region insides c_1. The closer to a_1, the higher channel capacity that u_1 can obtain from a_1. c_1, and c'_1 are concentric circles. If u_1 moves along c'_1, the channel capacity of (a_1, u_1) will not change [37]. Similarly, the coverage area of a_2 is the circular region insides c_2. If u_1 moves along c'_2, the channel capacity of (a_2, u_1) will not change. There is a line $\mathcal{L}$ goes through the intersections of c'_1 and c'_2. $\mathcal{L}$ is perpendicular to the line between a_1 and a_2. We consider the following two situations.

- If u_1 moves to the left of $\mathcal{L}$, which means u_1 has the tendency of moving closer to a_1. Because user was moving back to its current access point during the stability period, it does not need vertical handoff anymore, and the network selection result is cancelled.
- If u_1 moves to the right of $\mathcal{L}$, which means u_1 has the tendency of moving closer to a_2. Because user was moving away from its current access point during the stability period, it has to transfer the inter-network connection to the new access point at once.

Because the movement trend of user is important for the vertical handoff, some related work tried to predict the movement trend of a user. The existing work is based on location information [38], context [39] or historical record [40], and so on. Each of them requires a large amount of storage space. In this paper, we predict the movement trend of a user just based on the channel capacities of its current access point and new access point. Discussions are provided for the following nine cases.

(1) $q_{11}(t) < q_{11}(t+\tau)$ and $q_{21}(t) < q_{21}(t+\tau)$. Because the channel capacity of a_1 increases, u_1 must be inside c'_1. For the same reason, u_1 is also inside c'_2. That is to say, after a stability period, u_1 locates at the domain d_1 shown in Figure 8(a). Because the line $\mathcal{L}$ passing through d_1, we cannot determine the movement trend of u_1. As a result, u_1 should wait for another stability period and then analyze the situation again.

(2) $q_{11}(t) < q_{11}(t+\tau)$ and $q_{21}(t) = q_{21}(t+\tau)$. Because the channel capacity of a_2 does not change, u_1 must locate at c'_2. Furthermore, u_1 is inside c'_1. That is to say, after a stability period, u_1 locates at the line segment l_1 shown in Figure 8(b). l_1 is on the left of $\mathcal{L}$, which means u_1 moves back. As a result, u_1 does not need vertical handoff anymore.

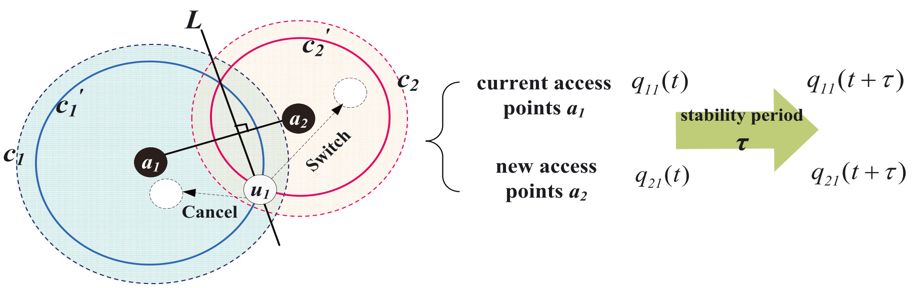
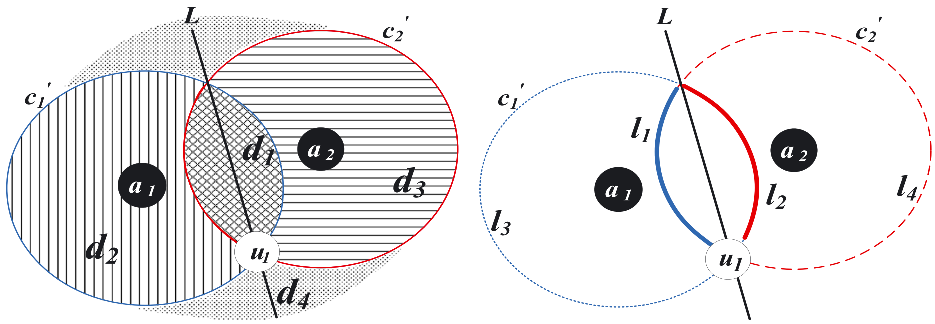


Figure 7. An example of the proposed handoff timing algorithm.



