[SMHDSt06] Christian Schwingenschlögl, Parag S. Mogre, Matthias Hollick, Volker Dastis, and Ralf Steinmetz; **Performance Analysis of the Real-time Capabilities of Coordinated Centralized Scheduling in 802.16 Mesh Mode**; 63th IEEE Semiannual Vehicular Technology Conference (VTC-Spring 2006), Melbourne, Australia, May 2006.

Performance Analysis of the Real-time Capabilities of Coordinated Centralized Scheduling in 802.16 Mesh Mode

Christian Schwingenschlögl and Volker Dastis
Siemens AG
Corporate Technology (CT IC 2)
Dept. of Information and Communications
Otto-Hahn-Ring 6, 81730 Munich, Germany
Contact Email: chris.schwingenschloegl@siemens.com

Parag S. Mogre, Matthias Hollick, and Ralf Steinmetz
Multimedia Communications Lab (KOM)
Technische Universität Darmstadt
Dept. of Electrical Engineering and Information Technology
Merckstrasse 25, 64283 Darmstadt, Germany
Contact Email: parag.mogre@KOM.tu-darmstadt.de

***Abstract*— The IEEE 802.16-2004 standard specifies wireless broadband networks with optional support for multi-hop mesh operation (mesh mode). The provision and support of high-quality real-time services such as voice over IP is crucial, if wireless networks based on the IEEE 802.16-2004 standard are to challenge wired network services. In this paper we investigate and identify critical factors in enabling real-time services in 802.16 based networks operating in the mesh mode. We present an analytical performance analysis and a simulation study investigating the coordinated centralized scheduling mechanism as specified in the 802.16-2004 standard. Our results show that the scalability and efficiency of such mesh networks with respect to real-time services are at stake. Our results, moreover, aid in the adjustment of critical system parameters allowing for optimized network performance.**

***Index Terms*— Broadband Wireless Access, IEEE 802.16, Modeling and Performance Evaluation, QoS, Dynamic Bandwidth Allocation, Ad hoc Networks.**

I. Introduction

Today, a burgeoning growth in the demand for broadband access networks can be witnessed. This growth is accompanied by the proliferation of real-time multimedia services such as voice over IP (VoIP). Traditionally such demands are met by wired networks, e.g., cable or DSL networks. Targeting the same applications and services, the IEEE 802.16-2004 standard [1] specifies a set of air interfaces as well as a medium access control (MAC) layer for fixed broadband wireless access systems. In contrast to widely deployed wireless technologies, the 802.16 standard is connection oriented and enables the specification of quality of service (QoS) and scheduling services on a per connection basis. This provides a powerful mechanism for enabling multimedia services with strict QoS requirements. As a result, networks based on the 802.16 standard are foreseen to provide a cost-effective and viable alternative to the traditional last-mile access networks.

The 802.16 standard identifies two primary modes of operation, namely, the point to multi-point (PMP) and the multi-hop mesh mode (MSH) of operation. The standard specifies that in the PMP mode, all the subscriber stations (SS) are to be in direct single-hop neighborhood of the base station (BS). As an extension to the PMP mode the MSH mode allows inclusion of SSs in the network that do not have a direct connection to the base station by means of relaying via intermediate nodes, thus, enabling multi-hop communication. To enable ubiquitous network coverage within broadband wireless community networks or enterprise wide mesh networks, 802.16's MSH mode of operation is considered to be of utmost importance.

The 802.16 standard defines that a BS is any node, which connects the rest of the network to external networks or the Internet. A majority of traffic in the real-time multimedia communications is expected to flow from clients in the mesh network to the external networks (via the BS) and vice versa. To support such traffic to and from the BS in the multi-hop MSH mode, the 802.16 standard specifies a coordinated centralized scheduling mechanism (MSH-CSCH). Recent work of Redana and Lott [2] shows that for traffic to and from the mesh BS, the centralized scheduling can outperform distributed scheduling in terms of throughput at the MAC layer.

