Ann. Telecommun. (2008) 63:137–148
DOI 10.1007/s12243-008-0017-6

ORIGINAL PAPER

Sensor-based architecture for QoS provisioning and fast handoff management in WLANs

Sonia Waharte · Jin Xiao · Raouf Boutaba

Received: 19 August 2007 / Accepted: 26 December 2007 / Published online: 21 February 2008

Abstract As wireless local area networks gain popularity from network access providers and customers, supporting multimedia applications becomes a crucial yet unresolved challenge. The need to maintain quality-of-service in the presence of bandwidth limitations, increasing traffic volume and user mobility entails radical rethinking in resource management design in WLANs. The unique capabilities of wireless sensor networks constitute a promising research direction to tackle these issues. In this paper, we present a new sensor-based resource management architecture for enhanced QoS provisioning and handoff management in WLANs. Through theoretical analysis and simulations, we show that the framework can maximize bandwidth utilization while satisfying applications' QoS requirements and significantly reduce handoff latency.

Keywords Wireless mesh networks · Quality-of-service

S. Waharte (✉) · J. Xiao · R. Boutaba
David R. Cheriton School of Computer Science,
University of Waterloo, 200, University Avenue West,
Waterloo, Ontario, Canada
e-mail: spwaharte@uwaterloo.ca

J. Xiao
e-mail: j2xiao@uwaterloo.ca

R. Boutaba
e-mail: rboutaba@uwaterloo.ca

1 Introduction

In recent years, the prevalence of networked electronic devices and the ensuing need for ubiquitous Internet accessibility have led to the growth of wireless access networks. Wireless LAN (WLAN), as the premiere access technology, fulfills this need for "anywhere, anytime" connectivity. Flexible and easy to deploy, WLANs are installed in abundance across public areas (e.g., hotels, cafes, shopping malls, etc.) and at home [22].

Although intended for best-effort services, the increasing presence of multimedia applications on the Internet has accentuated the need for end-to-end quality of service (QoS) support in WLAN. Unlike wired access networks, the capacity constraint of unlicensed frequency spectrum, the presence of environmental interference, and access competition of the shared transmission medium render the traditional bandwidth over-provisioning approach impractical. Even though advancements in hardware technology have progressively increased the nominal throughput in WLAN to as high as 54 Mbps (IEEE 802.11a, http://grouper.ieee.org/groups/802/11/), it is quickly outpaced by the rapid growth of mobile users. Efficient QoS provisioning is thus a key factor in supporting multimedia applications over wireless access networks. To be successful, such a provisioning mechanism must not only meet the QoS requirements of the individual applications, but also maximize the overall bandwidth utilization of the wireless channel. User roaming is yet another important consideration. As a user moves from the range of one access point to another, handoff must take place to reassociate the user. Handoff latency has a significant impact on the delay and jitter experienced at the application layer.

IEEE 802.11 Standard has been designed for best-effort traffic and lacks QoS provisioning mechanisms. Proposals have flourished to provide QoS at the MAC layer through service differentiation (e.g., IEEE 802.11e, http://grouper.ieee.org/groups/802/11/Reports/tge_update.htm), admission control, and bandwidth reservation, and at the physical layer with link adaptation [25]. However, to accommodate QoS-assured applications, hard QoS guarantees are necessary. Some attempts have been made to provide such guarantees through polling mechanisms [i.e., point coordination function (PCF) defined in IEEE 802.11 Standard [1]], but exhibited very poor bandwidth utilization [13].

Knowing the essence of QoS support for multimedia applications lies in the ability to accommodate the minute needs of each user, both in terms of its bandwidth requirements and its handoff requests, a much higher degree of omniscience is required in WLAN today. Such knowledge can only be obtained from a well-established monitoring and control infrastructure. To maximize bandwidth utilization, it is also necessary for such an infrastructure to be independent, alleviating the increased management message exchange from the data channel, leading to a clear control and data plane separation. Wireless sensor networks appear as a promising research direction to address these issues. Sensor nodes are tiny, low-cost devices, embedding limited communications, processing, and sensing capabilities [3]. The effectiveness of sensor networks for monitoring purposes has been demonstrated in many academic and industrial research projects [2, 14]. Sensor networks have been deployed for habitat monitoring [20], natural life observations [10], traffic and highway applications [8], and target detection [17]. The location tracking capabilities of sensor nodes can also aid in network resource management.

In this paper, we propose a novel sensor overlay architecture for WLAN resource management. We demonstrate that such sensor overlay design can drastically improve QoS provisioning and handoff management in WLAN through proactive network monitoring and efficient control message transport. The nonobtrusive nature of our design not only results in optimal utilization of the wireless channel but also offers full compatibility with existing WLAN protocols.

The remainder of this paper is organized as follows: Section 2 presents background and related work. The design of our architecture is described in Section 3. Section 4 gives details of our implementation. A theoretical analysis of our approach is conducted in Section 5 and simulation results are presented in Section 6. Section 7 concludes the paper.

2 Background and related work

An effective resource management framework in WLAN aims at providing QoS support while maximizing bandwidth utilization. It remains a challenging task due to limited bandwidth capacity in an intrinsically shared medium and the need to support user mobility. In this paper, we focus on two fundamental factors in WLAN resource management: QoS provisioning and handoff management.

2.1 QoS provisioning

Two general approaches have been proposed for QoS provisioning in WLAN. Soft QoS guarantees can be obtained by integrating traffic differentiation mechanisms based on priority levels, whereas hard QoS guarantee can be ascertained through resource reservation. IEEE 802.11 Standard [1] follows these approaches by defining two mechanisms, the best-effort distributed coordination function (DCF) and the QoS-assured PCF. The latter mechanism has never been implemented due to severe bandwidth wastage [13]. Current resource management implementations in WLANs have focused on offering differentiated QoS, based on the DiffServ approach initially developed for the Internet [9]. Priority levels can usually be enforced by setting different waiting periods before accessing the medium (i.e., IEEE 802.11e, http://grouper.ieee.org/groups/802/11/Reports/tge_update.htm). Our architecture offers hard QoS guarantees with class differentiation while still achieving maximum bandwidth utilization.

