

Performance of Dynamic Service Addition in Mobile WiMAX Networks

Filippo Meucci^o, Laura Pierucci^o, Isabella Cerutti*,
Piero Castoldi* and Enrico Del Re^o

* Scuola Superiore Sant'Anna, Pisa, Italy, E-mail: {isabella.cerutti,p.castoldi}@sssup.it

^o Università di Firenze, Firenze, Italy, E-mail: {filippo.meucci,laura.pierucci,enrico.delre}@unifi.it

Abstract—In WiMAX mobile networks (IEEE 802.16e-802.16m), connections, also referred to as service flows, can be dynamically activated between the base station and the mobile stations, by using a three-way handshake protocol referred to as Dynamic Service Addition (DSA). However, degraded channel quality and unreliable message retransmissions may lead to a delayed or even unsuccessful activation of a service flow.

In this paper, a thorough analysis of DSA protocol performance is carried out under a variety of scenarios. Blocking probability, admission control probability, and latency of DSA protocol are evaluated, by means of simulations, for different conditions of mobility and parameters of PHY and MAC layers. Results show the negative impact of a long channel coherence time caused by low mobility and of the loss of channel reciprocity. Results offer indications on how to compensate such effects.

I. INTRODUCTION

Wireless networks are facing a number of challenges imposed by user mobility, increased transmission rates, and quality of service (QoS) requirements. The recently standardized IEEE 802.16 protocol [1], Mobile WiMAX 2009, aims at providing high data rates for mobile users by exploiting Orthogonal Frequency Division Multiplexing (OFDM) transmissions. OFDM-based physical layer is robust against Inter Subcarrier Interference (ISI) and mitigates the negative effects of multipath fading and scattering, that are severe in metropolitan environment. Further advancements are going to be achieved by the IEEE 802.16m standard. Among others, fast mobility (i.e., for up to 350 km/h) will be supported.

At the MAC layer, IEEE 802.16 permits to easily manage and guarantee the requested QoS to each connection, referred to as *service flows*, established between the Base Station (BS) and the Mobile Stations (MSs) [2]. Service flows with QoS requirements can be dynamically established upon requests, by successfully completing a three-way handshake procedure, referred to as Dynamic Service Addition (DSA). The DSA protocol is based on a request message, a response message, and an acknowledgment message, exchanged between BS and MS. Service flow is activated and guaranteed the requested QoS, when the protocol terminates successfully (i.e., both the response and the acknowledgment are positive and the corresponding messages are received).

This work was supported by MIUR under FIRB project "Software and Communication Platforms for High-Performance Collaborative Grid" (RBIN043TKY).

However, the unreliable nature of the radio medium may seriously compromise DSA message exchange. Thus, to ensure a successful reception of DSA messages, DSA protocol is provided with a robust retransmission strategy. When a reply to a transmitted DSA message (e.g., the request or the response message) is not received correctly within a timeout, a retransmission of the same DSA message can be attempted. Maximum number of message transmission attempts and timeout duration can be flexibility selected within a wide range of values defined by the standard. Although robust, DSA signaling can still be blocked if all the multiple attempts of sending a request (or a response) message fail. Thus, service flow requests can be blocked due to unavailable resources to meet QoS (i.e., admission control blocking) as well as erroneous termination of DSA protocol (i.e., signaling blocking). Blocking and latency of DSA protocol are both affected by the selection of timeout values and maximum number of attempts.

The impact of the unreliable radio medium on the radio transmissions has been extensively studied. For instance, a number of works addressed the performance of data transmissions [3] and automatic repeat request protocol [4] in WiMAX networks. Also, strategies to ensure QoS of activated service flows [5], [6], cross-layer approaches that adapt to the channel conditions [4], [7] and service flow scheduling [8]–[10] have been proposed for WiMAX networks. Although relevant, these works are unable to capture the peculiar behavior of DSA protocol and its performance, especially in a mobile environment. The problem is that a degradation of the radio channel, for instance due to mobility, could lead to an increase of DSA protocol blocking and in turn to a reduction of the overall network performance. Similar problems occur also during the message exchange for handover procedure and are investigated in [11].