(a) user moves into a region (b) user moves on a line

Figure 8. The movement direction of a user.

(3) $q_{11}(t) < q_{11}(t+\tau)$ and $q_{21}(t) > q_{21}(t+\tau)$. Because the channel capacity of a_2 decreases, u_1 must be outside c_2'. Furthermore, u_1 is inside c_1'. That is to say, after a stability period, u_1 locates at the domain d_2 shown in Figure 8(a). d_2 is on the left of $\mathcal{L}$, which means u_1 moves back. As a result, u_1 does not need vertical handoff anymore.

(4) $q_{11}(t) = q_{11}(t+\tau)$ and $q_{21}(t) < q_{21}(t+\tau)$. Because the channel capacity of a_1 does not change, u_1 must locate at c_1'. Furthermore, u_1 is inside c_2'. That is to say, after a stability period, u_1 locates at the line segment l_2 shown in Figure 8(b). l_2 is on the right of $\mathcal{L}$, which means u_1 moves away. For this case, if the handoff overhead is less than the performance gain (i.e., $o_a + o_r < q_{21}(t+\tau) - q_{11}(t+\tau)$), u_1 should handoff to the new access point a_2 at once. Otherwise, u_1 will initialize another network selection.

(5) $q_{11}(t) = q_{11}(t+\tau)$ and $q_{21}(t) = q_{21}(t+\tau)$. Because the channel capacities of a_1 and a_2 have no change, u_1 still locates at the original point after a stability period. We cannot determine the movement trend of u_1. As a result, u_1 should wait for another stability period, and then analyze the situation again.

(6) $q_{11}(t) = q_{11}(t+\tau)$ and $q_{21}(t) > q_{21}(t+\tau)$. Because the channel capacity of a_2 decreases, u_1 must be outside c_2'. Furthermore, u_1 locates at c_1'. That is to say, after a stability period, u_1 locates at the line segment l_3 shown in Figure 8(b). l_3 is on the left of $\mathcal{L}$, so u_1 does not need vertical handoff anymore.

(7) $q_{11}(t) > q_{11}(t+\tau)$ and $q_{21}(t) < q_{21}(t+\tau)$. Because the channel capacity of a_1 decreases, u_1 must be outside c_1'. Furthermore, the channel capacity of a_2 increases, u_1 must be inside c_2'. That is to say, after a stability period, u_1 locates at the domain d_3 shown in Figure 8(a). Similar to case 4, if the handoff overhead is less than the performance gain (i.e., $o_a + o_r < q_{21}(t+\tau) - q_{11}(t+\tau)$), u_1 should handoff to the new access point a_2 at once. Otherwise, u_1 will initialize another network selection.

(8) $q_{11}(t) > q_{11}(t+\tau)$ and $q_{21}(t) = q_{21}(t+\tau)$. Because the channel capacity of a_2 does not change, u_1 must locate at c_2'. Furthermore, the channel capacity of a_1 decreases, u_1 is outside c_1'. That is to say, after a stability period, u_1 locates at the line segment l_4 shown in Figure 8(b). l_4 is on the right of $\mathcal{L}$, so u_1 takes the same measures as case 4 and case 7.

(9) $q_{11}(t) > q_{11}(t+\tau)$ and $q_{21}(t) > q_{21}(t+\tau)$. Because the channel capacity of a_2 decreases, u_1 must be outside c_2'. Furthermore, u_1 is outside c_1'. That is to say, after a stability period, u_1 locates at the domain d_4 shown in Figure 8(a). The line $\mathcal{L}$ passing through d_4, so we cannot determine the movement trend of u_1. As a result, u_1 should wait for another stability period, then analyze the situation again.

As we have explained that after the new access point is selected, a user should transfer its inter-network connection to the new access point at an appropriate time. For this purpose, we proposed a handoff timing algorithm. The SDN controller only needs to know the channel capacities of current access point and new access point for each handoff user, and the handoff overhead. Based on limited information, the SDN controller can determine the time when a handoff user should implement its network selection result. Following the aforementioned approach, we can see that the network selection result will be implemented only if the user is certain to move away from its current access point.

