Design and Analysis of Mobility-Aware Clustering Algorithms for Wireless Mesh Networks

Nizar Bouabdallah, Rami Langar, and Raouf Boutaba

***Abstract*—One of the major concerns in wireless mesh networks (WMNs) is the radio resource utilization efficiency, which can be enhanced by efficiently managing the mobility of users. To achieve this, we propose in this paper the use of mobility-aware clustering. The main idea behind WMN clustering is to restrict a major part of the exchanged signaling messages due to the mobility of users to a local area (i.e., inside a cluster). As such, less wireless links are used by the signaling messages, which reduces the resources occupied by a mobile user during its service and thus improves the total network capacity. Through analytical models and simulations, this work first identifies the cases where clustering is helpful. Building on these results, we propose two clustering schemes that take into consideration the mobility properties of the users in order to improve the WMN performance. We prove that both schemes can achieve significant gains in terms of radio resource utilization. Specifically, we show that the first scheme fits better both large and low-connected wireless mesh networks, whereas the second scheme is more suitable for both small to moderate and highly connected networks.**

***Index Terms*—Clustering, mobility management, performance analysis, radio resources, wireless mesh networks (WMNs).**

I. Introduction

Wireless mesh networks (WMNs) have emerged recently as a promising solution to support the increasing demand for mobile wireless access to the Internet. A variety of applications are expected to benefit from WMNs such as "community wireless networks" [1]–[3]. Typically, a WMN comprises static wireless mesh routers, also called access points (APs). Each AP serves multiple mobile users and connects them through multihop wireless routing to the wired network. The mesh nodes connected directly to the wired network (i.e., connecting the WMN to the wired network) are called gateways. They represent the traffic sinks and sources to the WMN.

As opposed to ad hoc networks, WMNs have a stable topology, which changes occasionally due to AP failures or when new APs are added to the system. As a second main distinguishing feature from ad hoc networks, the communications in WMNs are performed chiefly between the gateways and their associated APs and not directly between arbitrary pairs of mesh nodes.

In WMNs, two types of messages are exchanged between the gateway and a mobile user: data and signaling messages. Specifically, each time the mobile user moves to a new AP, it notifies the gateway with its new location. These signaling messages are relayed through multihop wireless links to the gateway. For convenience, we assume a multichannel WMN where interfering links operate on different channels, enabling multiple parallel transmissions. In order to achieve an efficient use of the radio resources in such multihop wireless networks, we need to minimize the average number of links occupied by a mobile user during its service (i.e., while exchanging both data and signaling packets). Henceforth, we refer to this metric as the radio resource utilization (RRU) cost.

Reducing the RRU cost per user improves the utilization efficiency of the WMN resources and thus increases the number of accepted subscribers to the WMN service. This means more profits to the service provider from the existing infrastructure. One way to achieve this is by means of clustering as demonstrated in our study. Accordingly, the WMN is divided into a set of virtual clusters, covering all the nodes in the network. In each cluster, a node would serve as a clusterhead (CH). It operates as an intermediate node between the gateway and the APs inside the cluster. With regard to the management operations, the CH substitutes to the gateway inside the cluster and manages the mobility of local users. As such, less wireless links are used by the signaling messages, which reduces the RRU cost and improves thus the total network throughput.

In this paper, we propose two clustering schemes in order to minimize the RRU cost in WMNs. In the first scheme, called Optimal Static Clustering (OSC), we assume that the clusters are static and disjoint. In this case, we determine the optimal cluster placement that minimizes the RRU cost by formulating the placement problem as an integer linear program (ILP). In the second scheme, called Distributed Clustering Algorithm for Mesh networks (DCAM), the clusters may overlap. In this case, the cluster placement is done in a distributed manner. Typically, each AP calculates its own cluster when it acts as a CH. In essence, the DCAM approach is proposed to alleviate the time complexity entailed by the OSC approach due to the time-consuming resolution of the associated ILP problem. In this context, the cases of OSC, classical WMNs (i.e., without clustering), and the well-known distributed clustering algorithm (DCA) [4], which is proposed originally for ad hoc networks, are used also to develop baselines to which the DCAM improvements could be compared since no clustering approach for mobility management in the context of WMNs has been previously proposed in

Manuscript received November 19, 2008; revised September 13, 2009; accepted March 22, 2010; approved by IEEE/ACM Transactions on Networking Editor E. Modiano. Date of publication June 14, 2010; date of current version December 17, 2010. This work was supported by the Natural Science and Engineering Council of Canada (NSERC) under its Discovery program and the WCU (World Class University) program through the Korea National Research Foundation funded by the Ministry of Education, Science and Technology (Project No. R31-2008-000-10100-0).

N. Bouabdallah is with INRIA, 35042 Rennes, France (e-mail: nizar.bouabdallah@inria.fr).

R. Langar is with the Computer Science Laboratory of Paris 6 (LIP6), University of Pierre and Marie Curie-Paris Universitas, 75005 Paris, France (e-mail: rami.langar@lip6.fr).

R. Boutaba is with the Computer Science Department, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: rboutaba@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2010.2049579

the literature. It is worth noting that in both OSC and DCAM clustering schemes, we take into account the mobility properties of the users. To the best of our knowledge, we are the first to introduce such constraint in the clustering process.

There are indeed several mobility-aware clustering algorithms that have been proposed in the context of ad hoc networks such as [4] and [5]. However, these works considered the mobility of APs (i.e., either pure APs or mobile users operating as APs) rather than that of ordinary mobile users. The mobility of ordinary users, not serving as APs, was not considered before.

To gauge the effectiveness of our proposals, a Markov-chain-based model is developed to analyze the DCAM algorithm. In the OSC case, the problem is formulated as an ILP. Building on these models, we derive the expressions of RRU cost, registration updates cost, data delivery cost, and load balancing for both cluster-based WMNs (using OSC, DCAM, and DCA methods) and classical WMNs (i.e., without clustering, denoted by WC). These results are also validated by simulations. We evaluate the performance of our algorithms using both regular and arbitrary meshed topologies under various mobility and traffic scenarios. We first study the benefit of clustering in WMNs. Explicitly, we identify the cases where clustering is useful. Then, we compare the proposed clustering schemes.

II. Related Work

Management of wireless multihop networks has been an active research area in the last few years. Numerous proposals for clustering and hierarchical routing schemes have been proposed in the literature, essentially in the context of ad hoc networks and more recently in the context of wireless sensor networks (WSNs) and WMNs.

Clustering in ad hoc networks is introduced mainly to efficiently handle the frequent network topological changes due to *ad hoc* nodes mobility [4]–[13]. The main objective of clustering has therefore been to adapt quickly to topological changes, which occur only occasionally in WMNs. Designed for ad hoc networks, these protocols are not suitable for typical WMN applications for two main reasons, as explained in the previous section: the static topologies of WMNs and the different communication patterns.

In [4], the author presents a clustering algorithm for "quasi-static" ad hoc networks, where nodes are static or moving at a very low speed. The proposed scheme is more adapted to the WMN environment. However, it is concerned with one-hop clustering, which defeats the purpose of clustering in WMNs.

[5] proposes a (α, t) cluster strategy for ad hoc networks that forms clusters using criteria based directly on node mobility. The algorithm dynamically organizes the nodes of an ad hoc network into clusters where probabilistic bounds can be maintained on the availability of paths to cluster destinations over a specific interval of time. Again, the main purpose of this clustering scheme is to simplify the routing operations in such mobile environment. Accordingly, the network is rearranged into clusters following each change in the topology due to node mobility. As before, this scheme is not appropriate for WMNs since it is based on the mobility of APs, which are static in WMNs.

Other works [9], [10] proposed d-hop clustering algorithms in wireless ad hoc networks such that a node in any cluster is at most d hops away from the CH. The proposed schemes are not restricted to one-hop clustering. However, the clusters have the same radius d, an additional constraint, which may lead to unsatisfactory results regarding the RRU cost minimization.

Another clustering approach in mobile ad hoc networks, based on graph theory, namely connected dominating set (CDS), is used in [11]–[13]. In this approach, the objective is to identify the smallest set of CHs that forms a CDS. This problem is known in graph-theoretic terminology as Minimum Connected Dominating Set (MCDS). Hence, by definition of a CDS, each *ad hoc* node not included in the set of CHs has at least one adjacent node belonging to the CH set. The set of CHs operates therefore as routers and forms a virtual backbone for the ad hoc network. Again, it is easy to see that the proposed scheme is concerned with one-hop clustering, which defeats the purpose of WMNs.

In addition to the above-discussed clustering algorithms used in ad hoc networks, more recent clustering algorithms have also been proposed in the context of WSNs [14]–[16]. The common criterion for the selection of CHs with these algorithms is based on the energy consumption constraint. Instead, efficiently using the wireless resources is the main concern in WMNs and is crucial to achieve acceptable performance. In this regard, the goal with our clustering schemes is geared more toward the efficient use of the scarce wireless resources rather than the reduction and balancing of the energy consumption.

In the context of WMNs, relevant works on clustering are [17]–[20]. These works attempt to integrate the WMNs with the wired backbone. More specifically, [17]–[20] investigated the well-known problem of gateway placement in WMNs. This consists of dividing the WMN into a minimum number of disjoint macro-clusters, where each macro-cluster is assigned to a gateway node that connects directly to the wired network. The objective, therefore, is to minimize the number of deployed expensive gateways (i.e., number of macro-clusters) required to connect all APs to the wired network subject to several QoS constraints such as the gateway capacity, the cluster radius, etc.

In our study, we focus rather on virtual clustering inside each macro-cluster. Our objective is to divide the macro-cluster into virtual micro-clusters in order to minimize the RRU cost in the WMN. Hence, while macro-clustering is performed to minimize the number of required gateways, micro-clustering is performed to optimize the RRU cost in WMNs. Both clustering approaches are complementary and needed to achieve cost-effective WMNs.

III. Model and Problem Description

A. Network Model

We represent a WMN by an undirected graph $G(V, E)$, called a connectivity graph. Each node $v \in V$ represents an AP with a circular transmission range R_t. The neighborhood of v, denoted by $N_e(v)$, is the set of nodes residing in its transmission range. A bidirectional wireless link exists between v and every neighbor $u \in N_e(v)$ and is represented by an edge $(u, v) \in E$. The number of neighbors of a vertex v is called the connectivity degree of v, denoted by $\delta(v)$. The average connectivity degree $\overline{\delta}$ of a graph G is called the graph degree and is defined by $(1/|V|) \sum_{v \in V} \delta(v)$. Moreover, we denote the distance $d(u, v)$ between two nodes u and v as the minimum number of hops between these nodes.