In this paper, the performance of DSA protocol are thoroughly evaluated in a mobile networks based on OFDM physical layer in Time Division Duplex (TDD). The Physical (PHY) layer performance has a strong impact on the MAC performance, especially in the presence of mobility, which cannot be easily derived numerically. For this reason, MAC layer simulation of DSA protocol is jointly combined with an accurate PHY layer simulation that accounts for both OFDM signal degradation due to Doppler shift generated by mobility

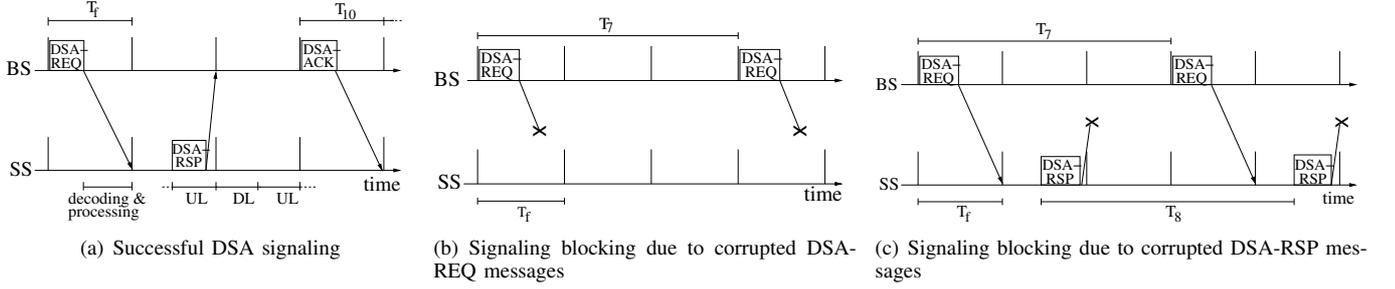


Fig. 1. Time chart of DSA signaling

and frequency selectivity caused due to multipath.

DSA is investigated in terms of signaling blocking, admission control blocking, and latency for a variety of scenarios, i.e., for different channel quality conditions, degrees of reciprocity between Downlink (DL) and Uplink (UL) radio channels, MS speeds, OFDM configurations and MAC layer parameters.

The main contribution of the paper is the comprehensive study of the impact of the different scenarios and parameters on DSA protocol performance. Simulation results aim at quantifying the impact of MS speeds and PHY layer parameters on MAC performance. This helps to derive useful considerations on how to counter-react the performance degradation due to high speeds, long channel coherence time at low (pedestrian) speeds, and low reciprocity between DL and UL channels.

II. DYNAMIC SERVICE FLOW ACTIVATION (DSA) SIGNALING

Consider a wireless network based on IEEE 802.16 operating in point-to-multipoint mode, i.e., the BS communicates with a number of MSs. DL and UL transmissions are scheduled by the BS in TDD mode, within each time frame. Time frame duration is fixed and indicated as T_f .

When a DL or UL service flow is requested, admission control operations are executed. If the service flow is admitted, the service flow can be activated using a three-way handshake between the BS and the MS. In the following, a BS-initiated handshake is considered¹. For simplicity, in the explanation a single MS is assumed.

Signaling is triggered by the BS by sending a DSA-REQ message (i.e., DSA request) to the MS. Upon correct reception of the DSA-REQ message, the MS replies with a DSA-RSP (i.e., DSA response). Upon correct reception of the DSA-RSP message, the BS replies with a DSA-ACK (i.e., DSA acknowledgement). Service flow is activated at the BS after the expiration of timeout T_{10} . This case of successful signaling is sketched in Fig. 1(a).

The messages are sent on the primary management connection, using the most robust PHY layer profile. However, due to the unreliability of the radio channel, one or more of the messages may be lost or incorrectly received. For this purpose,

¹According to IEEE 802.16 standard, support of BS-initiated signaling is a mandatory requirement, while support of MS-initiated signaling is optional.

DSA-REQ and DSA-RSP messages may be retransmitted multiple times, upon expiration of a timeout.

After sending the DSA-REQ message, the BS triggers a timer. If no DSA-RSP is correctly received within timeout T_7 , BS can retransmit another copy of the DSA-REQ message. BS can transmit up to n_R copies of a DSA-REQ message. This case is represented in Fig. 1(b), for $n_R = 2$.

After sending the DSA-RSP message, the MS triggers a timer. If no DSA-ACK is correctly received within timeout T_8 , the MS can retransmit another copy of the DSA-RSP message. MS can transmit up to n_S copies of a DSA-RSP message. This case is represented in Fig. 1(c), for $n_S = 2$.