6. PERFORMANCE EVALUATION

In this section, we provide the performance evaluation of our S-DNVH scheme. We compare the proposed scheme with two typical existing schemes: the ABC scheme [28] and the SASHA [15] under various network conditions. The concerned performance metrics are the number of handoffs, total throughput and the user served ratio. As we know, the SDN technique originated in and is commonly used in the campus networks. Furthermore, most of access points in the campus networks are the IEEE 802.11 standard [41] supported devices. Therefore, we will refer to the IEEE 802.11 standard to set the parameters of access points during the experiment. If other standards are required in the practical applications, this experiment procedure can be repeated by using the corresponding parameters. Simu-

Table II. Experimental parameters.

Parameter	Value
Simulator	Mininet, Floodlight, MATLAB
Number of access points	9
Coverage radii of access points	30 m, 50 m, 100 m
Transmission powers of access points	0.002 watts, 0.005 watts, 0.02 watts
Number of channels in access points	10, 20, 30
Basic bandwidth requirements of users	2 Mbps
Bandwidths of each channels in access points	2 MHz, 1 MHz, 0.6 MHz
Initial value of stability period τ_1	0.1 s
Time slot	1 s
Channel fading gain h	$h \sim exp(1)$
Noise power per channel n	$n \sim N(0, 1)$ watts
Moving velocities of users	$0 \sim 5$ m/s

lation experiments are repeated one thousand times and the results are presented with 95% confidence interval.

6.1. Experiment setup

Over a 500 m × 500 m rectangular flat space, there are nine access points and n users. These nine access points belong to three different types, and each type has three entities. Different types of access points have different coverage radii, number of channels, bandwidths and other different attributes. Furthermore, different types of access points are assumed to use different frequency bands, and same type of access points use the same frequency band. An access point is available to a user when the distance between them is smaller than the coverage radius of this access point. Users are moving around inside the considered area. If the current location of a user is denoted by a two-dimensional coordinate (x, y), this user will be inside $(x \pm \Delta t \cdot \delta, y \pm \Delta t \cdot \delta)$ after a period of time Δt, where δ is the maximal moving velocity of the user. Following this rule, the movement model of users used in our experiments can be illustrated as Figure 9, where (x, y) is the location of a user at the beginning of a time slot, (x', y') is the location of the user after a stability period, and (x'', y'') is the location of the user at the end of the time slot. We set the basic bandwidth requirements of users to be 2 Mbps, which corresponds to the video conference demanding. Based on references [42,43], we have the main experimental parameters, which are listed in Table II.

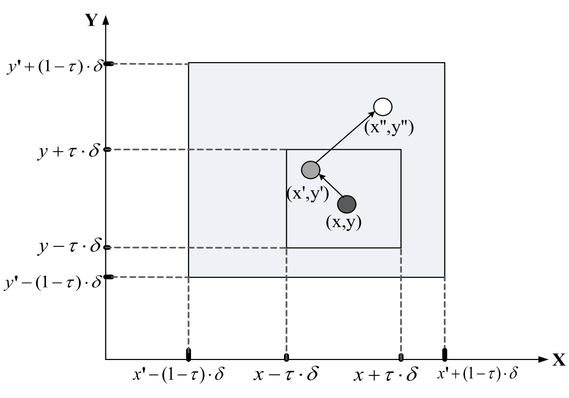


Figure 9. Movement model of users.

In order to construct the SDN scenario and evaluate the performance of proposed S-DNVH scheme, we simulate the OpenFlow switches by using the Mininet VM 2.2.1. The SDN controller is implemented through the Floodlight V 1.2. The SDN controller and OPENFLOW switches exchange information via OPENFLOW 1.3. In order to avoid too much change to the SDN controller, the calculation work is assigned to MATLAB. Hence, a logic complete SDN controller is composed by Floodlight and MATLAB. At first, we write a Python file that is able to insert the features of users into the data fields of Packet-In messages. We make use of the Packet-In message to send information from OPENFLOW switches to the SDN controller. Based on Eclipse, we encapsulate an I/O API for the SDN controller. This API is used for SDN controller to generate a file which contains the information received from switches, and read a file which contains the network selection results calculated by the CVX on MATLAB. Once the SDN controller obtain the network selection results, it will insert the network selection results into the data fields of Packet-Out messages, and send the Packet-Out messages to OPENFLOW switches. The architecture of our simulation system is shown in Figure 10. The switches in Figure 10 represent OPENFLOW enabled access points, and hosts represent users in this paper.