In our study, we represent the graph connectivity by a connectivity matrix (i.e., adjacency matrix). The connectivity matrix of $G(V, E)$ is a matrix with rows and columns labeled by the graph vertices V, with a 1 or 0 in position (m, n) according to whether v_m and v_n are directly connected or not. We associate to the connectivity matrix a distance matrix D representing the distance between every pair of nodes in the graph G. This matrix is simply derived from the connectivity matrix by using for instance the dijkstra algorithm.

B. Problem Description

In this paper, we focus on the efficient mobility management in WMNs by minimizing the RRU cost of a mobile user during its service. Reducing the RRU per user improves the utilization efficiency of the WMN resources and thus increases the number of accepted subscribers to the WMN service.

To achieve this, we use clustering. Three hierarchical levels in the WMN are identified in this case. At the top of the hierarchy is the gateway node, which is connected directly to the wired network for Internet access and serves all the mesh nodes (i.e., APs) in the WMN. The second level of hierarchy is the CH, serving the APs inside the cluster, and the third level is the AP offering IP-layer connectivity to mobile users located within its transmission range (i.e., inside the AP subnet).

For mobility management, we define two types of handoffs: intracluster and intercluster handoffs. An intracluster handoff occurs when a user moves between two APs that belong to the same cluster. On the other hand, an intercluster handoff occurs when a user moves between two APs belonging to different clusters. To maintain connectivity during user mobility, the gateway keeps a record of the current user cluster (i.e., the current CH's identity). Each time the user crosses a cluster boundary (i.e., intercluster handoff), it updates the system with its new location by sending a registration update message to the gateway through the CH of the new visited cluster. We call this kind of registration update *GW registration*. In contrast, when an intracluster handoff occurs, the update registration message will be sent only to the current CH and will not be forwarded to the gateway. This kind of registration is called *CH registration*.

With regard to data packets, an incoming packet from the backbone to the mobile user (i.e., downlink traffic) is first intercepted by the gateway. Then, the packet is forwarded to the current user CH, which relays the data packet to the corresponding AP for delivery. In turn, data packets transmitted by the mobile user to the wired network (i.e., uplink traffic) are directly routed to the gateway, for instance, according to the shortest path, without requiring to pass through the CH.

The RRU cost of a mobile user involves two terms, i.e., the first one related to the resources used by the data packets, and the second term is related to the resources used by the signaling messages necessary for managing user mobility. We refer to the first term as the *data delivery cost* and to the second term as the *registration updates cost*.

Clearly, an efficient clustering policy must minimize the sum of these two terms. It must achieve a balance regarding the cluster sizes in terms of number of APs. Specifically, a WMN with small-size clusters will result in an increasing number of expensive intercluster handoffs instead of low-cost local intracluster handoffs during the user mobility. As such, the registration updates cost increases. On the other hand, large-size clusters will reduce the number of intercluster handoffs. However, this will result in an increase of the data delivery cost. In view of this, the registration updates cost and the packet delivery cost are two opposite requirements. A tradeoff between these two requirements can be achieved by optimally constructing the clusters while minimizing the RRU cost.

To achieve the above tradeoff, we suggest two clustering schemes. We first assume that the clusters are static and disjoint. In this case, we determine the optimal cluster placement that minimizes the RRU cost by formulating the placement problem as an ILP. As an alternative to the static clustering (i.e., OSC), we propose a distributed algorithm to construct the virtual clusters. In this case, the clusters may overlap, and their placement depends on the user mobility (i.e., trajectory). This method (i.e., DCAM) is introduced to alleviate the long computing time required to resolve the static clustering ILP problem. A detailed description of both clustering schemes is given in the next sections.

In the remainder of this paper, we will compare between the following networks based on their RRU costs: classical WMN [i.e., without clustering (WC)] and cluster-enabled networks using DCA [4] as well as our OSC and DCAM schemes. To achieve this, both analytical models and simulations are used.

IV. Optimal Static Clustering

In this section, we address the optimal placement of clusters in WMNs using the OSC method. We formulate the clustering problem as follows. Given a WMN of N nodes, find the disjoint sets of APs (i.e., clusters) that minimizes the total radio resource utilization subject to the delay constraint. In our study, we assume a multichannel WMN where interfering wireless links operate on different channels, thus enabling multiple parallel transmissions. In such a multihop network, the delay is proportional to the number of hops between the gateway and the receiving node. The delay constraint is thus translated into an upper bound $D_{\max}$ on the number of hops that a packet can cross between the gateway and the AP connecting the mobile user. In other words, the indirect path between the gateway and the mobile user's AP through the current CH must be less or equal to $D_{\max}$.

A. ILP Formulation

Let $N = |V|$ be the number of APs. The APs are denoted by AP_i, $(i = 1, \ldots, N)$. We denote by $AP_1 \in V$ the gateway (GW) that connects the WMN to the wired network. We introduce a binary variable a_i to indicate whether an $AP_i \in V$ is set up as a CH or not. To represent CHs allocation for APs, we define another binary variable $b_{i,j}$, which takes the value of 1 whenever AP_j, $j = 1, \ldots, N$, is assigned to the CH AP_i, $i = 1, \ldots, N$. M is an upper bound on the number of clusters that can be formed. m_{sig} and m_{data} represent the average size of signaling messages used for registration updates and the average size of data packets, respectively. $1/\mu$ represents the mean sojourn time of a mobile user in a subnet (i.e., AP), and λ is the downlink packet transmission rate (in terms of packets/s). For each user inside the WMN, also let its mobility pattern be defined by the process $\{Y(t), t \geq 0\}$, where $Y(t) = AP_i$ represents the user's location at time t. Given the matrix of distances D between APs, the steady probability Π_i that the mobile user is located at the physical subnet AP_i, and the transition probability matrix P of the process Y, our objective function will be to minimize the RRU cost in the WMN. The RRU cost can be expressed as follows:

$$RRU\_Cost = \alpha \times Reg\_Update\_Cost + \beta \times Data\_Delivery\_Cost \quad (1)$$

where $\alpha = 2\,\mu\, m_{\rm sig}/(2\,\mu\, m_{\rm sig} + \lambda\, m_{\rm data})$ and $\beta = \lambda\, m_{\rm data}/(2\,\mu\, m_{\rm sig} + \lambda\, m_{\rm data})$ are the proportion of the amount of signaling messages and the proportion of data packets among the total traffic generated by a mobile user, respectively. The Reg_Update_Cost can be written as follows:

$$Reg\_Update\_Cost = \sum_{i=1}^{N} \Pi_i \times Update\_Cost(i) \quad (2)$$

where $Update\_Cost(i)$ is the cost of the registration updates when the mobile user leaves the AP_i. It is given by

$$\sum_{j=1}^{N}\left(P(i,j) \times \sum_{k=1}^{N} b_{k,j} \times [d(k,j) + (b_{k,j} - b_{k,i}) d(k,GW)]\right) \quad (3)$$

where $d(i,j) = d(AP_i, AP_j)$ is the distance between AP_i and AP_j, and $P(i,j) = P(AP_i, AP_j)$ denotes the transition probability from AP_i to AP_j.

Likewise, the $Data\_Delivery\_Cost$ can be expressed as follows:

$$Data\_Delivery\_Cost = \sum_{i=1}^{N} \Pi_i \times Delivery\_Cost(i) \quad (4)$$

where $Delivery\_Cost(i)$ is the data delivery cost of downlink traffic when the mobile user is connected to AP_i. It is given by

$$Delivery\_Cost(i) = \sum_{j=1}^{N} b_{j,i} \times [d(i,j) + d(j,GW)]. \quad (5)$$

Note that in our study, we consider only the packet delivery cost of the downlink traffic. This is because the packet delivery cost in the uplink direction (i.e., from the mobile user to the gateway) is the same for all the approaches (i.e., OSC, DCAM, DCA, and WC) since the shortest path is always used in this direction. Therefore, we omit this term as it is the same for all the compared approaches.

Hence, our ILP problem can be formulated as follows with the objective function:

$$\min\{RRU\_Cost\}$$

subject to :

(a) $\forall j = 1, \ldots, N : \sum_{i=1}^{N} b_{i,j} = 1$

(b) $\forall i,j = 1, \ldots, N : a_i \geq b_{i,j}$

(c) $a_1 = a_{GW} = 1$

(d) $\forall i = 1, \ldots, N : a_i = b_{i,i}$

(e) $\sum_{i=1}^{N} a_i \leq M$

(f) $\forall j = 1, \ldots, N : \sum_{i=1}^{N} b_{i,j} \times (d(j,i) + d(i,GW)) \leq D_{\max}$

(g) $\forall i = 1, \ldots, N : a_i \in \{0,1\}$

(h) $\forall i,j = 1, \ldots, N : b_{i,j} \in \{0,1\}$.

Condition (a) denotes that each AP is assigned to one and only one CH. Inequality (b) implies that a CH has to be set up before being assigned APs. Inequalities (c) and (d) ensure that the gateway cannot be assigned to another cluster and each CH belongs to the cluster that it manages. Inequality (e) provides an upper bound on the number of the constructed clusters that can be parameterized by the WMN administrator. For instance, assigning the value of 1 to M implies a WMN without clustering. On the other hand, putting $M = N$ implies that all the APs have the capability to operate as CHs. Inequality (f) traduces the delay constraint. The last two conditions indicate that a_i and $b_{i,j}$ are binary variables.

We will show in Section V-B how to derive the vector Π and the transition probability matrix P, which are used as input to our ILP formulation.

In practice, once the clusters are identified by solving the ILP problem, a *Cluster Table* at each CH is implemented. The table associated to a given CH contains the set of APs assigned to that CH. Each AP also retains the identity of its corresponding CH. Finally, the mobile user maintains the identity of the CH of its connecting AP during its movement. Once the mobile user moves to a new AP, it registers by sending a signaling message to the new AP containing the identity of its current CH. This registration message will be forwarded by the new AP to its CH. Accordingly, the CH achieves either a local registration (i.e., only a *CH registration* if the old and the new CH identities are the same) or a *GW registration* by forwarding the received message to the gateway.

V. Distributed Clustering Algorithm

Motivated by the dynamic and distributed nature of the clustering protocol operations, we now propose a distributed clustering policy. This policy is a new alternative to divide the WMN into clusters while avoiding the time complexity of the static approach based on the time-consuming ILP problem resolution. In the following, we present our distributed clustering algorithm for mesh networks (DCAM). Then, we develop a new analytical model using Markov chains to evaluate the performance of DCAM in terms of RRU cost, registration updates cost, packet delivery cost, and load balancing. The elaborated model will also be used to derive the performance metrics used as input to the ILP formulation presented in Section IV (i.e., the vector Π and the matrix P).