A service flow request can be blocked due to:

- admission control: BS may decide to not admit a service flow when the available bandwidth or QoS cannot be guaranteed;
- erroneous termination of signaling: when the BS does not receive any DSA-RSP from the MS, the requested service flow is not activated, as shown in Figs. 1(b) and 1(c);
- negative responses: for various reasons, MS (BS) can reply with a negative DSA-RSP (negative DSA-ACK). In such cases, the requested service flow is not activated.

After characterizing the PHY layer, the DSA performance are evaluated.

III. MOBILE OFDM CHANNEL

As standardized in [12], mobility is supported by PHY layer based on OFDM. In the OFDM physical layer, each time frame is composed of N_{OFDM} OFDM symbols. Each OFDM symbol consists of N_{sc} subcarriers. The frame duration is $T_f = N_{OFDM} \cdot T_{OFDM}$, where T_{OFDM} is the OFDM symbol duration and $T_{OFDM} = \frac{1}{\Delta f}$ where Δf is the subcarrier frequency separation. Given the maximum MS velocity and the maximum level of tolerable InterCarrier Interference (ICI), a minimum Δf (i.e., a maximum duration of the OFDM symbol) is derived according to [13]. The frequency separation is usually set to a value that ensures ICI below -27 dB and typical values of T_{OFDM} are in the order of tens of microseconds and time frame duration is in the range of [1,20] ms.

A DSA message can require one or more OFDM symbols depending on the PHY layer configuration. A DSA message is considered in error if the OFDM symbol/s is/are not decoded

correctly, i.e. if the BPSK symbol carried by any subcarrier is detected incorrectly. No FEC is assumed.

The probability that a single OFDM symbol is received incorrectly depends on the subcarrier SNR distribution over the frequency domain μ_{ij} which is related to the multipath channel delay spread. Multipath channels with Rayleigh distributed paths result in Rayleigh distributed μ_{ij} in the frequency domain, [14]. The Power Delay Profile (PDP) determines the subcarrier SNR distribution over the frequency domain: $\mu_{ij} = \frac{E_s}{N_0} \|\mathbf{H}(i, j)\|^2$, where $\frac{E_s}{N_0}$ is the SNR and $\mathbf{H}(i, j)$ is the channel response in the frequency domain for the i -th subcarrier for the j -th OFDM symbol.

Mobility affects the correlation of the instantaneous SNR distribution for each subcarrier in the time domain. For low and moderate MS speeds (e.g., pedestrian), the SNR of a subcarrier is highly correlated to the SNR of the same subcarrier experienced in the previous time-frame(s). This correlation leads to a high probability that retransmitted DSA messages are all incorrectly received. However, at high speeds, the correlation of subcarrier SNR in consecutive time frames becomes negligible. Next, the impact of mobility effects on the MAC layer performance is quantified.

IV. ANALYSIS OF DSA PROTOCOL PERFORMANCE

In this section, DSA protocol performance is evaluated in terms of:

- signaling blocking probability: it is the probability of erroneous termination of signaling. It is the probability that no DSA-RSP is correctly received by the BS within T_{10} timeout in response to DSA-REQ message, i.e., the probability that n_R DSA-REQ (or n_S DSA-RSP) are incorrectly received;
- admission control blocking probability: it is the probability that a service flow request cannot be admitted due to lack of available resources. It is derived under the assumption that a fixed number m of service flow requests can be accommodated, i.e., each service flow requests $1/m$ of the available bandwidth in the IEEE 802.16 network. Therefore, it is the probability that m service flows are already active in the network, when a new service flow request arrives.
- signaling latency: it is the time interval from the generation of a new service flow request to the instant in which data transmission can take place, i.e., it accounts for time that it takes to generate the first DSA-REQ, to receive correctly a DSA-REQ, a DSA-RSP, and a DSA-ACK, plus timeout T_{10} .

To quantify the performance, MAC and PHY layers are implemented as described in Section II and III in a C/C++ custom-made event-driven simulator. Simulation results are collected to achieve a confidence interval of 15% (or better) at 10% confidence level, using the following configuration, unless otherwise indicated.

Parameters of the OFDM physical layer are: $N_{sc} = 1024$, $\Delta f = 11.16071429$ kHz, $T_{OFDM} = 102.86$ μ s, $N_{OFDM} = 48$, $T_f = 5$ ms [15]. With this selection of N_{sc} , each DSA

message can be accommodated on a single OFDM symbol. Signal modulation is uncoded BPSK and carrier frequency is 3.5 GHz. The channel has analyzed with ITU Pedestrian and Vehicular Type A channel models [16].