2.2 Layer 2 handoff

Many studies have focused on maintaining IP layer connectivity by improving mobile IP efficiency [24]. However, reducing the handoff latency at the network layer can be effective only if the MAC layer (L2) handoff latency does not exceed acceptable thresholds. Studies [16] have demonstrated that, with current implementations, the scanning process of handoff can be as high as 400 ms [15]. ITU-T recommends that one-way delay should not exceed 400 ms for general network planning, and highly interactive applications (e.g., voice and multimedia applications) require even more stringent end-to-end latency, with an upper bound of 150 ms [6]. Improvements have been proposed by reducing the detection phase in considering that handoff should be initiated if the transmission of a frame and two successive attempts to retransmit this frame failed or if three successive beacon messages have not been received [23]. The scanning process can be enhanced by reducing the waiting period [11, 12, 15, 23] or by applying selective scanning [11, 19]. Supplying the mobile nodes with up-to-date information on the network condition, our architecture further reduces scanning time in *selective-scanning* mode, or completely eliminates the scanning process in *scan-free* mode.

Sensor networks have been shown as an effective monitoring and data gathering infrastructure [3, 8, 20], whose wide deployment is expected. As such, utilizing sensor networks for WLAN resource management constitutes a promising research direction. For example, during the layer-3 handoff, the registration request can be initiated by the sensor network upon detection of a mobile node's movement instead of being initiated by the mobile node itself [4]. The benefit of out-of-band signalling has been evidenced in many network architectures such as Public Switched Telephone Network, Integrated Service Digital Network, or satellite systems (e.g., INMARSAT-A). In WLAN, the use of a separate control plane is of significant interest as the shared transmission medium and the limited nominal capacity put forth to resource optimization strategies leading to limiting the number of control messages exchanges. This can be achieved by embedding mobile devices and access points with dual interfaces or, as presented in this paper, by using sensor networks as a control infrastructure. Dual interface transmitters present the advantage of not requiring additional infrastructure but at the cost of extra energy consumption due to longer transmission distance and higher probability of collision. At the opposite, a sensor-based control network, although relying on a the prior deployment of sensor nodes in the coverage area of the WLAN, offer enhanced services such as providing accurate information on the localization of the access points or on their signal strength with low-energy cost (short transmission distance, reduced contention). Moreover, the proposed framework does not necessitate a dedicated infrastructure as sensor networks are multipurpose and can be used in many applications (environment monitoring, target detection, etc.).

In this paper, we demonstrate how sensor networks can significantly enhance wireless network infrastructure with efficient QoS provisioning mechanisms and fast layer-2 handoff.

3 Architecture overview

IEEE 802.11 Standard medium access protocol is based on random deferred access with backoff mechanisms. This design constraint makes deriving an upper bound on the transmission delay very difficult. To overcome this limitation, a fixed transmission schedule can be established. However, not all of the users have the same level of guarantee, and the increased overhead of exchanging this information can significantly interfere with data transmissions. Handoff latency further affects the delay upper bound.

An efficient resource management framework should therefore have the following features: separate control transmission from data traffic, on-demand resource allocations with hard and soft QoS guarantees, and fast handoff.

Sensor networks are nonobtrusive, multipurpose, data gathering infrastructures capable of various environment sensing functions (building temperature, etc.), as well as tracking mobile node movements and access point locations. Its expected ubiquitous deployment can be utilized for WLAN resource management and to facilitate control information exchange.

Our architecture complements the existing IEEE 802.11 access protocol with a low-capacity sensor network acting as a control plane, operating on a different frequency band. It can provide users with accurate information on the environment and deliver control messages between mobile terminals and access points.

Three roles are defined in the sensor control plane:

- The manager node: This particular node is embedded in the access point and operates as a data aggregation server, responsible for gathering the resource requests (RREQ) of the mobile terminals. It is also in charge of establishing an appropriate schedule based on the received requests and sending back the corresponding resource notifications (RNOT) to the mobile terminals. Initializing relay sensors with information about the access point (supported data rate, channel used, etc.) is also part of the manager node's duty. As a data aggregation server, the manager node gathers authentication and reassociation requests from mobile nodes whenever a handoff process is initiated.
- The relay sensors: These fixed nodes embedding sensing facilities are uniformly dispersed in the vicinity of the access points. Contrarily to traditional sensor networks, the relay sensors are not likely to suffer from power outage or to fail due to hostile environments. Statically placed, they can draw power

from an electrical outlet. Their primary function is to route control packets (such as RREQ/RNOT or authentication/reassociation packets) between the agents and the manager node. Thanks to their sensing capabilities, they can also store information about the surrounding access points and provide this information to the agents upon request.

– The sensor agents: These sensors are attached to the mobile nodes. They are responsible for initiating RREQ according to the mobile nodes' QoS requirements and for processing RNOT upon reception.

4 Implementation details

4.1 QoS provisioning

Our design offers mobile users hard QoS guarantees. The general concept underlying our approach is as follows: A mobile node can request resource reservation from the access point by sending RREQ messages, which contain information on the capabilities of the mobile node in terms of throughput and on the desired bandwidth. During a predetermined period, the duration of which is to be dimensioned, the access point waits for the reception of such messages. Once the registration period is over, the access point establishes a schedule based on the received bandwidth requests and transmits this information back to the concerned nodes in a RNOT message. The advantage of this mechanism is the dynamic set up of a resource scheduling based on the user's need instead of a fixed resource allocation that introduces significant bandwidth wastage. Furthermore, each mobile node has a bounded transmission delay. Each mobile node can determine its bandwidth needs by monitoring its buffer occupancy at the application level.

Our scheme is a cyclic process consisting of two phases in the data plane: a new QoS-assured scheduling coordination function (SCF), which we introduce and the QoS-differentiated DCF, as depicted in Fig. 1. Instead of implementing a fair scheduling algorithm, which incurs the drawback of ignoring the specific needs of individual nodes, we propose an application-adaptive scheduling algorithm.