A. DCAM Algorithm

Like the OSC approach, the DCAM algorithm divides the WMN into virtual clusters where the mobile user limits its registration updates within this local area. The mobile user keeps registering with the current CH instead of the gateway as long as it moves inside the current virtual cluster. As a distinguishing key feature, the DCAM clusters are constructed in a distributed manner and may overlap, as opposed to the disjoint and centrally calculated OSC clusters. Moreover, the virtual cluster construction with DCAM depends on the mobile user trajectory. It de-

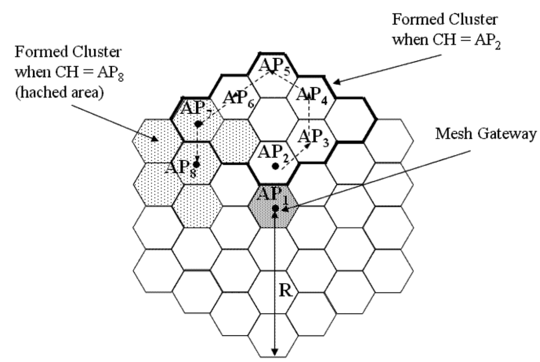

Fig. 1. The DCAM clustering scheme. An example of cluster placement in an hexagon-based regular wireless mesh network with $R = 3$ and $D_{\max} = 4$. Illustration of the clusters associated to CHs AP_2 and AP_8.

```
%Location registration procedures
IF (Mobile user enters a new subnet)
    New AP checks the existence of the current
    CH's IP address in its routing table;
    IF (the address exists)
        New AP finds the shortest distance to the CH
        and to the gateway;
        IF (d(newAP, CH) < d(newAP, GW))
        & (d(newAP, CH) + d(CH, GW) ≤ D_max)
            Perform a CH registration: new AP added to the cluster;
        ELSE
            Perform a GW registration: new cluster formation;
            CH renewal: new AP becomes the new CH;
        ENDIF
    ENDIF
ENDIF
```

Fig. 2. The DCAM algorithm.

pends on the relative position of the old and the new APs with respect to the gateway and on the delay constraint.

Specifically, assume that the maximum tolerable delay inside the WMN is $D_{\max}$, which is expressed in terms of hops. According to the DCAM approach, each time the mobile user moves to a new AP, it first compares the length of its indirect path to the gateway through the current CH with $D_{\max}$. We refer to this operation as the delay constraint verification. If this distance to the gateway is less than or equal to $D_{\max}$, the mobile user can register locally to the current CH. Otherwise, it registers directly to the gateway, and the new AP becomes the CH of the new cluster. As such, the new AP is considered as being outside the previous cluster. Moreover, to minimize the signaling cost, a second condition must be verified. Specifically, a *CH registration* is achieved as long as it is cheaper than a *GW registration*. Indeed, each time the mobile user moves to a new AP, the latter compares the signaling cost (in terms of occupied wireless links by the signaling messages) of a registration update to the CH with that to the gateway. In other words, once the distance between the new visited AP and the gateway is less than or equal to the distance between the new AP and the current CH, a *GW registration* is preferred. Thus, the new AP is considered as not belonging to the previous cluster.

To illustrate the DCAM algorithm, we consider the simple example presented in Fig. 1, where a hexagon-based regular wireless mesh topology is used. The gateway is placed at the center of the WMN. The WMN contains the gateway surrounded by three rings of subnets (i.e., APs). Hence, the radius R of the WMN is 3. We assume that $D_{\max} = 4$. It is worth noting that $D_{\max}$ must be at least equal to R.

Assume that the mobile user moves from AP_2 to AP_8 as depicted in Fig. 1. The DCAM algorithm operation can be described as follows. The mobile user begins its trajectory at AP_2, which is set up as the current CH. When the mobile user moves to AP_3, it compares first between the local and gateway registration costs. Explicitly, the new visited AP_3 compares its distance to the current CH (i.e., one hop) with that to the gateway (i.e., two hops). In this case, a *CH registration* is cheaper than the *GW registration*. In addition, the delay constraint is respected as the distance to the gateway through the CH (i.e., two hops) is less than $D_{\max}$. Consequently, the mobile user performs a *CH registration*. Likewise, the mobile user always achieves a local registration to the CH (i.e., AP_2) when it moves to AP_4, AP_5, AP_6, and AP_7 as it still fulfills the delay and the *CH registration* constraints. In fact, the associated cluster to the CH AP_2 is composed of nine APs, as shown in Fig. 1, since these APs satisfy both conditions.

The condition on the registration updates cost is no longer verified when the mobile user enters subnet AP_8. Hence, the mobile user registers directly to the gateway, and the new visited AP_8 becomes the new CH of the new cluster. In this case, the new cluster managed by AP_8 is composed of six APs as shown in Fig. 1.

We can see that AP_7 belongs to both clusters managed by CHs AP_8 and AP_2. Thus, in contrast to the OSC approach, the clusters are no longer disjoint. Indeed with DCAM, the mobile user can be attached to different CHs when visiting the AP_7. According to the mobile user trajectory, i.e., the tuple (old AP, old CH), the new visited AP assigns the new CH to the user.

In our example, when the mobile user visits the subnet managed by AP_7 with the tuple (AP_6, AP_2), it registers to the CH AP_2. The new state of the mobile user becomes (AP_7, AP_2). On the other hand, when the user visits AP_7 while having (AP_8, AP_8) as the current state, it registers to the CH AP_8. The new user state is therefore (AP_7, AP_8). This simple example clearly shows the dynamic and distributed properties of the DCAM algorithm. More formally, the distributed clustering algorithm is described by the pseudocode in Fig. 2.

It is important to note that the DCAM process does not require each CH to be aware of its virtual cluster. The clusters are indeed not preestablished in advance as with the OSC scheme. The operations of DCAM are completely distributed and can be described as follows: 1) The mobile user maintains the information about the current CH. 2) Each time the mobile user moves to a new AP, it notifies the new visited AP with the identity of the current CH. 3) Based on this information, the new visited AP verifies whether or not it belongs to the virtual cluster of the current CH. In other words, it verifies if the delay and registration cost conditions are still satisfied. 4) If the above conditions are true, the new visited AP notifies the current CH with the new mobile user location. Otherwise, the new visited AP registers directly to the GW and becomes the new mobile user CH. In this case, the mobile user updates the identity of the CH.

It is worth noting that DCAM needs all the APs to maintain the information regarding the distances between them. This information can be either set up statically in advance at each AP, or discovered and disseminated using a distance vector routing protocol such as DSDV [21] or AODV [22]. As APs in WMN are not mobile components, each one maintains a quasi-static table containing the distance that separates it from each of the other APs of the WMN.

This table changes and therefore needs to be recalculated only when the physical topology changes, notably in cases of AP failure or when new APs are added. In such cases, the DCAM algorithm needs only the distance table D to be updated at each AP. Then, each AP automatically identifies its corresponding virtual cluster when it operates as a CH. In contrast, using the OSC approach, the optimal placement of the clusters must be recalculated centrally at the gateway according to the new topology (i.e., connectivity graph). This operation is time-consuming and may take from several hours to several days according to the WMN size. Moreover, the information regarding the clusters placement must be disseminated to each AP, which introduces additional signaling cost compared to the DCAM algorithm. In view of this, the DCAM algorithm presents several advantages compared to the OSC method from an operational flexibility point of view.

B. Analytical Model

In this section, we introduce a mathematical model based on Markov chains to evaluate the performance of the DCAM method in terms of RRU cost, registration updates cost, data delivery cost, and load balancing. The elaborated model will be also used to derive some performance metrics that are required as input to the ILP formulation of the OSC method.

Assume an arbitrary meshed wireless network composed of N APs denoted by AP_i $(i = 1, \dots, N)$, where AP_1 is the gateway. In our study, we consider a general two-dimensional (2-D) random walk mobility model. Accordingly, a mobile user connected to AP_i moves to one of the neighboring subnets with equal probability p (i.e., $p = 1/\delta(AP_i)$, where $\delta(AP_i)$ is the AP_i connectivity degree). Using these probabilities, we construct the transition probability matrix $P = [p_{ij}]$ between pairs of mesh nodes. This matrix is used as an input parameter to solve the ILP problem described in the previous section.

Let $X(t)$ be the user's state within the WMN at time t defined by the tuple (AP_i, AP_j), where AP_i is the current AP and AP_j is the current user's CH. The sojourn time of a mobile user in a subnet AP_i has a general distribution (not necessarily exponential) with a mean $1/\mu$. Moreover, the sojourn times of a mobile user in different subnets are independent and have the same mean. Since the mobile user evolves as a 2-D random walk, the process $X = \{X(t), t \geq 0\}$ is a homogeneous semi-Markov process with state space $\mathcal{S} = \{(AP_i, AP_j) | 1 \leq i \leq N, AP_j \in E_{AP_i}\}$, where E_{AP_i} is the set of possible CHs that a mobile user can register to when it is connected to AP_i. In other words, $AP_j \in E_{AP_i}$ if and only if AP_i belongs to the cluster managed by AP_j, i.e., it satisfies the following relation:

$$AP_j \in E_{AP_i} \text{ if and only if } \begin{cases} d(AP_i, AP_j) < d(AP_i, GW) \\ \text{and} \\ d(AP_i, AP_j) + d(AP_j, GW) \leq D_{\max}. \end{cases} \tag{6}$$

Recall that $d(x, y)$ denotes the shortest path distance (in terms of number of hops) between x and y. Note that the first condition in (6) ensures a *CH registration* cost is cheaper than a *GW registration*. The second condition ensures that the mobile user fulfills the delay constraint.

We denote by $T_0, T_1, T_2, \dots$ the successive times of transitions for X, and by $Z_0, Z_1, Z_2, \dots$ the successive states visited between these transitions, i.e., for every $k \geq 0$, $Z(k) = X(t)$ if $T_k \leq t < T_{k+1}$, where $T_0 = 0$. According to [23], the embedded process $Z = \{Z(k), k \geq 0\}$ in the transition instants of X is a discrete-time homogeneous Markov chain with state space $\mathcal{S}$ and transition probability matrix denoted by Q whose transition probabilities are given below.

In the following, we calculate the transition probability matrix Q of the process Z when leaving a generic state (AP_i, AP_j). Let $AP_{i'}$ denote the next visited AP by the mobile user (i.e., $AP_{i'} \in N_e(AP_i)$). Hence, the mobile user moves to subnet $AP_{i'}$ with a probability $p = 1/\delta(AP_i)$. According to whether or not $AP_{i'}$ belongs to the current cluster managed by the CH AP_j, we can identify the next user's state.