At the MAC layer, the timeouts are set to: $T_7 = 3 \cdot T_f$, $T_8 = 3 \cdot T_f$, $T_{10} = 4 \cdot T_f$. Scheduling latency for transmitting DSA messages (i.e., buffering delay) is considered negligible. Latency due to signal and information processing at the PHY layer is accounted as indicated in Fig. 1(a), i.e., half time frame is required for decoding the message and generating a reply. Maximum number of copies of DSA-REQ and DSA-RSP messages are $n = n_R = n_S \in \{1, 3\}$. Responses to DSA-REQ and DSA-ACK are always positive, i.e., no blocking for negative responses. Ratio of inter-arrival rate (λ) of service flow requests over service flow duration is set to 20. Inter-arrival times ($1/\lambda$) and duration of the service flow requests are exponentially distributed. Inter-arrival rate of DSA requests is set to $\lambda = 20$ s^{-1} and expected duration of service flows is set to 1 s. Up to $m = 5$ service flows can be activated.

Effects of MS mobility are taken into account by implementing a channel model with Jakes Doppler spectrum. The ground speed, v , of the MS with respect to the BS is selected in the range [5, 300] km/h. For comparison purposes, a single-path block-fading channel without mobility is considered and indicated as $v = 0$ and used to assess the correctness of the simulator. DL and UL channels are assumed to be independent.

A. Impact of Mobility

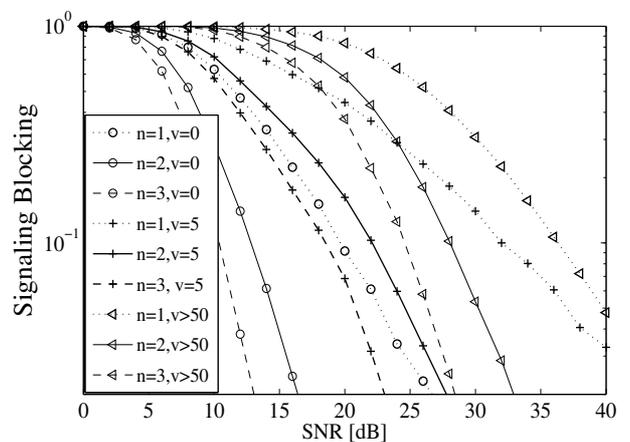


Fig. 2. Signaling blocking vs. SNR for different values of v and n

Figs. 2-4 show the impact of MS mobility in terms of signaling blocking, admission control blocking and latency for activating a service flow, respectively, when $n = 1$ (no retransmission), 2, and 3.

The latency is reported only for activated service flows. Therefore, when $n = 1$, the latency is constant. Mobility strongly impacts the DSA performance in terms of signaling blocking and latency, due to the deterioration of channel performance. However, when MS speed exceed 50 km/h (i.e., from 50 km/h to 300 km/h), DSA performance does not degrade any further. Signaling blocking can be reduced by

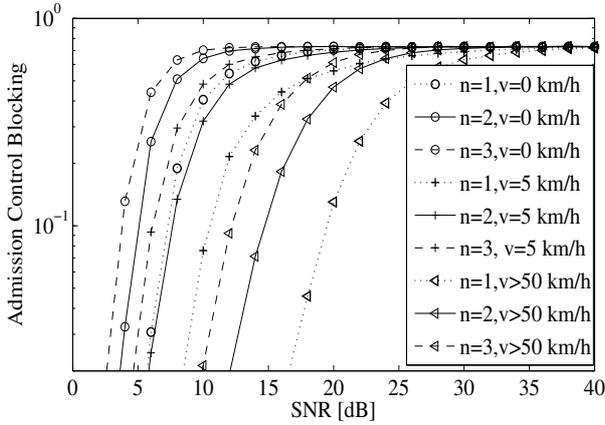


Fig. 3. Admission control blocking vs. SNR for different values of v and n

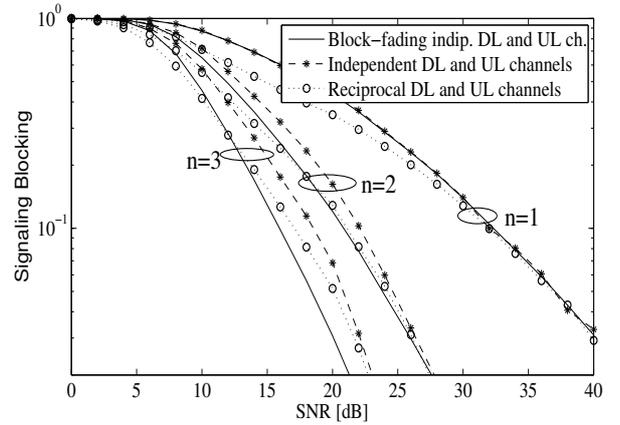


Fig. 5. Signaling blocking vs. SNR for $v = 5$ km/h, when UL and DL channels are either independent or fully reciprocal