6.2. Experiment results

6.2.1. Number of vertical handoffs.

We study the number of vertical handoffs when the number of users varies under free space propagation ($\alpha = 2$), flat-earth reflection ($\alpha = 3$), and diffraction losses ($\alpha = 4$) environment conditions in Figure 11. The general trend is that there will be more vertical handoffs in each time slot as the number of users increases. For the same number of users, larger pass loss exponent environments always have more handoffs. From the longitudinal comparison, we observe that the proposed S-DNVH scheme can reduce the number of handoffs significantly. This advantage is

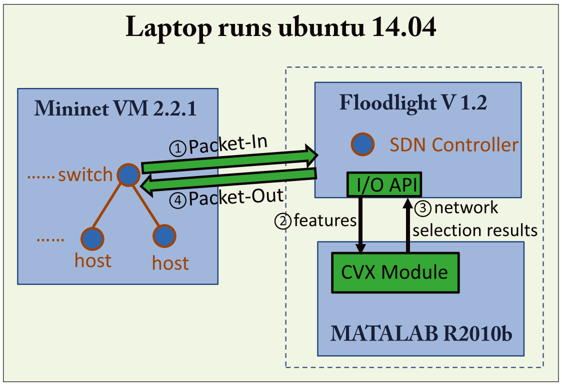


Figure 10. Structure of simulation system.

more remarkable in tough environments. For example in Figure 11(c), the number of handoffs in S-DNVH scheme is nearly 76% less than ABC scheme and 59.6% less than SASHA in the worst case.

Another interesting phenomenon is that the number of vertical handoffs will increase as more users join in. After the number of users is bigger than around 120, the number of vertical handoffs in the two contrast schemes becomes relative stable. However, this phenomenon does not exist in our proposed scheme. We also observe that the ABC scheme is particularly sensitive to environment condition. The number of handoffs in ABC scheme varies dramatically as the pass loss exponent changes. Compared with ABC, situations in other two schemes are much more peaceful. Furthermore, the number of handoffs in SASHA is getting closer to that in our proposed S-DNVH scheme when the pass loss exponent increases.

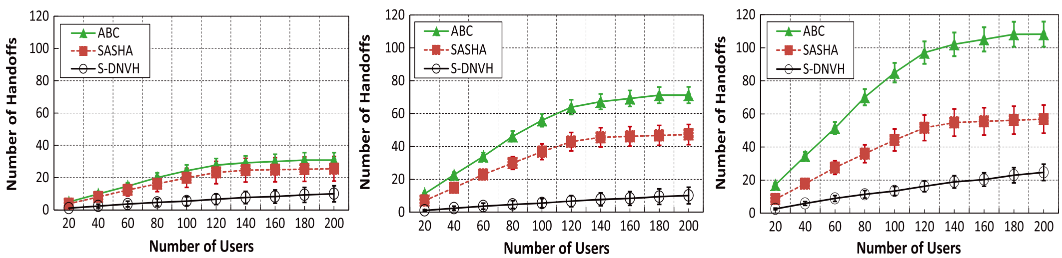
6.2.2. Throughput.

We compare the total throughput under different network settings in Figure 12. The total throughput is defined as the sum of channel capacities that non-handoff users and handoff users can obtain. Because each user is assumed to occupy at most one channel of access point, higher total throughput will be achieved as more users join in at first. Then, the value of total throughput has slower growth when the number of users is larger than a certain value. From Figure 12 we find that, this special value is around 160. Note that, the experimental nine access points theoretically can support maximum 180 users at the same time. Our experimental result is very close to this theoretical value.

From any sub-figure of Figure 12, we can find that S-DNVH scheme always has the highest total throughput. Figure 12 also reveals that an inverse proportion operates between the total throughput and the pass loss exponent. Furthermore, from the crosswise comparison, we can observe that the throughput in our proposed S-DNVH scheme is getting closer to that in SASHA when the pass loss exponent increases. However, even in the worst case, the throughput in S-DNVH scheme is 60% more than in SASHA and nearly 100% more than in ABC scheme.