Our goal is to investigate the performance of the MSH-CSCH mechanism for real-time traffic. We are particularly interested in the efficiency of the scheduling with respect to (a) scalability, i.e., supporting a large number of users, and (b) performance, i.e., guaranteeing an acceptable end-to-end delay. Our contribution is a thorough analysis of the parameterization of the MSH-CSCH mechanism. In particular, we investigate the partitioning of the individual frames into control subframe and data subframe, which is controlled by the MSH-CRTL-LEN parameter. To assess the performance and scalability of the network, we give both, analytical results as well as results obtained by means of a simulation study. Our results clearly show that there exists a trade-off between scalability and performance. Choosing the MSH-CRTL-LEN parameter appropriately can largely influence this trade-off.

The full paper is organized as follows. In Section II we discuss related work and introduce the different scheduling mechanisms specified in 802.16. Section III provides an analytical analysis of coordinated centralized scheduling for admitting real-time flows in the network. In Section IV we describe the simulation setup and environment and provide a critical analysis of the results. We discuss the performance and scalability of 802.16 mesh networks as far as support for real-time traffic is concerned. Section V summarizes our findings and concludes this work.

II. Background and Related Work

The 802.16 standard offers a promising alternative to traditional last-mile broadband wired access networks. Moreover, it promotes the possibility of organic growth of community mesh-networks. To serve these differing needs, the standard specifies a set of mechanisms on the MAC and physical (PHY) layer. The main subject of our investigation, namely the specified MAC layer scheduling mechanisms, can be distinguished into centralized and distributed scheduling. In mesh mode the network can use coordinated centralized scheduling (MSH-CSCH) and distributed scheduling (MSH-DSCH), which is further differentiated into coordinated distributed scheduling and uncoordinated distributed scheduling. While MSH-CSCH enables efficient scheduling over longer periods of time, MSH-DSCH enables flexibility in the setup of short term bandwidth requests/grants and for traffic not transmitted via the BS.

Given the high potential, the 802.16 standard has been the focus of many recent studies. Eklund et al. [3] present an overview of the 802.16 air interface with the major focus being on the PMP mode of operation. In [4], Wongthavarawat et al. present a QoS architecture for the PMP mode of operation of the 802.16 standard enabling the support of the different scheduling services namely unsolicited grant service (UGS), real-time polling service (rtPS), non-real-time polling service (nrtPS), and best effort (BE). The above scheduling services can be associated with individual connections (all data transfer in the 802.16 network is within the context of connections and identified by a connection identifier), thus, enabling the specification and provision of different QoS levels to different types of traffic. Lee et al. [5] describe an efficient scheduling algorithm to enable support of a higher number of VoIP users without degradation of the provided QoS of the VoIP calls. However, as is the case with most of the related work on 802.16, the focus of [5] is the PMP mode of operation.

The recent work of Redana and Lott [2] is one of the few papers in literature, which studies 802.16's mesh mode of operation. The authors present a detailed study and comparative analysis of the centralized scheduling and the distributed scheduling in the 802.16 mesh mode. They compare the efficiency of the above scheduling mechanisms in terms of throughput obtained at the MAC layer versus the raw throughput obtainable at the PHY layer. The conclusion of [2] is that for stable, long-term traffic to and from the BS, centralized scheduling gives better efficiency in terms of overhead as compared to distributed scheduling. Cao et al. [6] develop a stochastic model for the distributed scheduler for analytical investigation of the mesh mode. The results obtained in [6] provide insights into the efficiency of the scheduling mechanism, particularly, the effects induced by variation of parameters left open by the standard specification (e.g., XmtHoldoffExponent). The results mandate to identify and study the relevant parameters for the centralized scheduling mechanism, as well. To this end, we here present a systematic study of the MSH-CTRL-LEN parameter on the efficiency of 802.16 networks with respect to real-time traffic. In addition we analyze the scalability of 802.16 mesh networks as a function of number of subscriber stations.

III. Theoretical Worst-case Analysis

In this section we present an theoretical analysis of the worst-case and best-case behavior for the scheduling delay in 802.16 mesh networks. For our study, we identified the following parameters to be of utmost importance: network size, scheduling overhead, and scheduling delay. To control scheduling delay and overhead, the 802.16 standard defines the MSH-CTRL-LEN variable (4 bit length). This variable controls the amount of bandwidth that is reserved for network control and scheduling messages, thus, determining the allocation of MSH-CSCH request and grant messages. An increase in MSH-CTRL-LEN yields a faster propagation of control messages and decreases scheduling delay, however, at the expense of increased control overhead. The following analysis focuses on the influence of the MSH-CTRL-LEN parameter, which acts as the central parameter to control the scheduling performance of the network.