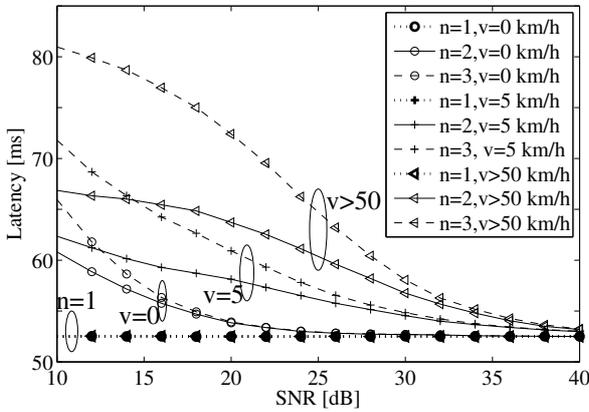


Fig. 4. Latency vs. SNR for different values of v and n

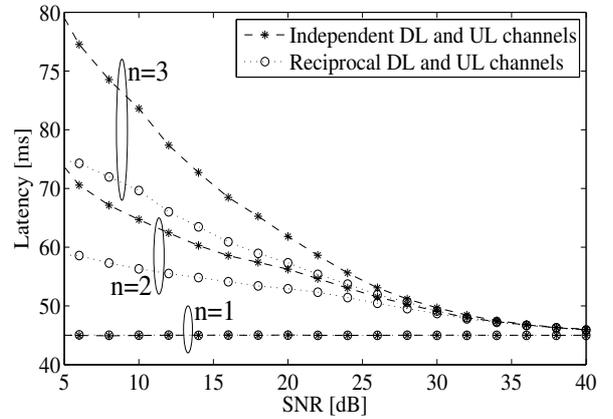


Fig. 6. Latency vs. SNR for $v = 5$ km/h, when UL and DL channels are either independent or fully reciprocal

resorting to higher number of transmission attempts (n), to detriment of an increased latency. For instance, to ensure a signaling blocking of 10^{-1} , a gain of about 6 dB can be achieved by increasing n to 2 and a further gain of about 2 dB can be achieved at $n = 3$, when $v = 0$. The gain is slightly higher (about 8 and 3 dB respectively) at high speed ($v > 50$ km/h) and is even higher at pedestrian speed (about 10 dB and 4 dB respectively). These improvements are achieved at the expenses of a latency increase of about 6 ms (for $n = 2$) and 12 ms (for $n = 3$), for $v = 0$ and $v > 50$. Interestingly, in addition to achieving higher SNR gains, pedestrian mobility suffers a smaller latency increase, i.e., 5 and 10 ms respectively. Contrary to signaling blocking, admission control blocking is reduced with mobility. This happens because the number of service flow requests blocked by erroneous termination of signaling increases with MS speed and, thus, the network load decreases.

B. Impact of Correlation between UL and DL Channels

Figs. 5 and 6 compare the signaling blocking and the latency, respectively, experienced at $v = 5$ km/h, when the DL and UL channels are reciprocal (i.e., DL and UL are considered as a single fading channel) or independent. In

Fig. 5, an independent block-fading channel is also considered for comparison and indicates that the blocking is negatively affected by the channel correlation.

Reciprocity is always beneficial when $n = 1$. At low SNR, it may even compensate the negative effects of channel correlation. Moreover, the correlation between the DL and UL channels greatly lowers the latency of the activated service flows, especially at low SNR as shown in Fig. 6, i.e. if a DSA-REQ is correctly received, with high probability the corresponding DSA-RSP is also correctly received.

