Novel Techniques for Improving the Throughput for the 802.11 MAC Sub-layer in Wireless Networks

M P Morrison, G A D Punchihewa, L De Silva, F M Al-Ali

Institute of Information Sciences & Technology
Massey University, Palmerston North, New Zealand
Emails: mpmorrison@iee.org, g.a.punchihewa@massey.ac.nz,
l.desilva@massey.ac.nz, f.m.al-ali@massey.ac.nz

Abstract: This paper proposes three novel improvements to the 802.11 MAC sub-layer that increase the throughput of 802.11 wireless networks. These improvements increase the efficiency of the MAC sub-layer. An overview of 802.11 standards is given (including 802.11a, 802.11b and 802.11g), with a synopsis of recent 802.11e, and 802.11i, and future 802.11n, extensions to 802.11. Justifications for improving the MAC sub-layer are presented, and then the improvements are detailed afterwards. Firstly, packet aggregation is described as a system that improves efficiency – particularly with small packets – along with simulation results, operational details and implementation issues. A stored frame header system is also introduced as a method specially suited to reducing overheads for large packets. Lastly, combined acknowledgments are presented as an improvement to the 802.11 acknowledgment scheme.

Keywords: aggregation, combined acknowledgment, MAC sub-layer, stored frame header, throughput.

1 INTRODUCTION

Both commercially and privately, 802.11 wireless networks have now become well known, and are capturing an increasingly larger portion of both the business and home network markets. The current growth in this sector is expected to continue for many years to come. Together with this growth comes the demand for higher data throughputs to cater for more demanding traffic loads.

Traditionally, the increase in throughput in wireless networks has been realized through increasing the physical layer data rate. The original 2 Mbps of 802.11 [1] of 1997 was extended to 11 Mbps with 802.11b [2], and 54 Mbps in the more expensive 802.11a [3], both in 1999. Then in 2003, 802.11g [4] combined features of 11a and 11b to provide a more affordable 54 Mbps alternative.

In the same timeframe (1997 to 2003), no major changes to the original 802.11 MAC sub-layer were ratified. Later extensions 802.11e and 802.11i have added functionality, but did not change the underlying operation. 802.11e added quality of service (QoS) that gives prioritisation for essential traffic, and changes the coordination functions to add these. However, these changes have decreased either the throughput, or the stability and fairness, of the MAC sub-layer [5]. 802.11i simply improves the security of wireless LANs, with negligible, if any, effect on throughput.

The next major milestone for the 802.11 family of standards is 802.11n – entitled "High Throughput Extension to 802.11", and targeted for ratification in

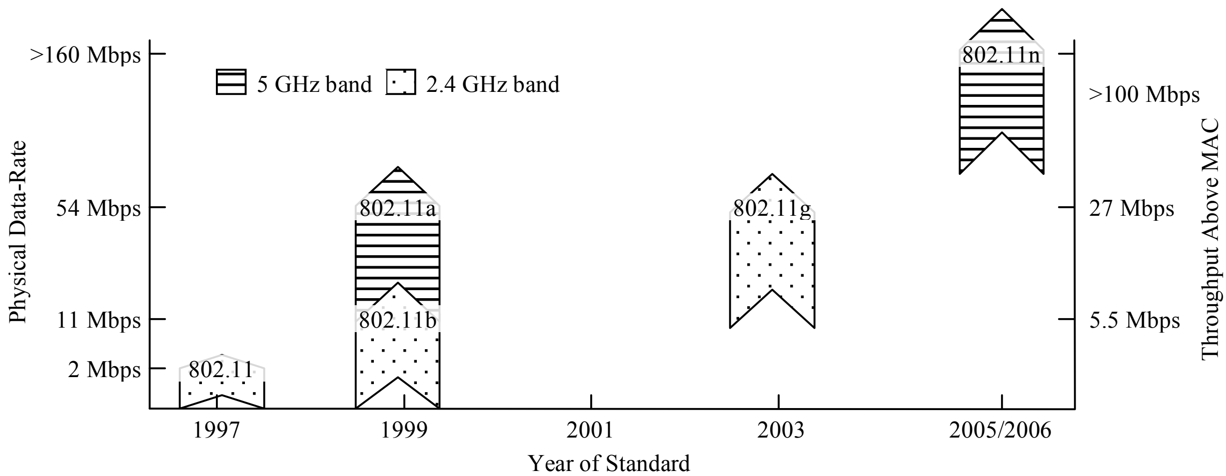

Figure 1: Comparison of different 802.11 standards

early 2006. 802.11n will extend the physical layer, as well as completely overhauling the MAC sub-layer. The main aim of 802.11n is to provide a maximum throughput – above the MAC sub-layer – of at least 100 Mbps [6]. The physical layer is intended to have a maximum speed of well over 200 Mbps. The authors have compiled a comparison of the different 802.11 standards in Figure 1.