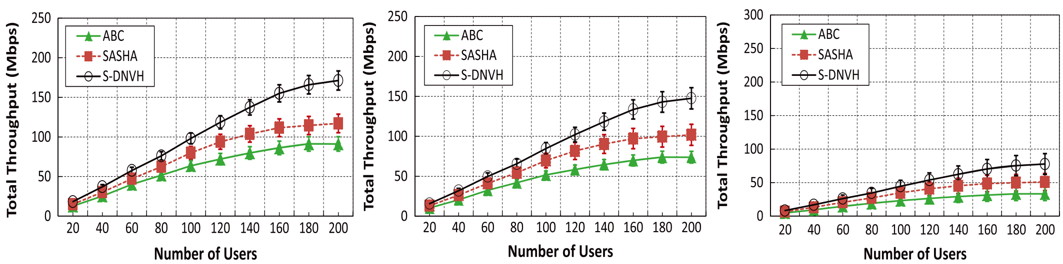
6.2.3. User served ratio.

The user served ratio is the ratio of users who have the network service. Following the definitions and assumptions made in Section 3, the user served ratio is equal to $\frac{\sum_{j=1}^{n} v_j(t)+\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}(t)}{n}$, where n is the number of users, $\sum_{j=1}^{n} v_j(t)$ is the number of non-handoff users and $\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}(t)$ is the number of handoff users whose vertical handoffs are successful. We compare the user



(a) pass loss exponent $\alpha = 2$ (b) pass loss exponent $\alpha = 3$ (c) pass loss exponent $\alpha = 4$

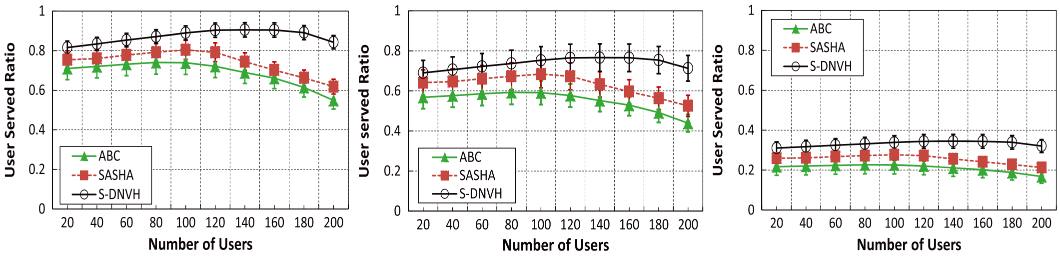
Figure 11. Number of users versus number of handoffs.



(a) pass loss exponent $\alpha = 2$ (b) pass loss exponent $\alpha = 3$ (c) pass loss exponent $\alpha = 4$

Figure 12. Number of users versus stotal throughput.

DOI: 10.1002/wcm



(a) pass loss exponent $\alpha = 2$ (b) pass loss exponent $\alpha = 3$ (c) pass loss exponent $\alpha = 4$

Figure 13. Number of users versus user served ratio.

served ratio under different network settings in Figure 13. It shows that the proposed S-DNVH scheme always has the highest user served ratio in different scenarios. As the number of users increases, the user served ratio sightly increases then decreases. Moreover, user served ratio will be less in higher pass loss exponent environments. As the pass loss exponent increases, the user served ratios in SASHA and the proposed S-DNVH scheme get closer to each other. When comparing our proposed S-DNVH scheme with the SASHA, we can find that the difference of their user server ratios is between 4% and 29%.

Another interesting observation is that the turning points in the two contrast schemes are around 120. This special value has also been found in Figure 11. Thus, we can infer that the maximal number of served users in the two contrast schemes should be 120. If there are more than 120 users are added to the scenario, the performance of two contrast schemes will degrade. However, our proposed scheme does not be affected by this limitation. Compared with these two schemes, our proposed scheme S-DNVH scheme can make fuller use of the system resources.

7. CONCLUSION

In this paper, we proposed a novel vertical handoff scheme with the support of SDN technique. The proposed scheme ensures that a user will transfer to the most appropriate new access point at the most appropriate time. From the standpoint of users, we choose the channel capacity as the performance metric. When the channel capacities cannot meet their basic bandwidth requirements, users will initialize vertical handoffs. We formulated the network selection process as a 0-1 integer programming problem, with the objective of maximizing the sum of channel capacities that handoff users can obtain from their new access points. After the network selection process is finished, users have to wait for a stability period and make a decision. Users will not handoff to the new access points unless the new access points are consistently more appropriate than their current access points during the stability period. We carried out comparison experiments under different network settings. Comparison results demonstrate that even in the worst case, the number of handoffs in the proposed S-DNVH scheme is 59.6% less than in SASHA, and the throughput in S-DNVH scheme is nearly 60% more than in SASHA. Furthermore, their difference of user served ratios is between 4% and 29%.