C. Impact of the Number of Subcarriers per OFDM Symbol

IEEE802.16e can be deployed over various spectrum bandwidths via flexible PHY layer configuration: scalable OFDMA [15]. The used spectrum depends on the number of subcarriers (N_{sc}). The impact of N_{sc} over the DSA performance is analyzed in this section. When $N_{sc} < 1024$, the DSA messages need to be sent over two or more consecutive OFDM symbols. A DSA message is received incorrectly when at least one of OFDM symbols is in error. In Figs. 7 and 8, the signaling blocking is evaluated as a function of SNR for various values of N_{sc} , when $v = 50$ km/h and $v = 300$ km/h,

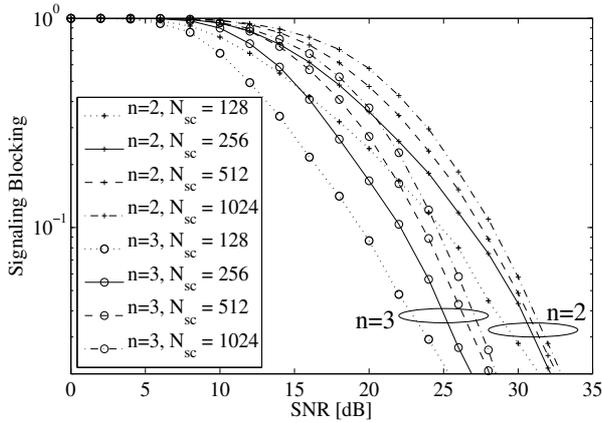


Fig. 7. Signaling blocking vs. SNR = 20 dB, for $v = 50$ km/h and different values of n and N_{sc}

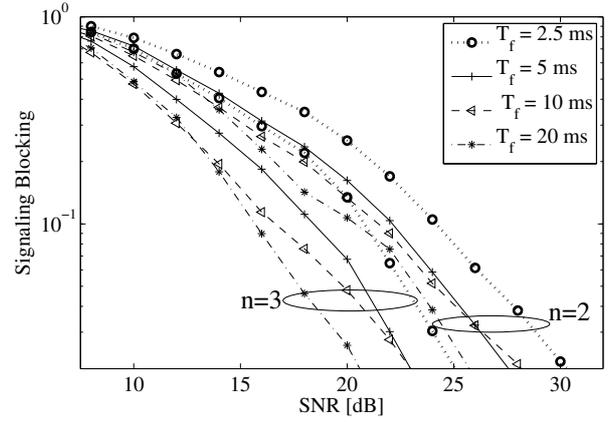


Fig. 9. Signaling blocking vs. SNR for $v = 5$ km/h, when $T_f = 2.5, 5, 10,$ and 20 ms

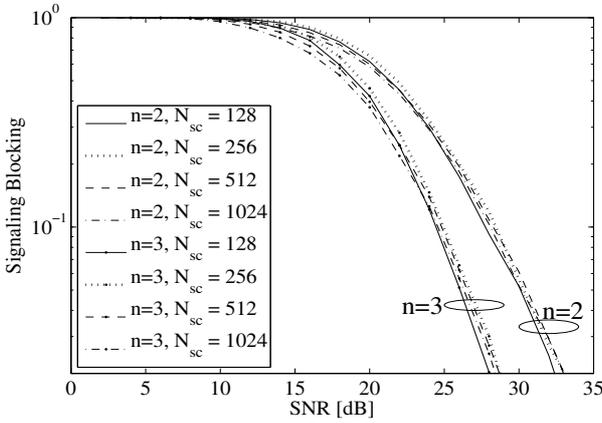


Fig. 8. Signaling blocking vs. SNR = 20 dB, for $v = 300$ km/h and different values of n and N_{sc}

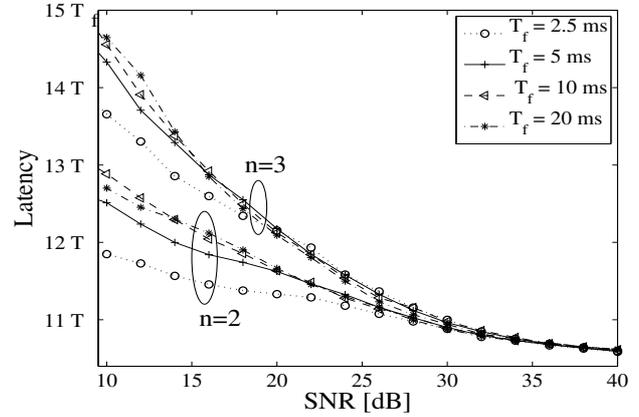


Fig. 10. Latency vs. SNR for $v = 5$ km/h, when $T_f = 2.5, 5, 10, 20$ ms

respectively. At 50 km/h, the correlation of the channel is favorable: if an OFDM symbol is received correctly, with high probability the consecutive OFDM symbols will be received correctly too, leading to a correct reception of a DSA message. Thus, the channel time correlation combined with the small values of N_{sc} , lowers the blocking.