Specifically, if this is the case (i.e., $AP_j \in E_{AP_{i'}}$), the mobile user will transit to the state $(AP_{i'}, AP_j)$. In this case, the mobile user performs a local registration to the current CH AP_j. Henceforth, we denote by $\mathcal{A}$ the event that $AP_j \in E_{AP_{i'}}$ [see (6)]. On the other hand, if $\mathcal{A}$ is not satisfied (i.e., $\bar{\mathcal{A}} = AP_j \notin E_{AP_{i'}}$), the mobile user registers to the gateway, and the new AP becomes the CH of the new cluster. As such, the mobile user transits to state $(AP_{i'}, AP_{i'})$. Consequently, we have

$$\begin{cases} Q\left((AP_i, AP_j), (AP_{i'}, AP_j)\right) = p \cdot 1_{\mathcal{A}} \\ Q\left((AP_i, AP_j), (AP_{i'}, AP_{i'})\right) = p \cdot 1_{\bar{\mathcal{A}}} \end{cases}$$

where $1_{\mathcal{A}}$ (respectively $1_{\bar{\mathcal{A}}}$) is the indicator function of the condition $\mathcal{A}$ (respectively $\bar{\mathcal{A}}$), i.e., it is equal to 1 if the condition $\mathcal{A}$ (respectively $\bar{\mathcal{A}}$) is true, and 0 otherwise.

Based on the above analysis, we derive the transition probability matrix Q. The Markov chain Z being irreductible and aperiodic and the sojourn times of X having the same mean, we have (see, for instance, [23]), for every $h, s \in \mathcal{S}$,

$$\lim_{t \longrightarrow \infty} \Pr\{X(t) = s | X(0) = h\} = \pi_s$$

where $\pi = [\pi_s]$ is the steady-state distribution of the Markov chain Z, which satisfies

$$\pi Q = \pi \quad \text{and} \quad \sum_{s \in \mathcal{S}} \pi_s = 1. \tag{7}$$

As such, we get the steady-state probabilities π of the process X. It is worth noting that using the vector π, we can derive the steady probability $\Pi = [\Pi_i]$ that the mobile user is connected to the AP_i $(i = 1, \dots, N)$. This vector Π, is used as an input parameter in the ILP formulation of the OSC approach. This vector can be derived as follows, using the steady-state probabilities $\pi_s = \pi(AP_i, AP_j)$ of the states $s = (AP_i, AP_j)$:

$$\Pi_i = \Pi(AP_i) = \sum_{AP_j \in E_{AP_i}} \pi(AP_i, AP_j). \tag{8}$$

Building on these results, we evaluate hereafter the performance of the DCAM and OSC methods. We will derive first the RRU cost. To do so, we calculate the data delivery and signaling costs. Moreover, we derive the load-balancing cost between the clusters (i.e., CH nodes).

1) Data Delivery Cost: It denotes the wireless link usage in the WMN by data packets. It is the average number of wireless links used for packet delivery between the gateway and the current serving AP (i.e., in the downlink direction). In both DCAM and OSC methods, packets destined to the mobile user have to pass through the CH due to the clustering process. Hence, the data delivery cost metric can be written as follows:

$$Data\_Delivery\_Cost(\mathrm{DCAM}) = \sum_{s\in\mathcal{S}} \pi_s \times \left(d\left(AP(s), CH(s)\right) + d\left(CH(s), GW\right)\right) \quad (9)$$

where $s = (AP(s), CH(s))$.

For the OSC method, the data delivery cost can be simply given by substituting (5) in (4)

$$Data\_Delivery\_Cost(\mathrm{OSC}) = \sum_{i=1}^{N} \left(\Pi_i \times \sum_{j=1}^{N} b_{j,i} \times [d(i,j) + d(j, GW)]\right) \quad (10)$$

where Π_i is given by (8), and $b_{j,i}$ is an output of the ILP problem resolution defined in Section IV.

In the case where clustering is not considered, packets are delivered using the shortest path between the gateway and the mobile user. Hence, the data delivery cost is given by

$$Data\_Delivery\_Cost(\mathrm{WC}) = \sum_{i=1}^{N} \Pi_i \times d(i, GW). \quad (11)$$

2) Registration Updates Cost: It denotes the signaling cost of registration updates when a handoff occurs. It is representative of the average number of occupied wireless links by the signaling messages exchanged in the WMN when the mobile user moves to a new AP. In both DCAM and OSC methods, a local registration (i.e., *CH registration*) is required as long as the mobile user remains in the same cluster. Otherwise, a *GW registration* is performed.

Considering the DCAM method, the average registration updates cost when leaving the state $s = (AP(s), CH(s))$ to a state $s' = (AP(s'), CH(s'))$ can be written as follows, using the transition probability matrix Q:

$$Update\_Cost(s) = \sum_{s'\in\mathcal{S}} Q(s, s') \times \mathcal{C}(s, s') \quad (12)$$

where

$$\mathcal{C}(s,s') = \begin{cases} cost_{CH}(s') & \text{if } CH(s') = CH(s) \\ cost_{GW}(s') & \text{if } CH(s') \neq CH(s) \end{cases}$$
$$\text{with} \quad \begin{cases} cost_{CH}(s') = d\left(AP(s'), CH(s')\right) \\ cost_{GW}(s') = d\left(AP(s'), GW\right). \end{cases}$$

The total registration updates cost can be thus written as

$$Reg\_Update\_Cost(\mathrm{DCAM}) = \sum_{s\in\mathcal{S}} \pi_s \times Update\_Cost(s). \quad (13)$$

Considering the OSC method, the registration updates cost can be expressed as follows by replacing (3) in (2):

$$Reg\_Update\_Cost(\mathrm{OSC}) = \sum_{i=1}^{N} \left(\Pi_i \times \sum_{j=1}^{N} P(i,j) \times \sum_{k=1}^{N} b_{k,j} \times [d(k,j) + (b_{k,j} - b_{k,i}) d(k, GW)]\right). \quad (14)$$

When the clustering policy is not considered, the registration updates cost is given by

$$Reg\_Update\_Cost(\mathrm{WC}) = \sum_{i=1}^{N} \Pi_i \times Update\_Cost(i) \quad (15)$$

where $Update\_Cost(i)$ is the cost of registration updates when the mobile user moves to AP_i, which is simply equal to $d(AP_i, GW)$. Hence, the registration updates cost can be written as follows:

$$Reg\_Update\_Cost(\mathrm{WC}) = \sum_{i=1}^{N} \Pi_i \times d(AP_i, GW). \quad (16)$$

It is easy to see that, in the WC case, the registration updates cost shown in (16) is equal to the data delivery cost shown in (11). This is simply because both packet delivery and registration updates are always performed with the same node, which is the gateway.

3) RRU Cost: It is the radio resource utilization of a WMN where the profile of its mobile users are characterized by the pair (λ, μ). λ measures the average rate of traffic exchanged by each mobile user, and μ describes the user mobility in the WMN. The expression of the RRU_Cost for the DCAM, OSC, and WC cases can be simply obtained by replacing (9) and (13) in (1), (10) and (14) in (1), and (11) and (16) in (1), respectively.

4) Load-Balancing Cost: We now analyze the load balancing in the WMN between clusters when clustering technique is used. We compare the load balancing in the OSC and the DCAM methods. The load balancing is defined as the variance of the traffic load handled by the different clusters. It is worth noting that the lower the variance we get, the more efficient the load balancing is. We denote by w_i the proportion of traffic handled by the cluster identified by its CH AP_i $(i = 1, \ldots, N)$. w_i can be written as follows according to whether the OSC or the DCAM approach is considered:

$$w_i^{\mathrm{OSC}} = \sum_{j=1}^{N} b_{i,j} \times \Pi_j \quad \forall i = 1, \ldots, N \quad (17)$$
$$w_i^{\mathrm{DCAM}} = \sum_{\substack{1\leq j\leq N \\ AP_i \in E_{AP_j}}} \pi(AP_j, AP_i) \quad \forall i = 1, \ldots, N. \quad (18)$$

Note that according to (17) and (18), $w_i \neq 0$ if the AP_i is a CH; otherwise $w_i = 0$. Hence, the number K of clusters according to each clustering scheme is

$$K^{\mathrm{OSC}} = \sum_{i=1}^{N} 1_{w_i^{\mathrm{OSC}}} \quad (19)$$
$$K^{\mathrm{DCAM}} = \sum_{i=1}^{N} 1_{w_i^{\mathrm{DCAM}}} \quad (20)$$

where 1_{w_i} is the indicator function of the condition $w_i > 0$. Let the vector $W' = [w'_i], i = 1, \ldots, K$, be the subset of $W = [w_i]$, $i = 1, \ldots, N$, representing only the traffic load of the CH APs. As stated before, the load balancing is defined as the variance of the variable w'_i, $i = 1, \ldots, K$. Note that $\sum_{i=1}^{K} w'_i = 1$.

Hence, $E[W'] = 1/K$, and the load balancing between clusters according to each clustering approach is given by

$$LB(\mathrm{OSC}) = \frac{1}{K^{\mathrm{OSC}}}\left(\sum_{i=1}^{K^{\mathrm{OSC}}}\left(w_i'^{\mathrm{OSC}}\right)^2\right) - \left(\frac{1}{K^{\mathrm{OSC}}}\right)^2$$
$$= \frac{1}{K^{\mathrm{OSC}}}\left(\sum_{i=1}^{N}\left(w_i^{\mathrm{OSC}}\right)^2\right) - \left(\frac{1}{K^{\mathrm{OSC}}}\right)^2 \quad (21)$$

$$LB(\mathrm{DCAM}) = \frac{1}{K^{\mathrm{DCAM}}}\left(\sum_{i=1}^{K^{\mathrm{DCAM}}}\left(w_i'^{\mathrm{DCAM}}\right)^2\right) - \left(\frac{1}{K^{\mathrm{DCAM}}}\right)^2$$
$$= \frac{1}{K^{\mathrm{DCAM}}}\left(\sum_{i=1}^{N}\left(w_i^{\mathrm{DCAM}}\right)^2\right) - \left(\frac{1}{K^{\mathrm{DCAM}}}\right)^2. \quad (22)$$

VI. Numerical and Simulation Results

In this section, we evaluate the efficiency of the proposed clustering algorithms (i.e., OSC and DCAM). Specifically, we study the gain they introduce compared to the DCA algorithm [4] as well as to the case where clustering is not used (i.e., the WC case), under various mobility and traffic scenarios. The OSC, DCA, and WC cases are also used as baselines to which the DCAM improvements are compared. The analysis is based on both regular and arbitrary meshed topologies. In the first case, we used a hexagon-based regular wireless mesh topology as depicted in Fig. 1. Accordingly, the gateway is located at the center of the WMN and is surrounded by R rings of subnets (i.e., APs). The number N of APs in the WMN is therefore $N = 1 + 3R(R+1)$. In our results, we considered different values of the radius R to represent small, medium, and large WMNs. Specifically, R is varied from 1 to 5, i.e., N is varied from 7 to 91. Moreover, the maximum tolerable delay $D_{\max}$ inside the WMN is set equal to R.