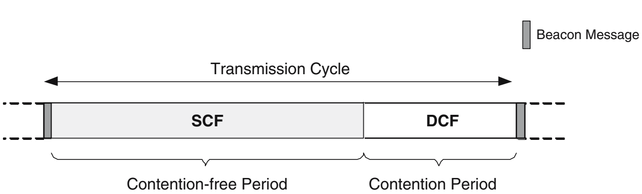

Fig. 1 Integration of QoS support in current architectures

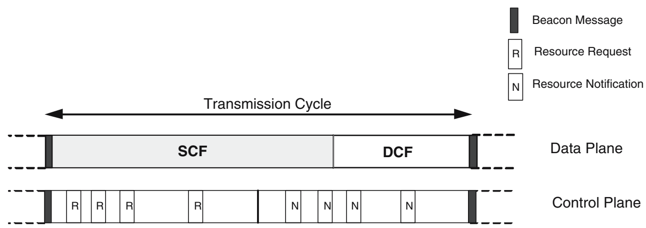

Fig. 2 Resource reservation process

In the control plane, the transmission cycle is divided into two periods. The first half of the transmission cycle is reserved for the registration of mobile nodes. Each mobile node wanting to transmit in the next transmission cycle (or for a set of successive transmission cycles) sends a RREQ to the access point (using the relay sensors as repeaters). The confirmation of time slots allocation occurs during the second period of the transmission cycle (Fig. 2). In the data plane, the mobile nodes with allocated time slots during prior transmission cycles conduct their data transmission according to the established schedule. This period, referred to as SCF, has a duration that varies according to the bandwidth requirement of the mobile nodes, but which cannot exceed a maximum αT_c, where T_c is the duration of the transmission cycle and α is a reservation factor, whose value can be used for service differentiation. A DCF period is included to provide backward compatibility with existing implementations and to support traffic that does not require hard QoS guarantees (e-mail, FTP, etc.). The scheduling algorithm is as follows: Let b be the buffer occupancy (in kb), T the transmission rate (in kb/s), and ϵ the transmission overhead (in s). The

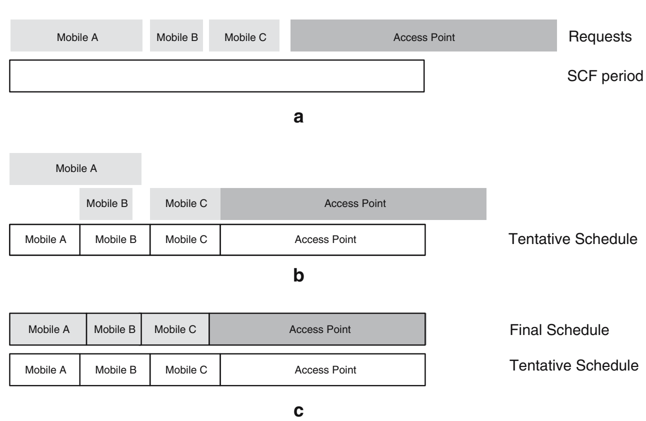

Fig. 3 Scheduling scheme. **a** Requests from mobile node access point. **b** Tentative schedule: even share of SCF period. **c** Final transmission schedule after excess bandwidth redistribution

for i=1..r **do**
  if $f(b_i)_{init} > \frac{\omega_i}{\sum_{i=1}^{r}\omega_i}\alpha Tc$ **then**
    $f(b_i) = \frac{\omega_i}{\sum_{i=1}^{r}\omega_i}\alpha Tc$
    count++
  else
    $f(b_i) = f(b_i)_{init}$
    $C = C + \frac{\omega_i}{\sum_{i=1}^{r}\omega_i}\alpha Tc - f(b_i)_{init}$
  end if
end for

while true **do**
  if count = 0 **then**
    exit
  end if
  r'=count
  count=0
  $C_{remain} = 0$
  for i=1..r' **do**
    if $(f(b_i)_{init} - f(b_i)) < \frac{\omega_i}{\sum_{i=1}^{r'}\omega_i}C$ **then**
      $f(b_i) = f(b_i) + \frac{\omega_i}{\sum_{i=1}^{r'}\omega_i}C$
      count++
    else
      $f(b_i) = f(b_i)_{init}$
      $C_{remain} = C_{remain} + \frac{\omega_i}{\sum_{i=1}^{r'}\omega_i}C - f(b_i)$
    end if
  end for
  $C = C_{remain}$
end while

Fig. 4 Scheduling algorithm

amount of channel time required to satisfy a request is then: $f(b) = b/T + \epsilon$.

Let r be the number of requests received and C the excess capacity. Class differentiation for a node i is enforced by a weighted resource allocation based on a factor ω_i. r' represents the nodes whose resource allocation did not satisfy their request. Our scheduling algorithm is illustrated in Fig. 3, and its implementation is given in Fig. 4. In this example, the access point is given the highest weight to account for the downlink traffic. Mobile A, mobile B, and mobile C have the same weight. The first step consists in assigning a weighted bandwidth allocation. In the second step, the excess capacity is redistributed among the nodes whose current allocation did not meet their initial request.

4.2 Fast handoff

Our proposed architecture also aims at reducing handoff latency. The handoff process can be decomposed into three phases: scanning, authentication, and reassociation. Details of the operation modes for each phase are depicted in Fig. 5a.

When the signal quality degrades below a certain threshold, the mobile node initiates a scanning process to detect the presence of other access points. In IEEE 802.11, two mechanisms are proposed to perform this task. The mobile node can successively scan all the channels and listen on each channel to detect the presence of access points through the reception of beacon messages. The other method consists in proactively sending probe requests to the access points and in waiting for the reception of probe responses. Passive scanning has the advantage of not generating extra packet overhead compared to the active scanning method, but the additional delay introduced can be detrimental to delay-sensitive applications.