We use the following parameterization for our analysis (see [1] for the explanation of the individual parameters): OFDM channelization parameters are MMDS with 24 MHz. With Tg=Tb/32, we obtain a frame length of 2.5 ms. For the chosen values, one frame consists of 252 minislots equaling 252 symbols. For control messages, the OFDM symbols are coded using 1/2 QPSK channel coding with RS-CC, which gives 24 bytes per symbol. Fig. 1 depicts the chosen parameterization. A numerical example illustrates the calculation of the worst-case and best-case delays. Our scenario is a small network

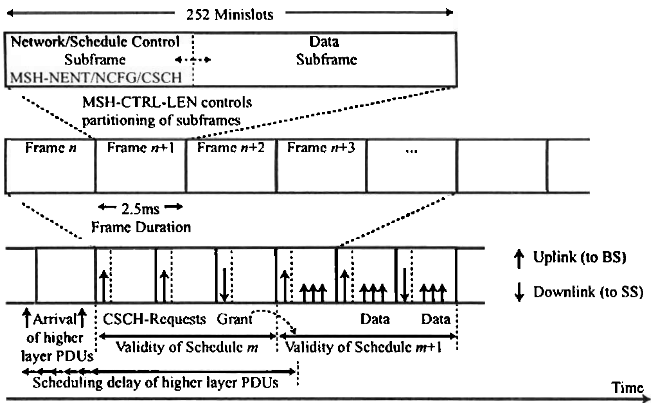

Fig. 1. Example for Coordinated Centralized Scheduling in 802.16.

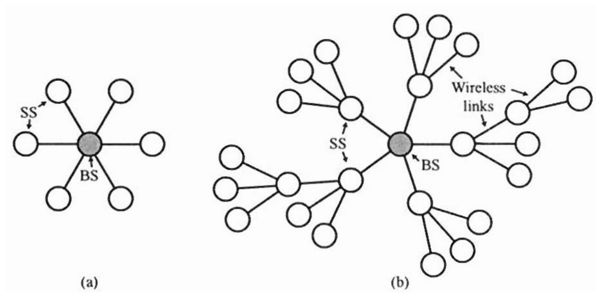

Fig. 2. Analyzed topologies with (a) 7 nodes and (b) 26 nodes.

with one BS and six SSs that are directly connected to the BS (see Fig. 2(a)). We assume that the BS can serve the bandwidth requests of all stations and neglect internal processing delays. Also, network control messages (MSH-NENT/NCFG) are not considered for our calculation. The MSH-CTRL-LEN is set to two yielding the first 14 symbols of the 252 symbols of every frame to be reserved for MSH-CSCH request and grant messages. Each message consists of two preamble symbols, one guard symbol, and one symbol for the MSH-CSCH request/grant. For the given network, the MSH-CSCH messages are smaller than 24 bytes and can be conveyed in one symbol. With increase in network size, the size of the message grows by one symbol every 20 nodes for the grant and every 40 nodes for the request.

The behavior of this sample network can be seen in Fig. 1. The system is repeating every three frames: in the first two frames the SSs send their uplink requests for bandwidth to the BS, in the third frame the BS sends the grant on the downlink. If an SS sends directly after receiving a message, a minCSForwardDelay of 5 Symbols has to be considered. To obtain the worst-case and best-case scheduling delay, we calculate the time from the arrival of a higher layer PDU, e.g. an IP packet, until its delivery (we assume that the packet is delivered during the validity of the next schedule granted). The worst case is marked if a higher layer PDU is queued directly after issuing a request, while the best case is marked if the PDU is queued directly before a request is issued. Moreover, the sequence of the requesting SS, which is determined by a network-internal index and the distance to the BS, influences the observed scheduling delay. For our numerical example (six SSs, MSH-CTRL-LEN = 2), the worst case delay is 25 ms. Fig. 3 shows the worst-case delays for three