In the second case of study, the performance of OSC, DCAM, DCA, and WC methods is evaluated using arbitrary meshed WMNs. Again, we considered different network sizes: 20, 60, and 100 APs, which are representative of small, medium, and large single-gateway WMNs. Due to space constraints, we present results only for 20-node WMNs. The N APs are distributed randomly in the network area. Then, based on the transmission range R_t, the connectivity graph is derived. It is worth noting that the resulting connectivity graph is considered only if it is connected, i.e., there is at least a path that connects each AP to the gateway. Otherwise, we keep generating random topologies until the graph connectivity condition is satisfied. Once a valid arbitrary meshed topology is obtained, we can also modify the connectivity graph by increasing R_t. Obviously, the resulting graph is always connected, and in doing so, we increase the average node degree $\bar{\delta}$. The rational behind this is to study the impact of the average node degree $\bar{\delta}$ on the evaluated metrics. In our experiments, the maximum tolerable delay $D_{\max}$ inside the network is set equal to the maximal distance between the gateway and any AP. In other words, $D_{\max} = \max_{i=1,\ldots,N} d(AP_i, GW)$.

TABLE I
Quantity of Downloaded Traffic per User per Day

λ (packets/s)	0.001	0.01	0.1	0.5	1	2	10
Traffic (Mbits/day)	0.318	3.18	31.8	159	318	636	3180

We recall that, in our study, we use a random walk mobility model. Accordingly, a mobile user connected to AP_i moves to one of the neighboring subnets with equal probability p (i.e., $p = 1/\delta(AP_i)$, where $\delta(AP_i)$ is the AP_i connectivity degree). In our experiments, the sojourn time $1/\mu$ within an AP subnet is set equal to 10, 100, or 1000 s to represent very fast, fast, or slow mobile users, respectively. In addition, we used different values of λ to represent different mobile user loads (i.e., light, moderate, and heavy-traffic mobile user loads). Specifically, the values of λ and the associated quantity of downloaded traffic per user and per day are reported in Table I. To get an estimate from λ of the downloaded traffic per user, we assume that the average packet size is equal to 460 B (see [24]).

Note that in order to evaluate the performance of the different clustering strategies by simulations and to validate the proposed analytical model, we both used ns-2 [25] and developed our own discrete-event simulator. Our simulations are run until a very narrow 97.5% confidence interval is achieved. Once this objective is accomplished, the simulations stop automatically. It is worth noting that to achieve such very narrow confidence intervals, simulations need to be run in certain cases over several hours up to more than one day, notably when the number of APs is large. We have noticed also that the simulation results converge in time to the analytical results, which are obtained in a few seconds. Hence, in addition to its accuracy and simple implementation compared to simulations, our elaborated analytical model enables significant time savings.

In this case, the perfect fit between the simulation and analytical results can be explained as follows.

- In our analysis, we did not make any assumption regarding the distributions of the sojourn times of the mobile users inside different subnets. The sojourn time indeed has a general distribution (not necessarily exponential) with a mean $1/\mu$.
- In our study, we considered a multichannel WMN where interfering wireless links operate on different channels, thus enabling multiple contentionless parallel transmissions. Obviously, frequencies can be reused in the network by noninterfering links. In such environment where collisions and interferences between transmissions over interferer links are avoided, the elaborated analytical model reflects exactly the real behavior of the WMN. It is worth noting that in a multichannel environment, interference may still exist between interferer links if the number of available orthogonal channels is insufficient. For example, according to the NetX testbed [26], five orthogonal channels among 12 are only available for use in the IEEE 802.11a standard. One way to ensure an interference-free environment is to therefore operate with the IEEE 802.16d or IEEE 802.16e standards, which provide 256 and up to 2048 orthogonal channels, respectively [27], [28]. A typical scenario to avoid interference between links in highly connected networks is indeed to use the IEEE

TABLE II
COMPUTATION TIMES OF OSC AND DCAM METHODS

$R(N)$	1(7)	2(19)	3(37)	4(61)	5(91)
OSC Time (seconds)	0.047	0.187	33.218	674.078	6570.77
DCAM Time (seconds)	0.31	0.98	1.97	7.81	26.05

802.16 standard (i.e., WIMAX) for the WMN backbone, and the IEEE 802.11 standard to carry traffic over the last hop to the user.

For the clarity of the presentation, in the reminder of this paper, unless otherwise mentioned, the reported curves for DCAM correspond to results obtained using the analytical model.

As mentioned before, we will first compare the OSC and DCAM clustering strategies based on their RRU costs and time complexities. To get an insight into the computation time needed by the clustering methods to identify the set of clusters and their associated CHs, let us consider Table II. The computation time of the OSC and DCAM methods is calculated for the regular hexagon-based wireless mesh topology case. These measurements are performed on a PC with 3.2 GHz of CPU and 2.00 GB of RAM. In the table, the network radius R is varied from 1 to 5, which corresponds to a variation of the number of APs from 7 to 91. The reported results show that the DCAM algorithm achieves great time saving compared to the OSC approach, notably when R is high. Indeed, the time needed to resolve the OSC optimization clustering problem increases dramatically with the network size N since the number of variables and equations in the ILP formulation increases exponentially with N. In contrast, with DCAM, the clusters are set up automatically and in a distributed manner according to both conditions regarding the delay constraint and the registration updates cost [see (6)]. In DCAM, each AP needs only to be aware of the distance table D to identify its virtual cluster when it operates as a CH. In view of this, the changes in the physical topology are more efficiently handled by the DCAM strategy since it reacts much more quickly than the OSC strategy.

Let us now focus on the comparison among the different strategies based on their optimal RRU costs. This comparison is achieved using both regular and arbitrary wireless mesh topologies, large and small networks (i.e., by varying N), low- and highly connected networks (i.e., by varying $\bar{\delta}$ as well as the variance in δ), and under various mobile user profiles characterized by the pair (λ, μ). The results are reported in Figs. 5–7. Before delving into the exploration of these graphs, let us start by analyzing the results regarding the data delivery and registration updates costs associated with the derived RRU cost according to the different strategies. Recall that the RRU cost of a mobile user is simply a weighted sum of these two terms [see (1)].

A. Data Delivery Cost Results

Fig. 3(a) and (b) show the returned data delivery cost following to the RRU cost minimization according to the different strategies for both the hexagon-based regular WMN and 20-node random WMN, respectively. Fig. 3(a) plots the data delivery cost as a function of the network radius R, which is varied between 1 and 5. In Fig. 3(b), the data delivery cost is plotted as a function of the average node degree $\bar{\delta}$.

The first thing to note is that the data delivery cost is insensitive to both the user mobility speed μ and its traffic load λ.

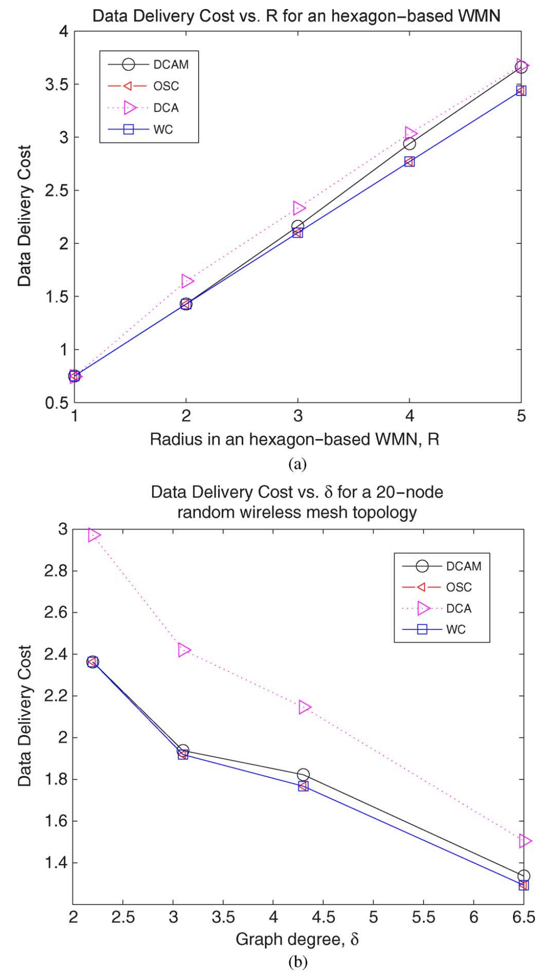

Fig. 3. Data delivery cost. (a) Hexagon-based regular WMN. (b) 20-node random WMN.

Moreover, with DCAM, OSC, and DCA methods, packets destined to the mobile user have to pass through the current CH before reaching their destination due to the clustering process. In contrast, when clustering is not considered, packets are delivered directly using the shortest path between the gateway and the mobile user. As such, the data delivery cost is minimal when clustering is ignored, as depicted in Fig. 3. This cost increases slightly with DCAM due to the additional cost introduced by the clustering process. However, we notice that, using the OSC approach, the data delivery cost does not increase compared to the WC optimal case although clustering is used. This means that with the OSC approach, the CH associated to a given AP_i is always chosen to be an AP in the shortest route between the AP_i and the gateway. In this regard, the optimal solution returned by the ILP resolution and that minimizes the RRU cost in the OSC case also minimizes the data delivery cost term included therein. In contrast, the *ad hoc* clustering algorithm DCA provides the worst data delivery cost, which confirms that it is not really appropriate for WMNs.