Sensor nodes can provide a better alternative by combining these two methods. With passive scanning, they can keep track of the surrounding access points and maintain information on their current status through a periodic scanning of the channel on which they have been registered. The sensor nodes can then provide the mobile nodes with accurate information on the number of surrounding access points and on parameters such as signal quality or number of users associated with each access point. The mobile node is therefore able to directly associate with the best access point (scan-free mode) or proceed with a selective scanning on the listed access points (selective-scan mode), to account for the discrepancies that can exist because of the difference of position between the mobile node and the sensor node. Figure 5b depicts the enhanced handoff process using our sensor-based architecture.

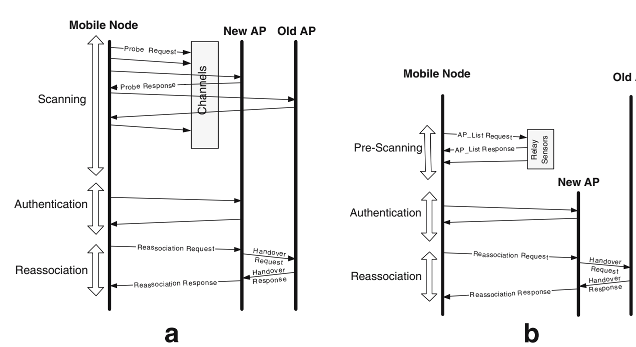

Fig. 5 **a** Handoff process. **b** Fast handoff using sensor networks

5 Computation of registration delay

To guarantee an on-time registration of the mobile nodes requesting time slot reservations (i.e., to avoid that the RREQ messages reach the access point after the end of the registration period), it becomes critical to properly estimate the impact of an increasing number of nodes on the overall registration delay.

The following analysis aims at providing an approximation of the registration delay. The accuracy of the results is then assessed through simulations. For the analysis, we make the following assumptions:

- Mobiles nodes are uniformly distributed at random over the coverage area of the access point (represented by a circle centered at the access point).
- The relay sensors are marshalled according to a hexagonal lattice (honey-grid) topology. Whereas our analysis is not restricted to any specific topology, our validation has been conducted with such a node distribution to maximize the network coverage while minimizing the number of nodes.
- All the devices in the control plane have similar transmission capabilities. This assumption makes the analysis more tractable by alleviating the computation from the specifics inherent to each mobile node. However, it does not impact the efficiency of the proposed architecture.
- The medium access is regulated according to a CSMA/CA protocol without the request to send/ clear to send (RTS/CTS) mechanism. Actually, with the data exchange being solely composed of control packets, the overhead pertaining to the RTS/CTS mechanism is not justified.
- Time is slotted and synchronization is maintained between the relay sensors. This can be achieved by the access point through the periodic transmission of beacon messages.

Our computation of the registration delay differs from previous approaches [7] in that, for a multitude of source nodes (mobile nodes), there exists only one destination: the access point. All the traffic generated by mobile nodes converges towards the access point, the nodes surrounding the access point constituting the bottleneck of the data transmission process. To compute the end-to-end transmission delay from the mobile node to the access point, we thus need to estimate: (1) the average number of hops and (2) the average transmission delay for one single hop.

5.1 Average number of hops

In the control plane, each packet has the access point as destination. Therefore, the computation of the number of hops can be simply derived by determining the total number of relay sensors present in the considered topology, and for each of these relays, by calculating the number of hops necessary to reach the access point. By adding one last hop to take into account the communication between the mobile node and the nearest relay sensor, the average number of hops can be derived. This obviously holds only for a uniform distribution of the mobile nodes.

Let k be the diameter of the network in terms of hops and h the average number of hops. h can be computed as follows:

$$h = \frac{\sum_{i=1}^{k+1} 6i(i-1)+1}{\sum_{i=1}^{k+1} 6(i-1)+1}$$

$$h = \frac{1+2(k+1)((k+1)^2-1)}{1+3k(k+1)}$$

We can then obtain:

$$h = \frac{1+2k(k+1)(k+2)}{1+3k(k+1)}$$

5.2 One-hop transmission delay

The second step of the analysis consists in deriving the average transmission delay between two neighboring relay sensors while considering the interference created by surrounding nodes. The impact of the number of mobile nodes on the transmission delay being negligible, it will not be considered in the derivation of the delay approximation. The rationale for this comes from the following observations: First, during the registration process, each mobile node transmits only one single packet. Therefore, the additional delay involved by a potentially increasing number of collisions can be alleviated by a fixed contention window size. Second, if a relay sensor has some traffic in its queue to be forwarded to the access point, it means that a mobile node successfully registered and thus will not contribute to the contention process anymore. The impact of a mobile node on the overall registration delay can therefore be neglected given that the relay sensors work at saturation (i.e., always have data to send). The computation follows a similar approach as in [7]. We denote R as the transmission radius of the access point, r the transmission radius of the relay sensor, N the number of mobile nodes, and S the number of relay sensors.

As previously described, two types of messages are exchanged during the registration process: RREQ and request acknowledgment (ACK). Each type of packet is associated with a transmission duration referred to as T_{RREQ} and T_{ACK}, respectively. An additional waiting period short interframe space (SIFS) is considered between the transmission of a RREQ and an ACK. The time unit, time slot, is referred to as T_{TS}. Each sensor relay can be in one of the four states represented in Fig. 6:

– Idle: No transmission occurs. The relay sensor decrements its backoff timer. The duration of this state is referred to as T_i.
– Transmit: Successful transmission. The duration of this state, T_t, corresponds to $T_{RREQ} + T_{SIFS} + T_{ACK}$.
– RREQ-Coll: Collision on RREQ. The duration of this state, T_r, corresponds to the transmission duration of a RREQ T_{RREQ}.
– ACK-Coll: Collision on the acknowledgment of a successfully received RREQ. The duration of this state T_a can be expanded as $T_{RREQ} + T_{SIFS} + T_{ACK}$.