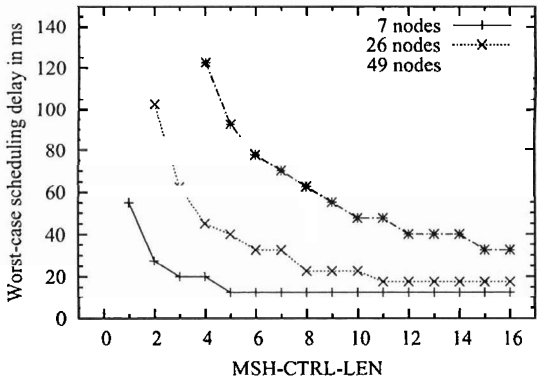

Fig. 3. Worst-case scheduling delay vs. MSH-CRTL-LEN for the investigated topologies.

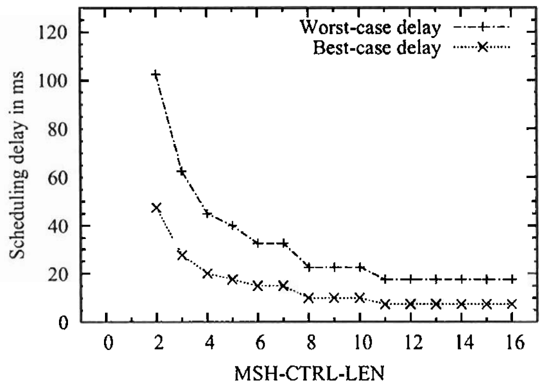

Fig. 4. Worst-case and best-case scheduling delay vs. MSH-CRTL-LEN for the 26 node setup.

different topologies (7 nodes, 26 nodes, 49 nodes). The x-axis gives the MSH-CTRL-LEN and the y-axis gives the scheduling delay in milliseconds. It is clearly visible that an increase in the MSH-CTRL-LEN parameter leads to a reduction of delay. This is because more request/grant messages can be transferred in a single frame. However, at the same time, the control subframe is increased at the expense of the data subframe, thus leading to a reduction of available bandwidth for data transmission. Also, the delay increases with network size, because the higher number of nodes lead to an increase in signalling traffic. Fig. 4 shows the worst-case and best-case delays for the 26 node setup given in Fig. 2(a). Again the effects of the variation of the MSH-CTRL-LEN parameter are clearly visible.

IV. Experimental Analysis

To enhance and complement our theoretical analysis of Section III, we performed a thorough simulation study. While the goal of our analysis remains the same, the simulation facilitates the reduction of the abstraction level for the modeling process. Thus, we are able to gain insights about average case results for the investigated scenarios. We describe our simulation model as well as the employed simulation platform in Subsection IV-A. The experimental design and the results are described and discussed in Subsection IV-B.