REFERENCES

1. Cao X, Wang X, Lu S. Function computation over heterogeneous wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems* 2013; **25**(7): 1756–1766.
2. Li J, Kameda H, Li K. Optimal dynamic mobility management for pcs networks. *IEEE/ACM Transactions on Networking (TON)* 2000; **8**(3): 319–327.
3. Li R, Li J, Wu K, Xiao Y, Xie J. An enhanced fast handover with low latency for mobile ipv6. *IEEE Transactions on Wireless Communications* 2008; **7**(1): 334–342.
4. Namal S, Ahmad I, Gurtov A, Ylianttila M. SDN Based inter-technology load balancing leveraged by flow admission control. In *Proceedings of the IEEE SDN for Future Networks and Services (SDN4FNS)*, Nov. 11, 2013; 1–5.
5. Song Q, Jamalipour A. A network selection mechanism for next generation networks. *IEEE International Conference on Communications (ICC)* 2005; **2**: 1418–1422.
6. Márquez-Barja J, Calafate CT, Cano JC, Manzoni P. An overview of vertical handover techniques: algorithms, protocols and tools. *Computer Communications* 2011; **36**(8): 985–997.
7. Yan X, Şekercioğlu YA, Narayanan S. A survey of vertical handover decision algorithms in fourth generation heterogeneous wireless networks. *Computer Networks* 2010; **54**(11): 1848–1863.
8. Stallings W. Software-defined networks and openflow. *The Internet Protocol* 2013; **16**(1): 2–14, Cisco.
9. Akhunzada A, Ahmed E, Gani A, Khan M, Imran M, Guizani S. Securing software defined networks: taxonomy, requirements, and open issues. *IEEE Communications Magazine* 2015; **53**(4): 36–44.
10. Cormen TH, Leiserson CE, Rivest RL, Stein C. *Introduction to algorithms: Chapter 34 NP-Completeness*. MIT press: London, England, 2001 Sep.