Let CW be the size of the contention window and p the probability that a saturated node transmits at a given time slot. It has been proven that, for a fixed contention window size [5], p can be derived as:

$$p = \frac{2}{CW + 1}$$

Let N' be the number of nodes in the coverage area of a relay sensor. N' can be expressed as $N' = \rho \Pi r^2$ with a network density $\rho = \frac{S}{\Pi R^2}$. By considering the equilibrium state, each transition can be straightforwardly computed as follows:

$$P_{ii} = (1 - p)^{N'}$$

$$P_{it} = N' \Pi_s (1 - p)^{N'-1}$$

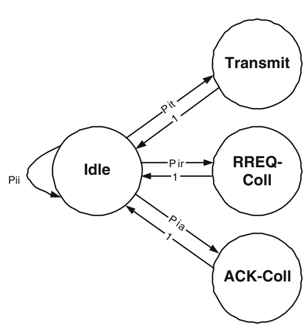

Fig. 6 FSM of data transmission during the registration process

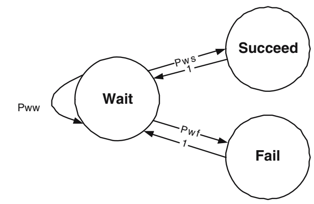

Fig. 7 Markov model of wireless transmission

with Π_s the probability that a node successfully sends its RREQ and receives its acknowledgment.

$$P_{id} = 1 - (1 - p)^{N'} - N' p (1 - p)^{N'-1}$$

$$P_{ia} = 1 - P_{ii} - P_{it} - P_{id}$$

The last step consists in computing Π_s. To this effect, a transmission process can be modelled as represented in Fig. 7. Basically, a node can be in three different states: wait (for the backoff timer to expire), if no transmission occurs; success, if the node accesses the medium and successively transmits its data; or fail, if after accessing the medium, the node experiences collision, which would necessitate data retransmission. The probability to remain in the wait state corresponds to the situation where no node initiates a data transmission process and can be computed as:

$$P_{ww} = (1 - p)^{N'}$$

For a transmission to be successful, a node should not experience any collision when sending its RREQ message, and there should not be any collision on the acknowledgment message sent back to the source node. Therefore, the potential presence of hidden terminals has to be taken into consideration. The nodes in the transmission range of the source node and the receiving node should not initiate a data transmission process at the same time as the source node and the receiving node (Fig. 8). For a given transmission, the area $B(r')$ containing nodes (hidden terminals) that may potentially interfere with the transmitting node can be computed as (r' is the internode distance) [21]:

$$B(r') = \Pi r^2 - 2r^2 \arccos\left(\frac{r'}{2r}\right) - \frac{r'}{2r}\sqrt{1 - \left(\frac{r'}{2r}\right)^2}$$

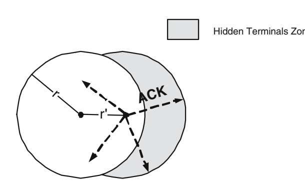

Fig. 8 Hidden terminals

Table 1 Control plane and data plane models for simulations

	Control plane	Data plane
Physical layer	Based on PHY802.11b	Based on PHY802.11a
	2.4 GHz comm.	5 GHz comm.
	Limited transmission range[a]	Increased transmission range[b]
	Data rate: 2 Mbps	Data rate: 54 Mbps
MAC layer	Based on MAC802.11	Based on MAC802.11
	CSMA/CA without RTS/CTS mechanism	CSMA/CA without RTS/CTS mechanism

[a]TX-Power is reduced to the minimum required for successful communications so that interferences are reduced
[b]TX-Power is increased so that 54 Mbps can be achieved wherever a mobile node is placed (no rate fallback)

As the distance between two nodes may vary between ϵ (we assume that two nodes are not exactly at the same location) and r, the average internode distance r_a can be computed as follows:

$$r_a = \sqrt{\frac{r^2 - \epsilon^2}{2}}$$

$$r_a \approx \frac{\sqrt{2}r}{2}$$

Consequently, the probability of a successful transmission can be derived as follows:

$$\Pi_s = \frac{P_{ws}}{2 - P_{ww}}$$

$$\text{with} \begin{cases} P_{ww} = (1-p)^{N'} \\ P_{ws} = p(1-p)^{N'-1}[(1-p)^{\rho B(r')}]^{\frac{(T_{RREQ}+T_{SIFS}+T_{TS})}{T_{TS}}} \end{cases}$$

5.3 Average end-to-end transmission delay

When computing the overall average end-to-end transmission delay, an important factor to consider is the possibility of simultaneous transmissions. As the transmission radius of the relay sensors is less than the transmission radius of the access point, it is feasible to envision that several nodes transmit data successfully. Therefore, by averaging this possibility over the coverage area of the access point, we can obtain a good approximation of the impact of this phenomenon. By estimating the average surface involved in a data transmission process, we can easily derive the possible number of simultaneous transmissions. Similar to before, we need to consider the area in which hidden terminals are located. The maximum area B_{MAX} is reached when $r' = r$. Therefore, we can derive B_{MAX} as:

$$B_{MAX} = \Pi r^2 - 2r^2 \left(\frac{\Pi}{3} - \frac{\sqrt{3}}{4} \right)$$

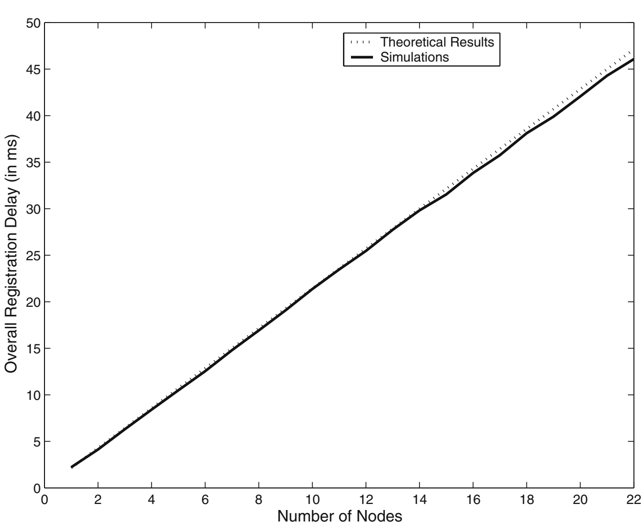

Fig. 9 Registration delay (theoretical results validated through simulations)