A. Simulation Model

As of today, we are not aware of standard-compliant implementations of the PHY/MAC of IEEE 802.16-2004 in network simulators. To be able to carry out our analysis, we implemented the functionality necessary to cover our problem. We have chosen the JiST [7] platform for our implementation. JiST allows for efficient discrete event simulation and follows a virtual machine-based simulation concept, i.e., JiST is based on a standard Java Virtual Machine (JVM). By embedding time semantics at byte-code level, the virtual machine-based simulation concept converts an existing JVM into a simulation platform. Since JiST capitalizes on the existing optimizations of the JAVA platform, time and memory consumption of the simulation environment are optimized as well. All nodes in our simulation are represented as entity-objects, thus, the event processing can be processed in a quasi-parallel manner. The simulation-time itself is independent from real-time or process advances and is accessible in every object. The handling of all time dependencies between the nodes is directly performed by JiST (a more detailled description of JIST as well as a performance comparisons with other platforms can be found in [8]). The MAC of 802.16 comprises three sublayers. The Service-Specific Convergence Sublayer (CS), the MAC Common Part Sublayer (MAC-CPS), and the Security Sublayer. The CS provides any transformation or mapping of external network received through the CS service access point (CS-SAP) into MAC service data units (MAC-SDU). The CPS processes these MAC-SDUs and performs the sizing of the SDUs to fit the PHY symbols (e.g., 96 Bytes per symbol in one-half QAM64 modulation). Moreover, the MAC-CPS provides the core MAC functionality of system access, bandwidth allocation, connection establishment, and connection maintenance. For our analysis we implemented only the necessary parts of the MAC of IEEE 802.16-2004 [1], i.e., parts of the Common Part Sublayer. In particular, we include the message handling of network control messages (NCFG) and coordinated centralized scheduling messages (MSH-CSCH). To simplify the simulation setup, we store basic configuration settings such as information about NCFG messages or the scheduling tree in central configuration files. As a result, simulations are static in the sense that it is not possible to make changes, e.g., addition or removal of nodes or choice of a different modulation scheme, during simulation time. It is obvious, that this simplification does not influence the accuracy of simulation results for our setup. The reservation of bandwidth is performed by means of the scheduling messages. MSH-CSCH messages have to be created to request/grant bandwidth. The exact reservation mechanism, i.e., the size of the reservation request formed by the SSs during the request phase, is left unspecified in the standard. Thus, we implemented a basic reservation mechanism that is based on the amount of queued data on the SS. In particular, our reservation mechanism estimates the number of frames necessary to transmit all queued packets. Since the allocation of requested bandwidth for the next schedule is performed during the lifetime of the existing grant, minor overbookings are possible using this scheme. While the standard allows for a decentralized computation of the corresponding schedule, we perform all related calculations at the Base Station object to improve simulation performance. Per Subscriber Station (SS) we introduce an object that holds timing information of the uplink transmission requests to its parent node as well as bandwidth information on granted bandwidth per SS. We introduce physical layer simplifications with respect to the propagation delay of individual frames. Overall, we estimate the cumulative error introduced with our model simplifications to be below 4 milliseconds.

B. Simulation Results

The experimental design of our simulation study reflects the setup of the theoretical analysis as far as possible (see Section III). Our simulation scenario comprises seven networks with differing node sizes. The smallest one is of size 7 nodes (each node within a one-hop distance from the BS) while the largest one is of size 121 nodes and has an average hop-distance of 5.16 to the BS. Because of space limitations we here only discuss the results that can be directly compared with the theoretical findings of Section III, however. The parameterization of our simulations is as follows: MMDS with 24 MHz, so we have 252 symbols per frame (Tg=Tb/32). In all of the simulated networks, the SS generates packets with a size of 1500 Byte destined to the BS in intervals of 500ms. The simulation duration has been set to 30 seconds. The packets arrive at the source node with a jitter of up to 3ms. The length of a frame is fixed to 2.5ms and the CSForwardingDelay is set to 3 symbols. We step through all feasible parameters for MSH-CTRL-LEN and perform 60 independent packet transmissions equaling 60 independent request/grant cycles to reserve bandwidth. For all simulations, we show the average as well as the worst-case and best-case results. Figure 5 shows the obtained results for all simulations, while Figures 6, 7, and 8 give a direct comparison between simulation results and analytical worst-case and best-case results for the individual setups. In Figure 5, the dependency between the number of nodes in the network and the scheduling delay is clearly visible. The higher the number of nodes, the higher the scheduling delay. Also, the decrease in delay with increase in MSH-CTRL-LEN is notable. As mentioned earlier, the MSH-CTRL-LEN variable controls the amount of bandwidth available for MSH-CSCH messages, which convey the request/grant messages. This directly translates into an increase in delay if a high number of nodes compete for bandwidth, because the MSH-CTRL-LEN value allows only for a limited amount of requests to be transmitted. With a higher value of MSH-CTRL-LEN, the messages are propagated faster along the tree thus, resulting in decreasing delays. Lower values of MSH-CTRL-LEN have the opposite effect. The acceleration, however, does not come without additional cost: with higher values of MSH-CTRL-LEN, the resources used for scheduling messages increase, thus, the bandwidth available for the transmission of data packets decreases. In figures 6, 7, and 8, we directly compare our theoretical results with the simulations. The figures show that our simulation results are within the envelope given by the analytical worst-case and best-case results. Minor violations of this envelope are due to the aforementioned simplifications of our simulation model. We, however, rarely approach the theoretical best-case results in the setups representing 26 and 49 nodes. These pessimistic results are due to our choice of the source nodes among the SSs: we only selected SS at the edge of the distribution tree in our simulation study, while the best-case results of the theoretical analysis also include SSs that are one-hop neigbors of the BS (see Fig. 2). Since our focus lies on the real-time capabilities of 802.16, we next compare the individual results with the delay requirements of VoIP traffic. A good starting point for this comparison is the Recommendation G.114 [9] from the International Telecommunication Union (ITU). This recommendation defines the following three bands of delay. 0-150ms: Acceptable for most user applications. 150-400ms: Acceptable in some situations. Above 400ms: Unacceptable. The values are one-way delay values and assume that echo cancellation is used. Echo cancellers are required according to ITU G.131 [10] when the one-way delay exceeds 25 ms. Our simulation setup reflects the real-time requirements as follows: the scheduling delays shown in figures 5, 6, 7, and 8 assumes that a packet arrives at the SS and no bandwidth has been reserved before. This can be translated in a setup, where no additional QoS mechanisms are used and per-packet reservations are used for voice traffic. If we compare our results with the VoIP delay requirements, the 49 node scenario is close to the limits for "acceptable" voice quality for certain MSH-CTRL-LEN settings. In larger networks, the obtained values are well beyond the acceptable limits for VoIP traffic for inappropriate settings of the MSH-CTRL-LEN parameter. To summarize, the simulations as well as the analytical results show a significant effect of MSH-CTRL-LEN settings on the transmission delay. In networks beyond 40 nodes, some MSH-CTRL-LEN settings are already critical in a sense that VoIP delay requirements can not be met. Is has to be emphasized that these results have been gained far below the capacity limits of the network. Additional data traffic further has a further impact on the delay and mandates for the use of QoS mechanisms to prioritize delay-sensitive traffic. Setting the MSH-CTRL-LEN parameter, however, has also to be optimized with regard to communication overhead. Let us consider the 26 node setting depicted in figure 7 as an example. It can be seen that settings from 8 to 15 show no effects on communication delay. However, a setting of MSH-CTRL-LEN=15 results in an increase of the scheduling control subframe size of 56 symbols compared with a setting of MSH-CTRL-LEN=8. Thus, for this particular setup, the available bandwidth is reduced by 56/252=22,2 percent.