11. Wang HJ, Katz RH, Giese J. Policy-enabled handoffs across heterogeneous wireless networks. In *Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, WMCSA'99*, Feb. 25-26, 1999; 51–60.
12. Qiang L, Li J, Huang C. A software-defined network based vertical handoff scheme for heterogeneous wireless networks. In *Proceedings of the IEEE Global Communications Conference (IEEE GLOBECOM)*, Dec.8, 2014; 4671–4676.
13. Hou J, O'Brien DC. Vertical handover-decision-making algorithm using fuzzy logic for the integrated radio-and-OW system. *IEEE Transactions on Wireless Communications* 2006; **5**(1): 176–185.
14. Lee W, Cho D-H. Enhanced group handover scheme in multiaccess networks. *IEEE Transactions on Vehicular Technology* 2011; **60**(5): 2389–2395.
15. Ciubotaru B, Muntean G-M. SASHA – a quality-oriented handover algorithm for multimedia content delivery to mobile users. *IEEE Transactions on Broadcasting* 2009; **55**(2): 437–450.
16. Agarwal S, Kodialam M, Lakshman TV. Traffic engineering in software defined networks. In *Proceedings of the IEEE INFOCOM*, Apr 14-19, 2013; 2211–2219.
17. Dabbagh M, Hamdaoui B, Guizani M, Rayes A. Software-defined networking security: pros and cons. *IEEE Communications Magazine* 2015; **53**(6): 73–79.
18. Yin H, Xie H, Tsou T, Lopez D, Aranda P, Sidi R. SDNi: A message exchange protocol for software defined networks (SDNs) across multiple domains. In *IETF Draft, Work in Progress*, Jun 27, 2012.
19. Jain S, Kumar A, Mandal S, Ong J, Poutievski L, Singh A, Venkata S, Wanderer J, Zhou J, Zhu M, Zolla J. B4: Experience with a globally-deployed software defined WAN. In *Proceedings of ACM SIGCOMM Computer Communication Review*, Aug. 12, 2013; 3–14.
20. Taniuchi K, Ohba Y, Fajardo V, Das S, Tauil M, Cheng Y-H, Dutta A, Baker D, Yajnik M, Famolari D. IEEE 802.21: Media independent handover: features, applicability, and realization. *IEEE Communications Magazine* 2009; **47**(1): 112–120.
21. Andrews JG, Baccelli F, Ganti RK. A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications* 2011; **59**(11): 3122–3134.
22. Novlan TD, Ganti RK, Ghosh A, Andrews JG. Analytical evaluation of fractional frequency reuse for OFDMA cellular networks. *IEEE Transactions on Wireless Communications* 2011; **10**(12): 4294–4305.
23. Yang K, Gondal I, Qiu B, Dooley LS. Combined SINR based vertical handoff algorithm for next generation heterogeneous wireless networks. *IEEE Golbecom* 2007: 4483–4487.
24. Sato H. The capacity of the Gaussian interference channel under strong interference (corresp.) *IEEE Transactions on Information Theory* 1981; **27**(6): 786–788.
25. Shannon CE. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review* 2001; **5**(1): 3–55.
26. *Pre-loaded Switches*. [Online] Available: http://www.pica8.com/products/pre-loaded-switches [Accessed June 2rd, 2016].
27. *Inter-Datacenter WAN with centralized TE using SDN and OpenFlow*. [Online] Available: https://www.opennetworking.org/images/stories/downloads/sdn-resources/customer-case-studies/cs-googlesdn.pdf [Accessed June 2rd, 2016].
28. Louta M, Bellavista P. Bringing always best connectivity vision a step closer: challenges and perspectives. *IEEE Communications Magazine* 2013; **51**(2): 158–166.
29. *Nettle OpenFlow Messages*. [Online] Available: http://hackage.haskell.org/package/nettle-openflow-0.2.0/docs/Nettle-OpenFlow-Messages.html#t:CSMessage [Accessed June 2rd, 2016].
30. Li J, Chen H-H. Mobility support for ip-based networks. *IEEE Communication Magazine* 2005; **43**(10): 127–132.
31. Mishra A, Shin M, Arbaugh W. An empirical analysis of the IEEE 802.11 MAC layer handoff process. *ACM SIGCOMM Computer Communication Review* 2003; **33**(2): 93–102.
32. Jain A. *Handoff delay for 802.11 b wireless lans. Project Report (2003)*. [Online]. Available: http://www.cs.columbia.edu/~hgs/papers/2003/AnshulJain80211.ppt [Accessed Jan. 10th, 2016].
33. Gast M. *802.11 wireless networks: the definitive guide. Chapter 4–802.11 Framing in Detail O'Reilly Media, Inc*. Available: http://osnet.cs.nchu.edu.tw/powpoint/seminar/802.11/802.11Framing%20in%20Detail.pdf [Accessed June 2rd, 2016].
34. Sharma S, Zhu N, Chiueh T. Low-latency mobile IP handoff for infrastructure-mode wireless LANs. *IEEE Journal on Selected Areas in Communications* 2004; **22**(4): 643–652.
35. Li J, Kubota A, Kameda H. Location management for pcs networks with consideration of mobility patterns. In *Proceedings of the IEEE International Conference on Computer Communications (IEEE INFOCOM)*, March 13-17, 2005; 1207–1216.
36. Ma D, Ma M. A QoS oriented vertical handoff scheme for WiMAX/WLAN overlay networks. *IEEE Transactions on Parallel and Distributed Systems* 2012; **23**(4): 598–606.
37. Aristomenopoulos G et al. An autonomic QoS-centric architecture for integrated heterogeneous wireless net-

works. *Mobile Networks and Applications* 2011; **16**(4): 490–504.
38. Mumtaz S, Monteiro V, Rodriguez J, Politis C. Positioning for performance enhancement in shared LTE and HSPA networks. In *Proceedings of the International Conference on Telecommunications and Multimedia (TEMU)*, Jul 30, 2012; 169–174.
39. Ahmed T, Kyamakya K, Ludwig M, Anne KR, Schroeder J, Galler S, Kyamakya K, Jobmann K. A context-aware vertical handover decision algorithm for multimode mobile terminals and its performance. In *Proceedings of the Euro American Conference on Telematics and Information Systems (EATIS)*, Feb. 7-10, 2006; 19–28.
40. Kim T, Han S, Han Y. A QoS-aware vertical handoff algorithm based on service history information. *IEEE Communications Letters* 2010; **14**(6): 527–529.
41. Hiertz GR, Denteneer D, Stibor PL, Zang Y, Costa XP, Walke B. The IEEE 802.11 universe. *IEEE Communications Magazine* 2010; **48**(1): 62–70.
42. *Effect of Transmit Power Changes on AP Cell Sizing*. Available: http://wifinigel.blogspot.jp/2014/11/effect-of-transmit-power-changes-on-ap.html [Accessed June 2rd, 2016].
43. *Best Practices for High Density Wireless Network Design in Education and Small/Medium Businesses*. Available: http://www.netgear.com/images/pdf/High_Density_Best_Practices.pdf [Accessed June 2rd, 2016].