At 300 km/h, the difference among the curves at different N_{sc} is narrowed. In this case, the performance for a low number of subcarriers can worsen. This is because, at high velocities, the channel varies sensibly during consecutive OFDM symbols and thus the probability that two or more consecutive OFDM symbols are received correctly can be lower than in the single symbol case (i.e., $N_{sc} = 1024$).

D. Impact of Time Frame Duration

Fig. 9 shows the impact of time frame duration (T_f) on the signal blocking probability as a function of the SNR, for $v = 5$ km/h. The signal blocking probability decreases not only with a higher number of retransmissions but also with longer frame duration. Indeed, by using long time frames, the channel correlation can be reduced, with a beneficial effect on the signal blocking probability. The effect of T_f

duration is particularly evident at high SNR, where the channel performance is dominated by the fading coherence time with respect to the additive white noise.

Fig. 10 shows the impact of T_f on the latency normalized in number of time frames, for $v = 5$ km/h. Since the normalized latency for various T_f is comparable while T_f may up to 8 times longer, it is evident that low values of T_f are preferable for a fast activation of service flows. However, in a practical case, the value of T_f is lower bounded to avoid a loss of throughput due to the overheads (e.g., for synchronization and data maps).

These results indicate that, at pedestrian speed, by using longer time frame durations, it is possible to exploit the time diversity and, thus, to better diversify the probability of a successful reception. This permits to achieve lower signaling blocking to detriment of a higher latency. At higher speeds ($v \geq 50$) or at $v = 0$, timeout duration affects the latency only (not shown here).

E. Impact of Timeout Duration

In addition to time frame duration, the channel correlation experienced by successive transmission of DSA messages can

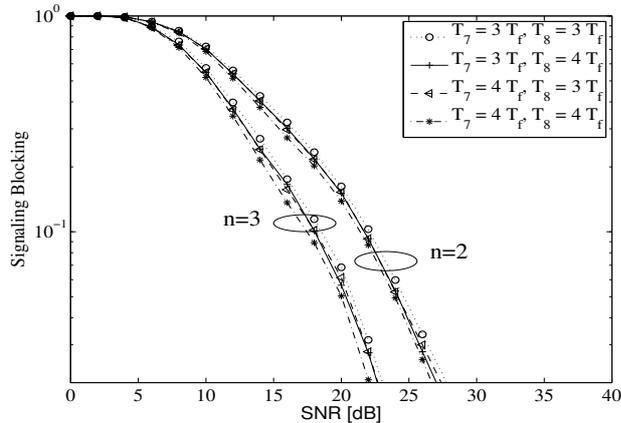


Fig. 11. Signaling blocking vs. SNR for different values of timeouts and $v = 5$ km/h

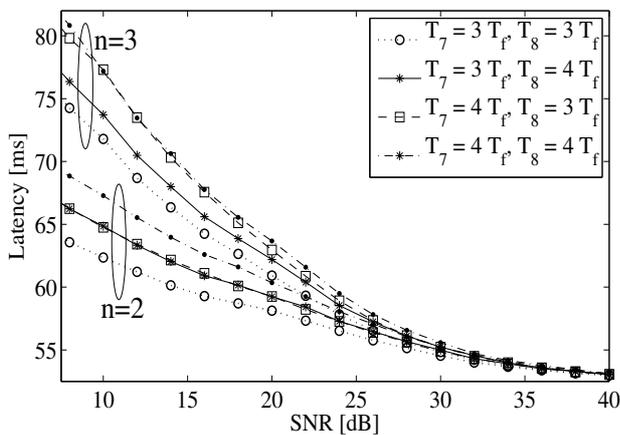


Fig. 12. Latency vs. SNR for different values of timeouts and $v = 5$ km/h

be controlled by varying the duration of DSA timeouts. Figs. 11 and 12 show the impact of timeout duration on DSA performance, for increasing SNR and $v = 5$ km/h. Consistently with the results in Fig. 9, the lower signaling blocking is achieved for longer timeout duration (i.e., $T_7 = T_8 = 4T_f$ in the figures) as time diversity can be exploited. Even in this case, the price to pay for the reduced blocking is an increase of the latency. At MS speeds of $v = 0$ and $v \geq 50$ km/h, timeout duration affects the latency performance only (not shown here).