Finally, we can observe in Fig. 3(a) that the data delivery cost increases with the network size under all the strategies. This is because the average distance between the gateway and an AP increases with the network radius R. In addition, we can observe from Fig. 3(b) that the data delivery cost decreases when the

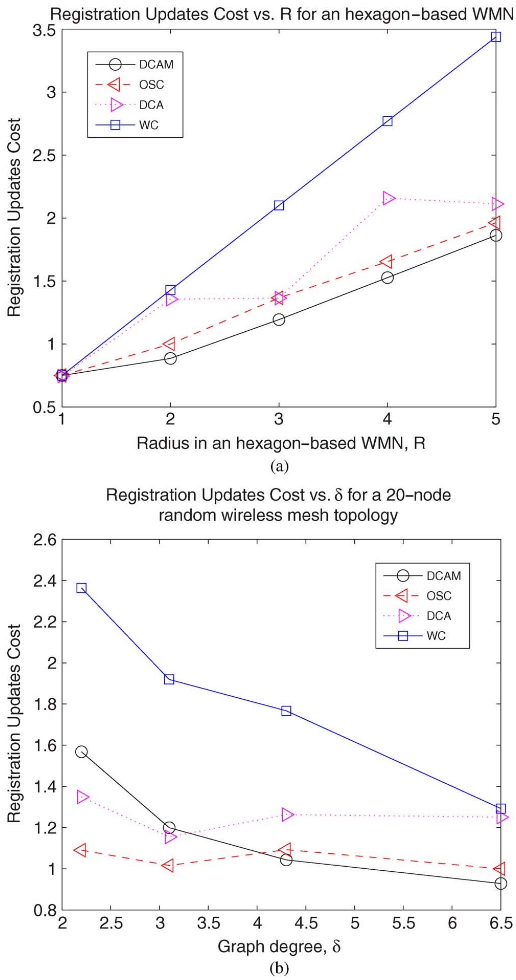

Fig. 4. Registration updates cost. (a) Hexagon-based regular WMN. (b) 20-node random WMN.

average node degree $\bar{\delta}$ increases. This is simply because the average distance between an AP and the gateway decreases with the increase of $\bar{\delta}$. This can also be verified through the variation of the Moore bound derived in [29], which relates the average shortest path distance (h) between two nodes in an arbitrary meshed network with an average node degree $\bar{\delta} \geq 2$

$$h \approx \frac{\ln\left[(N-1)\frac{(\bar{\delta}-2)}{\bar{\delta}}+1\right]}{\ln(\bar{\delta}-1)}. \tag{23}$$

B. Registration Updates Cost Results

Fig. 4(a) and (b) plot the obtained registration updates costs following the RRU minimization when using the regular hexagon-based and the arbitrary meshed topologies, respectively. Similar to the data delivery cost, the registration updates cost is insensitive to λ and μ.

Fig. 4(a) plots the registration updates cost as a function of the network radius R. We can observe that the WC approach has the highest registration updates cost due to the required expensive *GW registration* at every handoff. In this case, the registration updates cost is equal to the data delivery cost. Indeed, both packet delivery and registration updates are always performed with the gateway. Clustering alleviates this issue since some expensive *GW registrations* are replaced by low-cost *CH registrations*. In this regard, the registration updates cost decreases significantly with OSC and DCAM approaches as well as with DCA, notably when R is large. This is because the difference in cost between a *CH* and a *GW registration* increases with R. Moreover, we can see that the registration updates cost increases with R for all the strategies since the average *CH* and *GW registration* costs increase with the radius R.

With regard to the comparison between the OSC and DCAM approaches, we can see that the DCAM algorithm outperforms the OSC method in this particular case (i.e., the hexagon-based regular wireless mesh topology). This is because for mesh topologies with relatively high connectivity degrees $\bar{\delta}$, the DCAM algorithm provides the lowest registration updates cost as shown in Fig. 4(b), where we plot the registration updates cost as a function of the average node degree $\bar{\delta}$. Indeed, we can see in this figure that this metric decreases with $\bar{\delta}$ for all the different strategies. This is simply because the average distance between nodes decreases with $\bar{\delta}$.

In addition, Fig. 4(b) shows that the minimal registration updates cost is obtained either by DCAM or OSC schemes according to the value of $\bar{\delta}$. Specifically, for small and moderate values of $\bar{\delta}$, the OSC method stands out as the best choice; otherwise, for high values of $\bar{\delta}$, the DCAM algorithm is the best choice. For instance, when $N = 20$, the DCAM algorithm provides the lowest cost when $\bar{\delta} > 4$, otherwise the OSC method gives the best cost. The rational behind this can be explained as follows. Using the DCAM approach, the cluster size in terms of number of APs around each CH increases with the average node degree $\bar{\delta}$ since more and more APs satisfy the delay and the registration updates cost constraints shown in (6) with respect to the considered CH. As such, more and more expensive *GW registrations* are replaced by the cheap *CH registrations* when $\bar{\delta}$ increases. Hence, the registration updates cost decreases considerably with $\bar{\delta}$ for the DCAM case as shown in Fig. 4(b). However, the decrease of this cost with $\bar{\delta}$ is less significant in the OSC case since the OSC approach tries to minimize the registration updates cost without deteriorating the data delivery cost. In this regard, the gain in terms of registration updates cost for DCAM over OSC when $\bar{\delta}$ is high is achieved at the expense of a worse data delivery cost.

C. RRU Cost Results

In this subsection, we analyze the RRU cost results. To achieve this, we will proceed as follows. We will first identify the cases where clustering is useful. Then, when this is the case, we will compare among the OSC, DCAM, and DCA clustering approaches.

First, we recall that the RRU cost is simply a weighted sum of the already discussed data delivery and registration updates costs. In the WC case, the RRU cost equals the data delivery cost and the registration updates cost. It is simply the average length of the shortest paths between each AP_i $(i = 1, \ldots, N)$ and the gateway weighted by the probability Π_i that the mobile user is connected to AP_i during its movement within the WMN. Hence, as shown in Figs. 5 and 6, the RRU cost with the WC approach is insensitive to the mobile user profiles (i.e., λ and μ) and is a constant metric that depends only of the physical topology (i.e., graph connectivity properties).

In the previous subsections, we have proven that ignoring clustering enables the lowest data delivery cost. Moreover, we

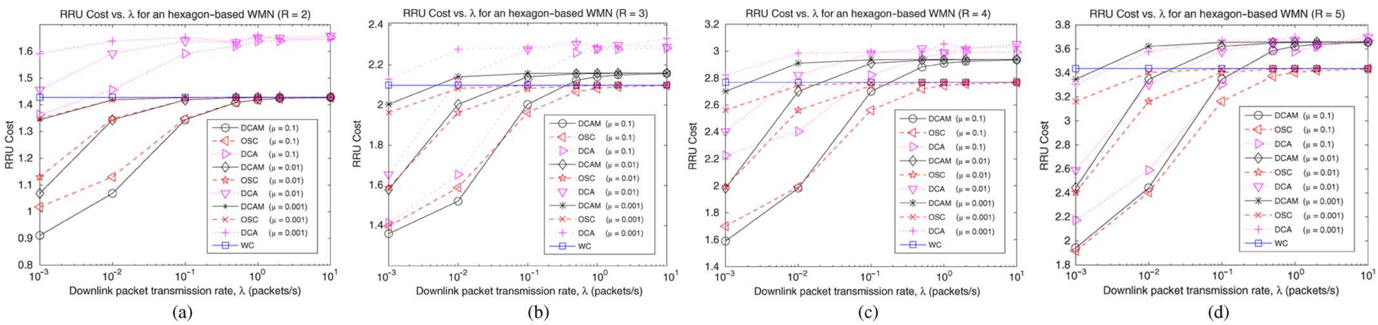

Fig. 5. RRU cost in a hexagon-based regular wireless mesh network. (a) Radius $R = 2$. (b) Radius $R = 3$. (c) Radius $R = 4$. (d) Radius $R = 5$.

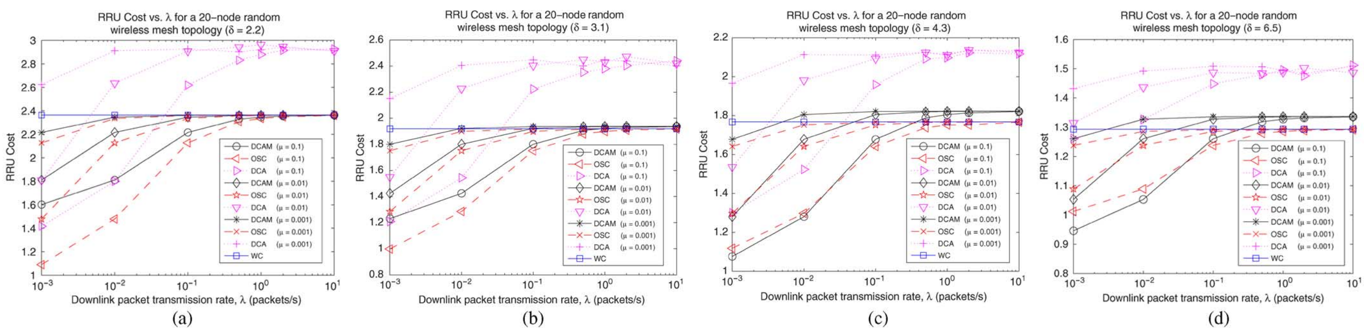

Fig. 6. RRU cost in a 20-node random wireless mesh network. (a) $\bar{\delta} = 2.2$. (b) $\bar{\delta} = 3.1$. (c) $\bar{\delta} = 4.3$. (d) $\bar{\delta} = 6.5$.

have shown that the main aim of clustering is to reduce the registration updates cost, probably at the expense of a slight increase of the data delivery cost. Intuitively, when the amount of signaling messages that a mobile user generates due to its mobility is negligible, i.e., the registration updates cost has a minor impact on the total RRU cost, then clustering is not helpful. In other words, if $2\,\mu \times m_{\mathrm{sig}} \ll \lambda \times m_{\mathrm{data}}$, clustering does not provide any gain. In view of this and as a first main finding, we can see through Figs. 5 and 6 that when λ and $1/\mu$ are high, the WC approach stands out as the best solution. Clustering becomes interesting only for small and moderate values of λ and $1/\mu$. This range of μ implies short sojourn times of the mobile users in each AP subnet, which results in frequent handoffs, thus increasing the signaling traffic load. Accordingly, the registration updates cost is no longer a negligible cost, and thus both OSC and DCAM achieve a significant gain compared to the WC case.

It is worth noting that the OSC approach always achieves better results than the WC case. In this context, we point out that the WC strategy can be seen as a particular case of the OSC one where the number of clusters is set equal to 1 and where the gateway is the only CH. In this regard, both strategies exhibit similar results when λ and $1/\mu$ have relatively large values. This is shown clearly in Figs. 5 and 6 by the asymptotic behavior of the OSC curves with respect to the WC constant (i.e., horizontal) one. Obviously, in this range of λ and μ, the WC technique is preferred thanks to its simplicity compared to the OSC method since they yield similar results.

Let us now focus on the comparison among the OSC, DCAM, and DCA clustering techniques. Beforehand, let us recall that OSC provides the optimal RRU cost when the WMN is divided into disjoint clusters. With DCAM, clusters may overlap. Essentially, the DCAM approach is proposed to alleviate the time complexity of the OSC approach as proven at the beginning of this section. In what follows, we will see if DCAM gives also reasonable results compared to the OSC one.