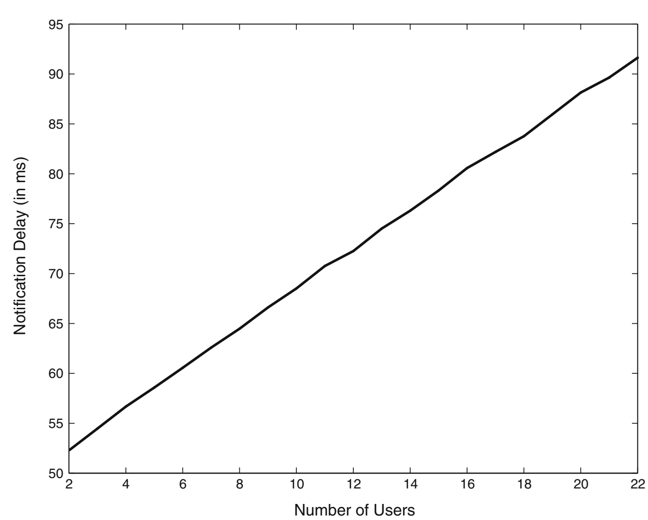

Fig. 10 Notification delay

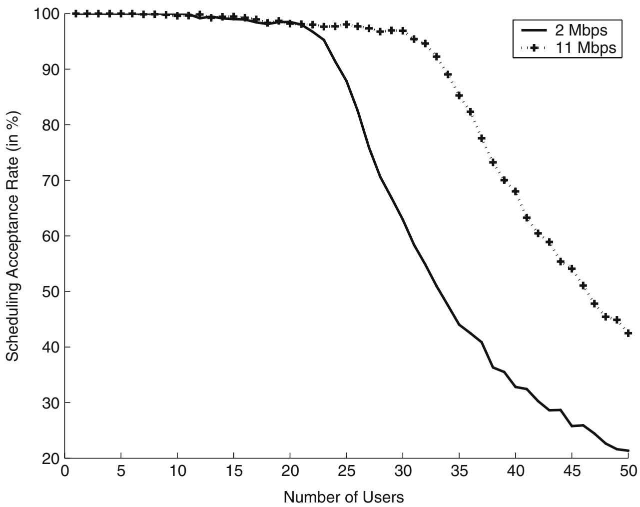

Fig. 11 Scheduling acceptance rate

The average number of simultaneous transmissions n_{sim} can then be computed as follows:

$$n_{sim} = \frac{\Pi R^2}{\Pi r^2 + \frac{B_{MAX}}{2}}$$

$$n_{sim} = \frac{2\Pi R^2}{3\Pi r^2 - 2r^2\left(\frac{\Pi}{3} - \frac{\sqrt{3}}{4}\right)}$$

Finally, by considering the time between two successive transmissions and by averaging this result over the number of mobile nodes, the number of successive hops, and the number of simultaneous transmissions, we obtain the target estimation of the registration delay. The percentage of time in the transmit state is derived as the time spent in the transmit state over the time spent in all the possible states

$$perc = \frac{P_{it}T_t}{P_{it}T_t + P_{ir}T_r + P_{ia}T_a + T_i}$$

So in total, the average end-to-end transmission delay can be obtained as:

$$delay = \frac{1 - perc}{perc}\frac{T_t * N * h}{n_{sim}}$$

The mathematical analysis has been validated through simulations under various scenarios. The results are presented in the following section.

6 Evaluation

In this section, we assess the benefits of our architecture for QoS provisioning and handoff management through simulations implemented on top of QualNet 3.6.1 [18].

The efficiency of our architecture relies on a proper dimensioning of parameters including the registration duration $T_{REGISTRATION}$, the transmission cycle duration T_c, and the control plane data rate $DataRate_{CP}$. The simulation results can be used for such dimensioning. The benefits of our architecture for QoS provisioning are assessed by comparing our SCF to a traditional DCF for different types of traffic. The second set of simulations demonstrates the effectiveness of a selective scanning over an active scanning whenever a handoff is initiated and the significant scanning delay reduction that can be achieved. We performed only a comparison with DCF, as the design of our scheme, by nature, outperforms PCF by providing a dynamic resource allocation based on users' need instead of a systematic polling of the registered users (who may not have data to send).

6.1 QoS provisioning

Table 1 summarizes the parameters of our simulations. We use two frequency bands for our simulations, but a subset of the band supporting the data transmission

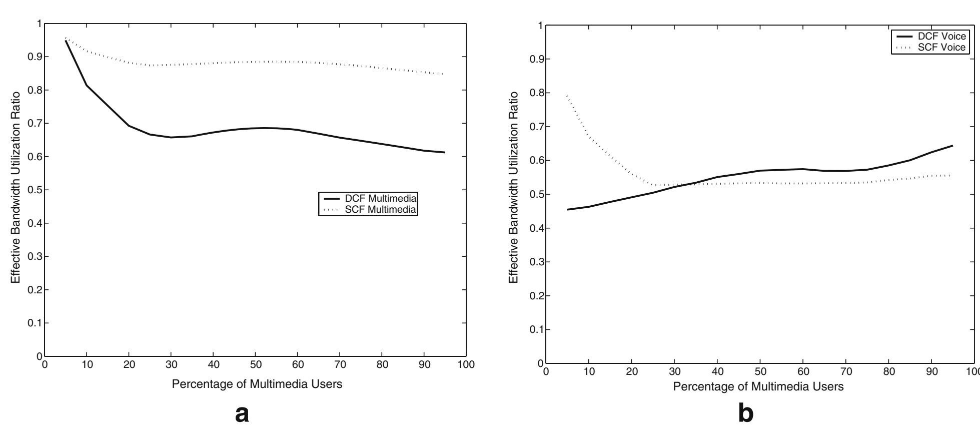

Fig. 12 Effective bandwidth utilization ratio of **a** multimedia users **b** voice users

could be used for the control plane. All mobile nodes are considered as belonging to the same class of service. Synchronization is maintained by the access point through the periodic transmission of beacon messages. The simulation results are averaged over 100 runs with different seed values and mobile node placements.