AUTHORS' BIOGRAPHIES

Li Qiang is a PhD candidate at the Graduate School of Systems and Information Engineering, University of Tsukuba, Japan. She received the BE degree from Anhui University, China, in 2009, and ME Degree from University of Science and Technology of China, China, in 2012. Her main research interests include mobility management, heterogeneous wireless networks, 5G, and software-defined networking (SDN).

Jie Li received the BE degree in Computer Science from Zhejiang University, Hangzhou, China, and the ME degree in Electronic Engineering and Communication Systems from China Academy of Posts and Telecommunications, Beijing, China. He received the Dr. Eng. degree from the University of Electro-communications, Tokyo, Japan. He is with the University of Tsukuba, Japan, where he is a professor. He has been a visiting professor in Yale University, USA, and in Inria, France. His current research interests are in mobile-distributed computing and networking, big data and cloud computing, IoT, OS, and modeling and performance evaluation of information systems. He is a senior member of IEEE and ACM and a member of IPSJ (Information Processing Society of Japan). He is the chair of Technical Sub-committee on Big Data (TSCBD), IEEE Communications Society. He has served as a secretary for Study Group on System Evaluation of IPSJ and on several editorial boards for the international journals, and on Steering Committees of the SIG of System EVAluation (EVA) of IPSJ, the SIG of DataBase System (DBS) of IPSJ, and the SIG of MoBiLe computing and ubiquitous communications of IPSJ. He has also served on the program committees for several international conferences such as IEEE INFOCOM, IEEE GLOBECOM, and IEEE MASS.

Yusheng Ji received BE, ME, and DE degrees in Electrical Engineering from the University of Tokyo. She joined the National Center for Science Information Systems, Japan (NACSIS) in 1990. Currently, she is a professor at the National Institute of Informatics, Japan (NII), and the Graduate University for Advanced Studies (SOKENDAI). Her research interests include network architecture, resource management, and quality of service provisioning in wired and wireless communication networks. She has been a Board Member of Trustees of IEICE and is an Expert Member of IEICE Technical Committees on Internet Architecture, and Communication Quality, and a Committee Member of Quality Aware Internet (QAI) SIG and Internet and Operation Technologies (IOT) SIG of IPSJ. She is an editor of IEEE Transactions of Vehicular Technology and has been an associate editor of IEICE Transactions and IPSJ Journal, an editor of KSII Transactions on Internet and Information Systems, a guest editor-in-chief, guest editor, guest associate editor of Special Sections of IEICE Transactions, guest associate editor of Special Issues of IPSJ Journal, and so on. She has also served as a TPC member of many conferences, including IEEE INFOCOM, ICC, GLOBECOM, VTC, and so on, and symposium co-chair of IEEE GLOBECOM 2012 and 2014, track co-chair of IEEE VTC 2016 Fall.

Changcheng Huang received his B Eng. in 1985 and M Eng. in 1988 both in Electronic Engineering from Tsinghua University, Beijing, China. He received a PhD degree in Electrical Engineering from Carleton University, Ottawa, Canada in 1997. From 1996 to 1998, he worked for Nortel Networks, Ottawa, Canada where he was a systems engineering specialist. He was a systems engineer and network architect in the Optical Networking Group of Tellabs, Illinois, USA during the period of 1998–2000. Since July 2000, he has been with the Department of Systems and Computer Engineering at Carleton University, Ottawa, Canada where he is currently a full professor. Dr. Huang won the CFI new opportunity award for building an optical network laboratory in 2001. He is an associate editor of Springer Photonic Network Communications. Dr. Huang is a senior member of IEEE.

DOI: 10.1002/wcm