V. CONCLUSIONS

The paper analyzed the performance of the Dynamic Service Addition (DSA) protocol. An extensive performance evaluation of the DSA protocol has been carried out through simulations for a variety of scenarios. Signaling blocking, admission control blocking, and latency experienced by DSA protocol have been quantified for different channel and mobility conditions and various PHY and MAC parameters.

Results indicate that robustness of DSA protocol can be increased by increasing the number of message transmission attempts, at the expenses of an increased latency. This is

especially important for fast moving MS. Indeed, mobility has been shown to have a detrimental effect on the DSA performance.

Contrary to fast moving MS, slowly moving MS (e.g., pedestrian speed) suffers from the effect of the long channel coherence times and from the independence of DL and UL channels. Loss of reciprocity between DL and UL channels can be recovered with calibration techniques [17] which could be available in future networks. To compensate the negative impact of the long channel coherence time on the signaling blocking, time diversity could be exploited, equivalently attainable by selecting longer time frames or longer duration of timeouts, at the expenses of the latency. In addition, a careful selection of the number of OFDM subcarriers (e.g., as in OFDMA), that are used to carry the DSA message, may help to reduce both signaling blocking and latency.

REFERENCES

- [1] "IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems," *IEEE Std 802.16-2009 (Revision of IEEE Std 802.16-2004)*, pp. C1–2004, 29 2009.
- [2] C. Cicconetti, L. Lenzini, E. Mingozzi, and C. Eklund, "Quality of service support in IEEE 802.16 networks," *IEEE Network*, vol. 20, no. 2, pp. 50–55, March–April 2006.
- [3] M. Tran, A. Doufexi, and A. Nix, "Mobile WiMAX MIMO performance analysis: Downlink and uplink," in *Proc. PIMRC*, Sept. 2008, pp. 1–5.
- [4] I. Cerutti, F. Meucci, P. Castoldi, and L. Pierucci, "An adaptive cross-layer strategy for QoS-guaranteed links in 4G networks," in *Proc. IEEE GLOBECOM*, 2008.
- [5] Q. Liu, X. Wang, and G. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, pp. 839–847, May 2006.
- [6] X. Bai, A. Shami, and Y. Ye, "Robust QoS control for single carrier PMP mode IEEE 802.16 systems," *IEEE Transactions on Mobile Computing*, vol. 7, no. 4, pp. 416–429, April 2008.
- [7] S.-E. Elayoubi and B. Foubrestie, "Performance evaluation of admission control and adaptive modulation in OFDMA WiMax systems," *IEEE/ACM Trans. on Networking*, vol. 16, no. 5, Oct. 2008.
- [8] C. So-In, R. Jain, and A.-K. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 2, pp. 156–171, 2009.
- [9] J. Borin and N. da Fonseca, "Scheduler for IEEE 802.16 networks," *IEEE Communications Letters*, vol. 12, no. 4, pp. 274–276, April 2008.
- [10] A. Lera, A. Molinaro, and S. Pizzi, "Channel-aware scheduling for QoS and fairness provisioning in IEEE 802.16/WiMAX broadband wireless access systems," *IEEE Network*, vol. 21, no. 5, pp. 34–41, 2007.
- [11] Z. Yan, L. Huang, and C.-C. Kuo, "Seamless high-velocity handover support in mobile WiMAX networks," in *IEEE Singapore International Conference on Communication Systems*, Nov. 2008, pp. 1680–1684.
- [12] IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005, *Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, 2005.
- [13] Y. Li and J. Cimini, L.J., "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 401–404, 2001.
- [14] Z. Kang, K. Yao, and F. Lorenzelli, "Nakagami-m fading modeling in the frequency domain for OFDM system analysis," *IEEE Communications Letters*, vol. 7, no. 10, pp. 484–486, Oct. 2003.
- [15] H. Yaghoobi, "Scalable OFDMA physical layer in IEEE 802.16 WirelessMAN," *Intel Technology Journal*, vol. 8, no. 3, pp. 201–212, 2004.
- [16] T. B. Sorensen, P. E. Mogensen, and F. Frederiksen, "Throughput enhancement in WiMax mesh networks using concurrent transmission," in *Proc. IEEE Vehicular Technology Conference (VTC)*, October 2005.
- [17] M. Guillaud, D. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. Signal Processing and Its Applications*, vol. 1, 2005, pp. 403–406.