2 REALISTIC NUMBERS

2.1 Realistic Throughput

The 54 Mbps 'maximum data rate' of 802.11a/11g is the maximum possible speed that the data in a frame is transmitted. However, lower rates are used for both the physical and MAC sub-layer headers, and the control frames used to synchronize all stations. Therefore, the maximum data rate is only used for the actual packet data contained within the frame, with relatively long periods of slower rate headers and synchronization.

This means that the actual data throughput for an 802.11 wireless network is much lower than the maximum data transmission rate, as the maximum data rates are only used for a fraction of the time. The maximum data rate for 802.11b is approximately 5.5 Mbps, and 26 Mbps for 802.11a and 802.11g.

Figure 2 shows the transmission of a TCP Ack packet over 802.11b (the backoff period has been left out for simplicity). Only the MAC frame data (the TCP Ack packet) is sent at the maximum 11 Mbps, while everything else uses the lowest speed – 1 Mbps – for backwards compatibility and decreased susceptibility to errors. In this example, the maximum data rate is only used about 10% of the time.

2.2 Realistic Packet Sizes

Additionally, these maximum throughput speeds are only obtainable if the packet contained within the frame are close to, or equal to, the maximum allowable packet size (usually 1500 bytes). Realistically however, a large proportion of frames have a much smaller size.

In a study of a university campus wireless network, Baker and Tang examined over 70 million frames and found that over 70% of packets were smaller than 200 bytes [7]. (Figure 3 shows the cumulative packet distribution from the study.) With the lower average packet size, the throughput drops accordingly, as the per-packet overhead percentage increases.

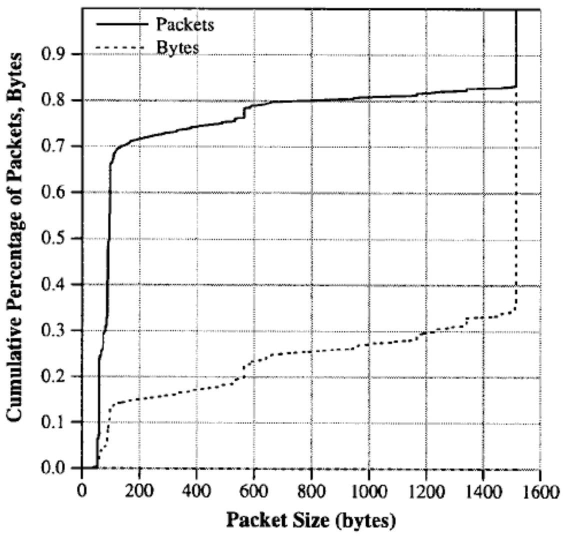

Figure 3. Packet Distribution from Baker & Tang [7]

2.3 MAC sub-layer Efficiency

Because of the fact that maximum speeds are used for only part of each transmission, and that a large proportion of frames are relatively small, the MAC sub-layer efficiency is very important. The authors believe that the current MAC sub-layer is too inefficient – especially for the smaller packets. Additionally, it has been found that as the physical layer speed increases, the MAC sub-layer efficiency decreases – even with 1500 byte packets the 802.11e MAC sub-layer is always less than 64% efficient with physical layer speeds higher than 140 Mbps [8].

With packets smaller than 1500 bytes, the efficiency will be even lower. By reducing the per-packet overheads, the efficiency of the MAC sub-layer can be improved. Improving the coordination functions and medium access operation could achieve this, but these have been the only part of the MAC sub-layer so far to be improved. Also, improvements in these parts of the MAC sub-layer would probably lose functionality such as QoS and security. Therefore, our research has targeted improving the per-packet efficiency by modifying the MAC sub-layer frame system.

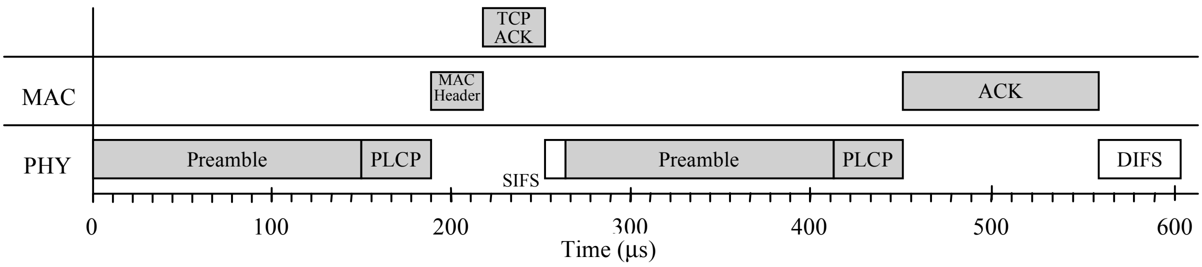

Figure 2. TCP ACK Transmission showing Physical and MAC sub-layer Overheads

3 PACKET AGGREGATION

3.1 *Aggregating Several Packets*

The most significant development to the 802.11 MAC sub-layer developed by our research is aggregating several packets into a single frame. Simulation results prove that this improves the MAC sub-layer efficiency by increasing the average size of the frame payload. It also means that several packets can be sent for the cost of only a single MAC sub-layer header, one coordination function round to gain access to the medium, and a single physical header. This lowers the overhead proportion and increases the percentage time the maximum physical layer transmission speed is used.