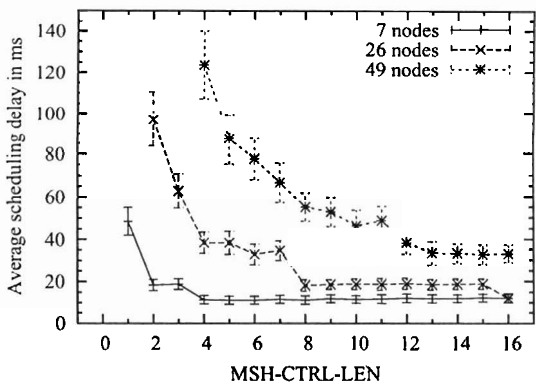

Fig. 5. Simulated scheduling delay vs. MSH-CRTL-LEN for all setups.

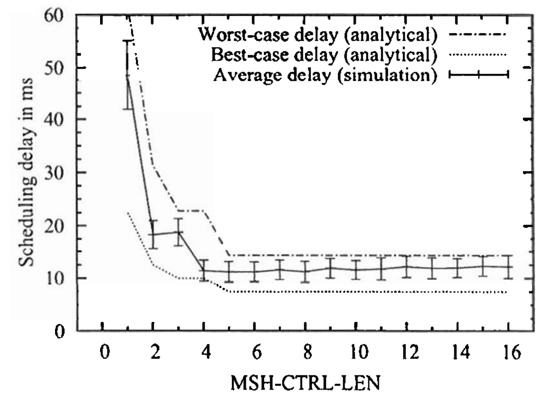

Fig. 6. Simulation vs. analysis of scheduling delay for the 7 node setup.