6.1.1 Registration phase

We consider *n* randomly scattered mobile nodes in the vicinity of an access point. All of them send RREQ packets after they receive a beacon message from the access point. RREQ are then routed to the access point through relay sensors. Figure 9 depicts the registration delays for both the mathematical analysis in Section 5 and the performed simulations.

From these results, we can observe that if $T_{REGISTRATION}$ is set to 50 ms, we ensure that our architecture provides QoS guarantees to a maximum of about 20 users, which should be sufficient for a 54-Mbps data rate.

6.1.2 Notification phase

Once established, the schedule is communicated to mobile nodes with RNOT messages and informs them of their respective allocated time slots. These packets have to successfully reach the destination sensor agents before the end of the current cycle. For $T_{REGISTRATION} = 50ms$, Fig. 10 shows the notification delays (delay of the last RNOT being received by an agent) for *n* users. We can conclude that a transmission cycle of 100 ms guarantees that all the packets are routed back to the agents with a leeway of around 10 ms.

6.1.3 Scalability of the architecture

To evaluate the scalability of our architecture, we need to estimate the number of users that can successfully register while considering the contention for the medium access in the control plane. If a large number of mobile nodes desire to register, the probability of collision will increase and therefore impact the registration delay. The consequence can be a reduced number of registered users. Figure 11 depicts the acceptance rate in the scheduling when different data rates (2 and 11 Mbps) are used for the control plane. Even if the architecture is more scalable (higher acceptance rates) with higher data rates, the improvement is not significant. Several reasons can explain this. First, the expected number of multimedia users is not expected to exceed 25. Beyond this limit, the QoS provided to each user is lessened by the small amount of bandwidth that can be guaranteed. The impact of the overhead transmission becomes more significant especially when considering that the Physical Layer Convergence Protocol overhead (24 bytes) is always transmitted at 1 Mbps. Second, the waiting periods defined in IEEE 802.11 Standard, such as SIFS, DIFS, and EIFS, are independent of the data rate. Third, higher data rates require higher transmission power, increasing the probability of collision. This explains the small difference in the acceptance rate between configurations with data rates of 2 and 11 Mbps. As a consequence, for energy considerations, we set $DataRate_{CP}$ at 2 Mbps.

Table 2 Standard deviation for 20 users (12 multimedia users and 8 voice users)

	Mult. 2 Mbps	Voice	Mult. 3 Mbps	Voice
DCF	0.8676	0.0011	0.9777	0.0006
SCF	0.0387	0.0067	0.13957	0.0056

Table 3 Simulation parameters for handoff

	Reference values (ms) (IEEE 802.11)	Optimized values (ms)
MinChannelTime	17	6.5
MaxChannelTime	38	11

6.1.4 Benefits for multimedia users

To assess the benefits of our architecture, we implemented different scenarios and analyzed for each case the average transmission delay and the number of bits transmitted. Our architecture, referred to as SCF in the figures, has been dimensioned with 80% of the transmission cycle for SCF and 20% for DCF.[1] Note that alternative dimensioning can be used depending on how much the network designers want to prioritize multimedia traffic over best-effort traffic. In Fig. 12a, we observe that the benefits of our architecture are significant whatever the number of multimedia users is. The impact on voice traffic (Fig. 12b) is negligible, as the traffic is sporadic and does not require as much bandwidth capacity as multimedia traffic (mainly video traffic). Note that voice and video traffic source characteristics are those provided in Qualnet (with 160-

[1] The bandwidth utilization comparison considers a full transmission cycle.

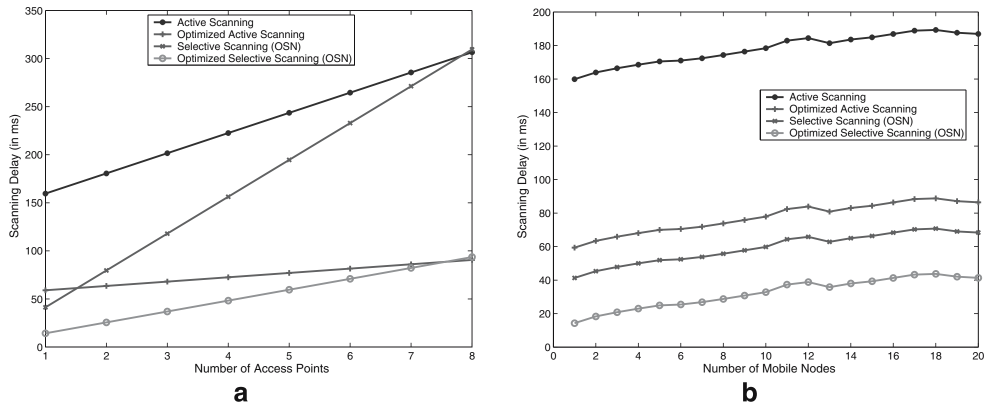

Fig. 13 **a** Scanning delays for one user with an increasing number of access points. **b** Scanning delays for one mobile user with two active access points

and 1,280-byte packets generated, respectively, every 20 and 1 ms with a probability of 0.352 and 0.25 to model the traffic fluctuations). Simulations conducted to evaluate the performance of our architecture on the transmission delay showed a significant enhancement over DCF especially when the number of multimedia users increases. Our architecture also exhibits a more fair share of the medium utilization as reflected by the standard deviation (Table 2).

6.2 Handoff management

In this section, we evaluate the delays pertaining to the scanning process when a handoff is initiated. Simulations are performed with the active scanning mechanism defined in the IEEE 802.11 standard and with the selective scanning described in Section 4. Both standard and optimized parameters for MinChannelTime (the minimum waiting period before considering that the channel is idle) and MaxChannelTime (the maximum waiting period after a probe response has been successfully received) are used as shown in Table 3.

In the simulations, we assume that $N_{CHANNELS} = 8$. To maximize the transmission distance, probe requests and probe responses are sent in the data plane using the lowest possible data rate (6 Mbps) defined in the IEEE 802.11a standard.