The first thing to note through Figs. 5 and 6 is that the minimal RRU cost is always obtained either by our OSC or DCAM schemes according to the network graph properties (i.e., network size N and average node degree $\bar{\delta}$). Our proposed clustering protocols indeed outperform the DCA scheme, which is proposed originally for ad hoc networks. Hereafter, we investigate the impact of the WMN topology properties on the relative performance of the OSC and DCAM strategies.

1) Impact of the Network Size N: Fig. 5 shows that the DCAM approach provides the best RRU cost when the network size is small or moderate (i.e., cases where $R = 2$ and $R = 3$). Recall that the DCAM algorithm almost always provides the lowest registration updates cost as shown in Fig. 4. However, this is achieved at the expense of a higher data delivery cost compared to the OSC case as depicted in Fig. 3. Hence, the clustering technique that accomplishes the best tradeoff between the registration updates and data delivery costs, will provide the lowest RRU cost. Typically, when the network size is small or moderate, the loss of the DCAM approach over the OSC approach in terms of data delivery cost is tiny since the distances between nodes are relatively short. As such, the gain that the DCAM approach achieves regarding the registration updates cost is prevailing, and thus DCAM provides a better RRU cost.

2) Impact of the Average Node Degree $\bar{\delta}$: The same reasoning holds when analyzing the impact of the average node degree $\bar{\delta}$. For instance, let us consider Fig. 6, where we have an arbitrary meshed WMN of $N = 20$ nodes. We can observe that for small values of the average node degree $\bar{\delta}$ (i.e., $\bar{\delta} = 2.2$ and $\bar{\delta} = 3.1$), the OSC approach provides the best results, and when $\bar{\delta}$ gets moderate to high values (i.e., $\bar{\delta} = 4.3$ and $\bar{\delta} = 6.5$), the DCAM approach is the best solution. Indeed, increasing $\bar{\delta}$ reduces the distances in terms of hops between the mesh nodes. Hence, the loss of the DCAM algorithm over the OSC approach in terms of data delivery cost becomes tiny. As a result, the gain that the DCAM approach achieves regarding the registra-

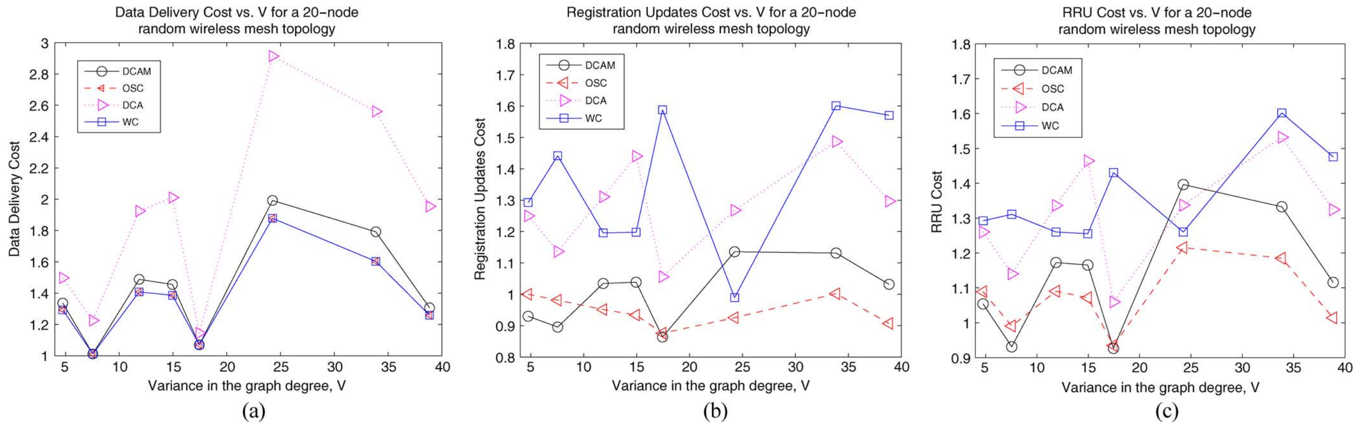

Fig. 7. Impact of the variance in the connectivity degree of nodes on the performance metrics of a 20-node random WMN. (a) Data delivery cost. (b) Registration updates cost. (c) RRU cost.

tion updates cost dominates the loss in terms of data delivery cost. Thus, DCAM provides a lower RRU cost.

3) Impact of the Variance in the Connectivity Degree of Nodes: To provide an in depth analysis of the impact of the network topology on the performance of the different clustering schemes, we also studied the impact of the variance in the connectivity degree of nodes. This indeed enables a more meticulous specification of the network topology and therefore gives more accurate results for the service provider to choose among the clustering schemes according to its specific network topology.

To achieve this, given a WMN with a set of static APs, we varied the variance in the nodes' connectivity degrees while conserving the same resulting average node degree $\bar{\delta}$. Obviously, following each modification in the network topology, we verify that it is still connected. The results are shown in Fig. 7, where we considered a 20-node arbitrary meshed WMN. In this case, $\lambda = 0.01$ and $\mu = 0.1$, and the average node degree $\bar{\delta}$ is maintained equal to 6.5. Two main observations can be made. First, for networks with small variance in the connectivity degree of nodes, the OSC approach stands out as the best choice for service providers with relatively low-connected networks, and the DCAM approach is the best choice for moderate- and highly connected networks. Second, increasing the variance in the connectivity degree of nodes favors the OSC scheme. Indeed, the gain of the OSC approach over the DCAM approach achieved at the highly connected areas of the network in terms of data delivery cost [see Fig. 7(a)] dominates the loss in terms of registration updates cost [see Fig. 7(b)] notably at the low-connected areas of the network.

4) Summary of the Results: In summary, we can state that for the special cases of high values of λ and $1/\mu$, clustering is not useful. In the remaining ranges of λ and $1/\mu$, clustering achieves significant gain regarding the radio resource utilization. The best cost is always provided either by our OSC or DCAM strategies thus outperforming the DCA scheme. Specifically, for both small-to-moderate-sized and highly connected networks, the DCAM clustering approach stands out as the best clustering solution. For large-sized and low-connected networks, the OSC clustering technique provides the best RRU cost. However, such solution is time-consuming and may not react efficiently to the physical network topology changes (AP failure, new AP addition, etc.), as opposed to the DCAM approach, which adapts instantaneously to network changes. To conclude, we can see that by relaxing the constraint on the disjoint clustering, the DCAM approach does not limit its gain over the OSC approach to the reduction of the computation time, but it can achieve further gain regarding the RRU cost.

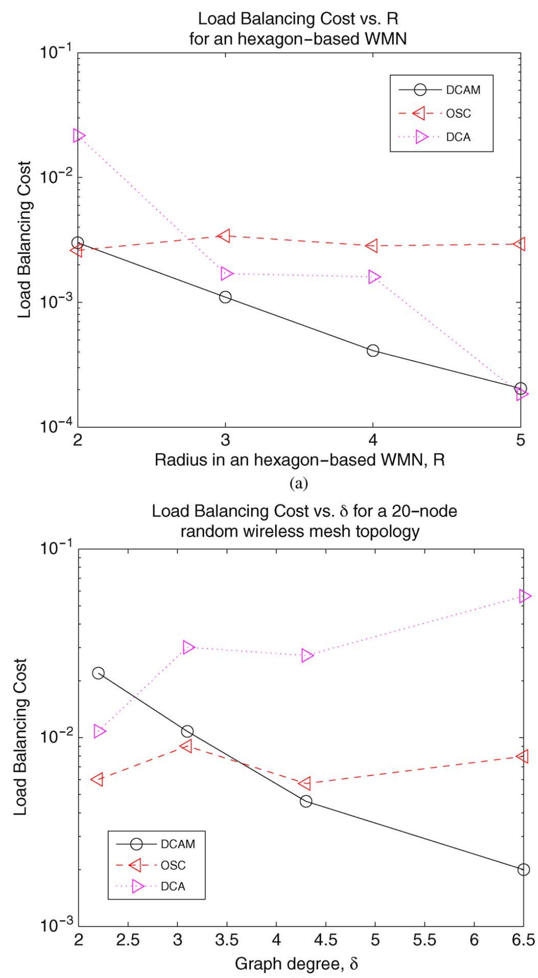

Fig. 8. Load balancing. (a) Hexagon-based regular WMN. (b) 20-node random WMN.

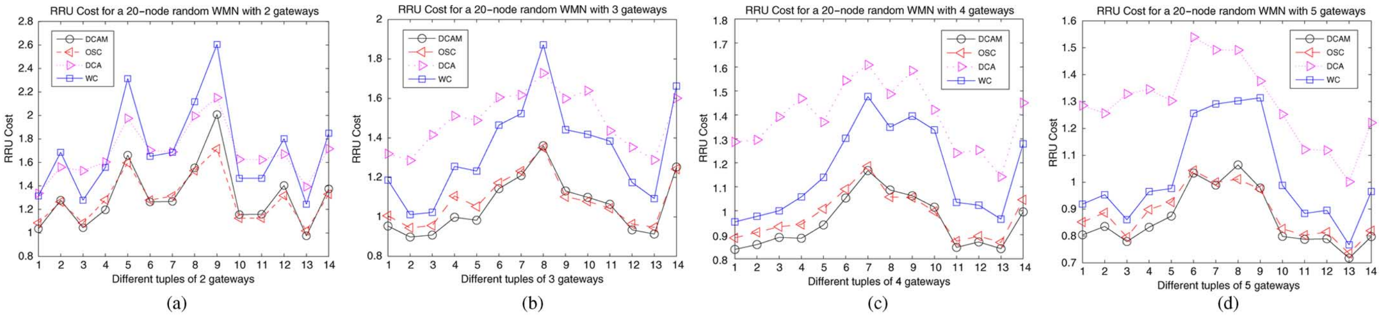

Fig. 9. RRU cost in a 20-node random WMN using multiple gateways. (a) WMN with two GWs. (b) WMN with three GWs. (c) WMN with four GWs. (d) WMN with five GWs.

D. Load-Balancing Results

This subsection investigates the resulting load balancing in the WMN following the utilization of the OSC, DCAM, and DCA clustering schemes. Fig. 8(a) and 8(b) illustrate the load balancing for regular hexagon-based and arbitrary meshed topologies, respectively. Recall that this metric is defined as the variance of the traffic load handled by the different clusters. We can observe from these figures that the DCAM algorithm often provides the lowest load-balancing cost. This is simply because the number of clusters in the DCAM approach is much greater than in the OSC and DCA cases. Typically, each AP can act as a CH in the DCAM case. As such, the traffic is better distributed and balanced among the network APs. As a result, the traffic load in DCAM is distributed more efficiently among the clusters than in the OSC and DCA cases, which reduces the probability of congestions in the WMN.