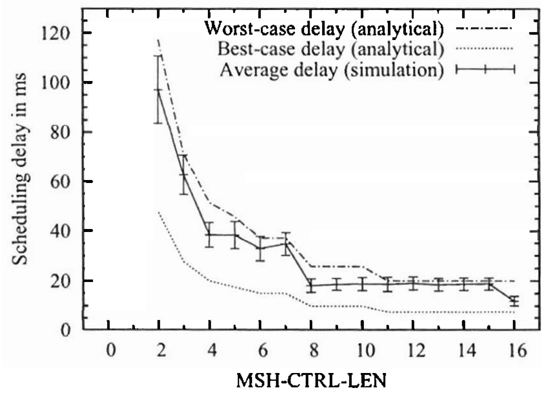

Fig. 7. Simulation vs. analysis of scheduling delay for the 26 node setup.

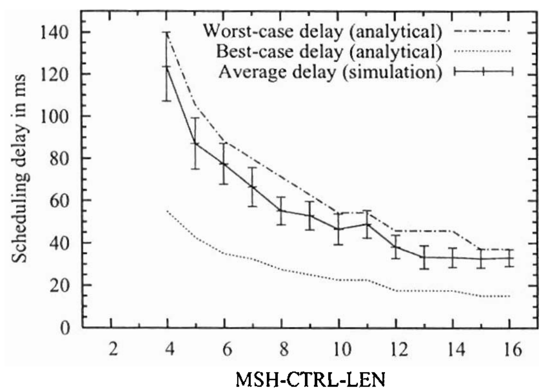

Fig. 8. Simulation vs. analysis of scheduling delay for the 49 node setup

V. Conclusion

The promise of wireless broadband access networks is to reduce costs, while providing adequate QoS to allow for the huge range of services currently supported in wireline access networks.

The QoS of real-time multimedia applications present a yardstick to measure the performance of upcoming wireless technologies such as the IEEE 802.16-2004 standard. For our investigation, we have chosen the transmission scheduling mechanisms of the 802.16 standard. In particular, we have carried out a theoretical performance analysis as well as a simulation study to investigate the coordinated centralized scheduling mechanism of the 802.16-2004 standard. Our results include theoretical worst-case, best-case, as well as simulative average-case results showing the trade-off between scalability and achievable performance. This trade-off can be optimized using the MSH-CTRL-LEN parameter, which controls the partitioning of the frames into control subframe and data subframe.

References

[1] IEEE Computer Society and IEEE Microwave Theory and Techniques Society, "802.16 IEEE Standard for Local and metropolitan area networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems," IEEE Std. 802.16-2004, October 2004.

[2] S. Redana and M. Lott, "Performance Analysis of IEEE 802.16a in Mesh Mode," in *Proc. of IST SUMMIT 2004, Lyon, France*, June 2004.

[3] C. Eklund, R. Marks, K. Stanwood, and S. Wang, "IEEE standard 802.16: a technical overview of the WirelessMAN air interface for broadband wireless access," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 98–107, June 2002.

[4] K. Wongthavarawat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *International Journal of Communication Systems*, vol. 16, pp. 81–96, 2003.

[5] H. Lee, T. Kwon, and D. Cho, "An Efficient Uplink Scheduling Algorithm for VoIP Services in IEEE 802.16 BWA Systems," in *IEEE VTC2004-Fall*, vol. 5, Korea Advanced Institute of Science and Technology (KAIST). IEEE, September 2004, pp. 3070–3074.

[6] M. Cao, W. Ma, Q. Zhang, X. Wang, and W. Zhu, "Modelling and performance analysis of the distributed scheduler in IEEE 802.16 mesh mode," in *MobiHoc '05*, 2005, pp. 78–89.

[7] JiST/SWANS: Java in Simulation Time/Scalable Wireless Ad hoc Network Simulator. [Online]. Available: http://jist.ece.cornell.edu/

[8] (2006) JiST- Java in Simulation Time User Guide. [Online]. Available: http://jist.ece.cornell.edu/jist-user/index.html

[9] (2003) Recommendation G.114 - One-way Transmission Time. [Online]. Available: http://www.itu.int/itudoc/itu-t/aap/sg12aap/history/g.114/index.html

[10] (2003) Recommendation G.131 (11/03) Talker echo and its control. [Online]. Available: http://www.itu.int/rec/T-REC-G.131/en