6.2.1 Scanning delays for multiple busy channels

In this scenario, a single user initiates a handoff. As opposed to active scanning where every channel needs to be scanned, selective scanning relies on a prescanning step performed in the control plane by the relay sensors, restricting the active scanning to the busy channels. Figure 13a shows the delays pertaining to these different approaches. We can observe that the delays are greatly reduced with a limited number of access points. The benefits of our architecture are lessened in the case where all channels are busy.

6.2.2 Scanning delays for a single busy channel under load

In this configuration, only one access point to which the user can handoff is active. An increasing number of users, randomly placed in the service area of the access point, generate CBR traffic at a 10-Mbps data rate. Given the shared nature of the medium, a defer access to the medium may occur before a user sends a probe request on the channel during the handoff process. According to the simulations results shown in Fig. 13b, this delay is still negligible compared to the time wasted waiting for probe responses on every channel (active scanning). Therefore, the performance of the proposed selective scanning still outperforms traditional handoff techniques.

7 Conclusion

QoS support remains a critical feature absent in WLANs today. Although research efforts have been pushing towards the development of solutions based on differentiated QoS, the lack of hard guarantee is highly detrimental to time-sensitive applications.

In this paper, we proposed a novel architecture based on sensor networks. We addressed two major issues: resource reservation for QoS provisioning and handoff management. QoS provisioning can be efficiently and dynamically established on-demand, based on users' minute needs. Mobile nodes can also benefit from a reduction of the handoff delay by obtaining accurate information on the surrounding environment through the relay sensors. We demonstrated the effectiveness of such a sensor-based approach through theoretical validation, as well as simulations.

Sensor networks have salient advantages for WLAN QoS management. With the ever-growing ubiquitousness of sensor deployment, sensor-enhanced resource management offers an effective and feasible solution. A real-world implementation of our architecture is planned for further validation.

References

1. IEEE (1997) IEEE Std 802.11 Information technology—telecommunications and information exchange between systems—local and metropolitan area networks—specific requirements—part 11: wireless lan Medium Access Control (MAC) and Physical Layer (PHY) specifications. IEEE, New York
2. FE Solutions (2008) AirMagnet. http://fe-solutions.com
3. Akyildiz IF, Weilian S, Sankarasubramaniam Y, Cayirci E (2002) A survey on sensor networks. IEEE Commun Mag 40:102–114 (August)
4. Bahety V, Pendse R (2004) Scalable QoS provisioning for mobile networks using wireless sensors. In: Wireless communications and networking conference, vol 3. IEEE, New York, pp 1528–1533 (March)
5. Bianchi G, Fratta L, Oliveri M (1996) Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LAN. In: Seventh IEEE international symposium on personal, indoor and mobile radio communications, PIMRC'96, vol 2. IEEE, New York, pp 392–396 (October)
6. I.-T. Recommendation G.114 (2003) General characteristics of international telephone connections and international telephone circuits: oneway transmission time. IEEE, New York (January)
7. Gobriel S, Melhem R, Mosse D (2004) A unified interference/collision analysis for power-aware adhoc networks. In: Twenty-third annual joint conference of the ieee computer and communications societies (INFOCOM 2004), vol 1. IEEE, New York, p 608 (March)
8. Hsieh TT (2004) Using sensor networks for highway and traffic applications. IEEE Potentials 23(2):13–16 (April–May)
9. IETF (1998) RFC 2475: an architecture for differentiated services. IETF, Fremont
10. Kumagai J (2004) Life of birds. IEEE Spectrum 41(4):42–49
11. Kwon K, Lee C (2004) A fast handoff algorithm using intelligent channel scan for IEEE 802.11 WLAN. In: The 6th international conference on advanced communication technology, vol 1. IEEE, New York, pp 46–50 (February)
12. Wei Li, Zeng Q-A, Agrawal DP (2003) A reliable active scanning scheme for the IEEE 802.11 MAC layer handoff. In: Radio and wireless conference. IEEE, New York, pp 71–74 (August)
13. Lindgren A, Almquist A, Schelen O (2001) Evaluation of quality of service schemes for IEEE 802.11 wireless LANs. In: 26th annual IEEE conference on local computer networks. IEEE, New York, pp 348–351 (November)
14. MeshDynamics (2008) http://www.meshdynamics.com/index.html.
15. Mishra A, Shin M, Arbaugh W (2003) An empirical analysis of the IEEE 802.11 MAC layer handoff process. SIGCOMM Comput Commun Rev 33(2):93–102
16. Montavont N, Noel T (2002) Handover management for mobile nodes in ipv6 networks. IEEE Commun Mag 40(8):38–43 (August)
17. Nemeroff J, Garcia L, Hampel D, DiPierro S (2001) Application of sensor network communications. In: Military communications conference (MILCOM), vol 1. IEEE, New York, pp 336–341 (October)
18. Scalable Networks (2008) Qualnet. http://www.scalable-networks.com
19. Shin S, Forte AG, Rawat AS, Schulzrinne H (2004) Reducing MAC layer handoff latency in IEEE 802.11 wireless LANs. In: Proceedings of the second international workshop on mobility management & wireless access protocols. Association for Computing Machinery, New York, pp 19–26
20. Szewczyk R, Osterweil E, Polastre J, Hamilton M, Mainwaring A, Estrin D (2004) Habitat monitoring with sensor networks. Commun ACM 47(6):34–40
21. Takagi H, Kleinrock L (1984) Optimal transmission ranges for randomly distributed packet radio terminals. IEEE Trans Commun 32:246–257 (March)
22. Varshney U (2003) The status and future of 802.11-based WLANs. Computer 36(6):102–105 (June)
23. Velayos H, Karlsson G (2004) Techniques to reduce the IEEE 802.11b handoff time. In: IEEE international conference on communications, vol 7. IEEE, New York, pp 3844–3848 (June)
24. Yu L, Min-hua Y, Hui-min Z (2003) The handoff schemes in mobile ip. In: The 57th IEEE semiannual vehicular technology conference, vol 1. IEEE, New York, pp 485–489 (April)
25. Zhu H, Li M, Chlamtac I, Prabhakaran B (2004) A survey of quality of service in IEEE 802.11 networks. IEEE Wireless Commun 11(4):6–14