So far, this work focused on virtual clustering inside single-gateway WMNs and can be viewed as a complementary study to our previous work [20], which addressed the gateway placement problem. Next, we extend the results to the case of multiple-gateway WMNs.

E. Extension to the Multiple-Gateway WMN Case

Fig. 9 shows the RRU cost provided by the different clustering schemes when varying the number of gateways from two to five in a 20-node random WMN. For instance, when the number of gateways is set equal to two [Fig. 9(a)], we randomly turn different pairs of APs into gateways and evaluate the algorithms. In this particular case, $\lambda = 0.01$ and $\mu = 0.1$.

Based on these results, two main observations can be made. First, our clustering algorithms always achieve significant gain compared to the DCA and WC cases. The DCAM and OSC schemes provide comparable results. As such, DCAM represents a sensible solution for WMNs since it alleviates the time complexity entailed by the OSC approach. As a second finding, we can observe that the RRU costs of the different algorithms decrease with the increase of the number of gateways since, in doing so, shorter and better paths can be found to connect each AP to a wired backbone.

It is worth noting that considering the multiple-gateway case, the OSC ILP formulation and the DCAM Markov-chain-based model should be changed as follows. The variable $d(i, GW)$ should be replaced by $d(i, GW(i))$ in all the formulations, where $GW(i)$ denotes the gateway associated to AP_i, typically the closest gateway to AP_i.

F. Results Discussion

Based on the above performance evaluation study, we can see that clustering is not helpful when the signaling overhead represents an insignificant portion of the overall exchanged traffic generated in a WMN, typically for WMNs with slow mobile users and high data traffic loads. The key question here is: What does insignificant portion represent?

Figs. 5 and 6 show that clustering is not useful when the mobile user traffic load λ exceeds 1 (i.e., the mobile user exchange more than 318 Mbit/s per day) even with relatively fast mobile users (i.e., the sojourn time $1/\mu$ within an AP subnet is set equal to 10 seconds). In such cases, the signaling overhead represents at most $\alpha = 2.04\%$ of the total generated traffic, where α is given by (1).

In other words, in the range of λ and μ where the resulting $\alpha \leq 2\%$, clustering is not meaningful. On the other hand, we can see that clustering starts achieving gains when $\lambda \leq 0.1$ and $\mu \geq 0.01$, i.e., when the proportion of signaling messages among the total generated traffic α exceeds the threshold 2%.

The question that arises now is whether a threshold of 2% is reasonable in real deployments? To answer this question, we can refer to the measurements of user activity on Roofnet [30], where data is collected by monitoring packets in a 24-h period. Accordingly, the signaling overhead represents almost 6% of the total generated traffic. In this range of $\alpha \approx 6\% \gg 2\%$, using our clustering schemes achieves significant gain regarding the radio resource utilization. This demonstrates indeed that clustering can be a viable solution for wireless mesh networking.

VII. Conclusion

In this paper, we investigated the radio resource utilization efficiency in wireless mesh networks. We proposed two clustering schemes to improve the resource utilization in such networks. Based on both analytical models and simulations, we proved that clustering is not helpful for slow mobile users with high data traffic loads. In the remaining cases, we demonstrated that our proposed clustering schemes achieve significant gains. Specifically, we showed that the distributed clustering algorithm DCAM stands out as the best solution for both small-to-moderate and highly connected wireless mesh networks, whereas the optimal static clustering (OSC) scheme is the best solution for both large and low-connected networks. Finally, we showed that the DCAM clustering technique handles better the changes in the physical topology due to AP failures or addition. Typically, the OSC clustering scheme needs several hours up to days to reconfigure the network according to the new optimal disjoint-cluster placement as opposed to the instantaneous DCAM reconfiguration.

REFERENCES

[1] MIT Roofnet [Online]. Available: http://www.pdos.lcs.mit.edu/roofnet/

[2] Seattle Wireless [Online]. Available: http://www.seattlewireless.net/

[3] R. Karrer, A. Sabharwal, and E. Knighly, "Enabling large-scale wireless broadband: The case for TAPs," *ACM Comput. Commun. Rev.*, vol. 34, no. 1, pp. 27–32, Jan. 2004.

[4] S. Basagni, "Distributed clustering for ad-hoc networks," in *Proc. ACM ISPAN*, Jun. 1999, pp. 310–315.

[5] A. B. McDonald and T. F. Znati, "A mobility-based framework for adaptive clustering in wireless ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 8, pp. 1466–1487, Aug. 1999.

[6] S. Basagni *et al.*, "Localized protocols for ad hoc clustering and backbone formation: A performance comparison," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 4, pp. 292–306, Apr. 2006.

[7] C. R. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 15, no. 7, pp. 1265–1275, Sep. 1997.

[8] N. H. Vaidya, P. Krishna, M. Chatterjee, and D. K. Pradhan, "A cluster-based approach for routing in dynamic networks," *ACM Comput. Commun. Rev.*, vol. 27, no. 2, pp. 49–64, Apr. 1997.

[9] A. Antis, R. Prakash, T. Vuong, and D. Huynh, "Max-min d-cluster formation in wireless ad-hoc networks," in *Proc. IEEE INFOCOM*, 2000, vol. 1, pp. 32–41.

[10] S. Banerjee and S. Khuller, "A clustering scheme for hierarchical control in multi-hop wireless networks," in *Proc. IEEE INFOCOM*, 2001, vol. 2, pp. 1028–1037.

[11] V. Bharghavan and B. Das, "Routing in ad-hoc networks using minimum connected dominating sets," in *Proc. IEEE ICC*, Jun. 1997, vol. 1, pp. 376–380.

[12] J. Wu and H. Li, "On calculating connected dominating set for efficient routing in ad-hoc wireless networks," in *Proc. Workshop Discrete Algor. Methods Mobile Comput. Commun.*, 1999, pp. 7–14.

[13] Y. Chen and A. Liestman, "Approximating minimum size weakly-connected dominating sets for clustering mobile ad-hoc networks," in *Proc. ACM MobiHoc*, Jun. 2002, pp. 165–172.

[14] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[15] O. Younis and S. Fahmy, "Distributed clustering in ad-hoc sensor networks: A hybrid, energy-efficient approach," in *Proc. IEEE INFOCOM*, 2004, pp. 629–640.

[16] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in *IEEE INFOCOM*, 2003, vol. 3, pp. 1713–1723.

[17] J. Wong, R. Jafari, and M. Potkonjak, "Gateway placement for latency and energy efficient data aggregation," in *Proc. IEEE LCN*, 2004, pp. 490–497.

[18] R. Chandra, L. Qiu, K. Jain, and M. Mahdian, "Optimizing the placement of Internet TAPs in wireless neighborhood networks," in *Proc. IEEE ICNP*, 2004, pp. 271–282.

[19] Y. Bejerano, "Efficient integration of multi-hop wireless and wired networks with QoS constraints," *IEEE/ACM Trans. Netw.*, vol. 12, no. 6, pp. 1064–1078, Dec. 2004.

[20] B. Aoun, R. Boutaba, Y. Iraqi, and G. Kenward, "Gateway placement optimization in wireless mesh networks with QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 11, pp. 2127–2136, Nov. 2006.

[21] C. E. Perkins and P. Bhagwat, "Highly dynamic destination sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. ACM SIGCOMM*, 1994, pp. 234–244.

[22] C. E. Perkins and E. B. Royer, "Ad-hoc on-demand distance vector routing," in *Proc. IEEE Workshop Mobile Comput. Syst. Appl.*, 1999, pp. 90–100.

[23] E. Cinlar, *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[24] Cooperative Association for Internet Data Analysis [Online]. Available: http://www.caida.org/

[25] ns2 home page [Online]. Available: http://www.isi.edu/nsnam/ns

[26] P. Kyasanur and N. H. Vaidya, "Routing and link-layer protocols for multi-channel multi-interface ad hoc wireless networks," *ACM Sigmobile Mobile Comput. Commun. Rev.*, vol. 10, no. 1, pp. 31–43, Jan. 2006.

[27] *Part 16: Air Interface for Fixed Broadband Wireless Access System*, IEEE Std. 802.16-2004, Jul. 2004.

[28] *Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE Std. 802.16e-2005, Feb. 2006.

[29] J. F. Labourdette *et al.*, "Fast approximate dimensioning and performance analysis of mesh optical networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 4, pp. 906–917, Aug. 2005.

[30] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *Proc. ACM MobiCom*, 2005, pp. 31–42.

Nizar Bouabdallah received the B.S. degree in telecommunications engineering from Ecole Superieur des Communications, Tunis, Tunisia, in 2001, and the M.S. and Ph.D. degrees in network and computer science from the University of Pierre and Marie Curie-Paris 6, Paris, France, in 2002 and 2004, respectively.

He joined Alcatel Research Laboratories, Marcoussis, France, in 2002 while pursuing the Ph.D. degree. In 2005, he was with North Carolina State University, Raleigh, as a Post-Doctoral Fellow. He is currently a Researcher with INRIA, Rennes, France. In 2007, he spent six months as a Visiting Researcher with the University of Waterloo, Waterloo, ON, Canada. His research interests include optical networking, wireless and sensor networks, and performance evaluation.

Rami Langar received the B.S. degree in telecommunications engineering from the Ecole Superieure des Communications, Tunis, Tunisia, in 2001; the M.S. degree in network and computer science from the University of Pierre and Marie Curie-Paris 6, Paris, France, in 2002; and the Ph.D. degree in network and computer science from the Ecole Nationale Superieure des Telecommunications de Paris, Paris, France, in 2006.

In 2007 and 2008, he was with the School of Computer Science, University of Waterloo, Waterloo, ON, Canada, as a Post-Doctoral Research Fellow. He is currently an Associate Professor of computer science with the University of Pierre and Marie Curie-Paris 6. His research interests include post-Internet protocol architecture, mobility and resource management in wireless mesh and vehicular ad hoc networks, performance evaluation, and quality-of-service support.

Raouf Boutaba received the M.Sc. and Ph.D. degrees in computer science from the University of Pierre and Marie Curie-Paris 6, Paris, France, in 1990 and 1994, respectively.

He is currently a Professor of computer science with the University of Waterloo, Waterloo, ON, Canada. His research interests include network, resource, and service management in wired and wireless networks.

Dr. Boutaba has received several best paper awards and other recognitions such as the Premier's Research Excellence Award. He is the founder and Editor-in-Chief of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and serves on the editorial boards of several other journals. He is currently a distinguished lecturer of the IEEE Communications Society, the Chairman of the IEEE Technical Committee on Information Infrastructure, and the Director of the ComSoc Conference Publications Board.