3.2 *Aggregation Types*

Aggregation can be implemented in two forms – forced delay aggregation (FDA) and congestion triggered aggregation (CTA) – identified by Jain et al. [9]. Figure 4 illustrates the differences between the two forms of aggregation.

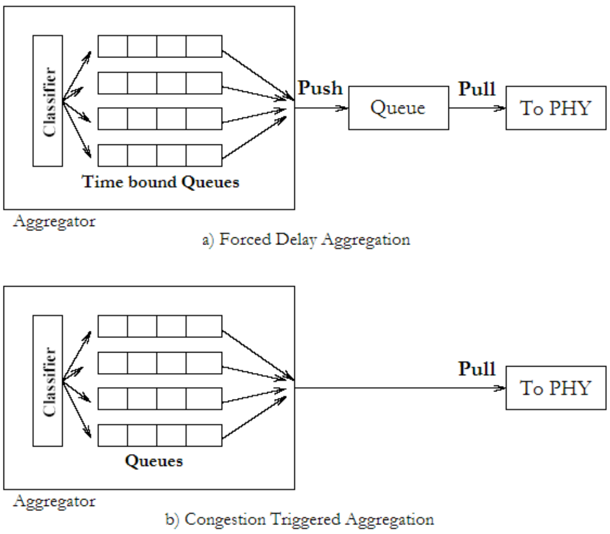

Figure 4. Aggregation Types [9]

A packet entering an FDA MAC sub-layer will be put into a time bounded queue according to some criterion, and pushed (within an aggregated packet) to the transmitter when the packet at the head of the queue has undergone a set delay. This type of aggregation is easier to implement, and is intended to introduce a high level of aggregation, but increases the latency for almost all packets.

CTA, however, uses the same queuing criterion, but the transmitter 'pulls' an aggregated packet when it can transmit. In this way, aggregation only occurs when the network becomes congested, but no extra latency occurs as with FDA. Studies have shown that CTA is just as effective as FDA in improving throughput through aggregating packets, and with no extra latency compared with existing 802.11 [9].

3.3 *Queuing Control*

With CTA, the decision on which queue to take is based solely on the timestamp of the packets at the head of the queues. With FDA, however, a decision needs to be made on which queue the packets are pulled from for transmission. For our research, we compared six queuing decision algorithms: round robin; priority based; weighted fair queuing (combination of priority and round-robin); load based; FIFO (first in, first out); and maximise aggregation (in term of packets aggregated). These simulations showed that the best overall queuing decisions used a load based algorithm, although in some traffic conditions FIFO performed better.

The simulations used the same packet distribution as in Figure 3, and were run using 802.11 DCF mode. They were carried out using MATLAB, with a custom-made simulation script, as the researcher could not find an existing simulation tool that could handle variable frame sizes. The existing 802.11 MAC sub-layer was included for comparison purposes, and an additional "sub-header" of 6 bytes per packet was included to simulate the additional headers needed for aggregation.

3.4 *Simulation Results*

The simulations of aggregation have been carried out using both 802.11a and 802.11b physical layer speeds. Figure 5 shows the results of the simulations for both physical layer speeds, using different numbers of stations active. All stations kept a similar level of traffic – about 6 Mbps per station for 802.11, and about 1.2 Mbps per station for 802.11b.

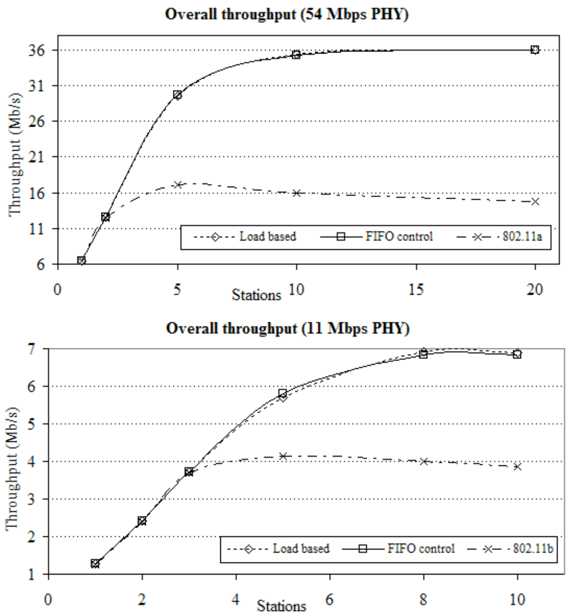

Figure 5. Results for 802.11a and 802.11b Physical Layers

The results show the large increase in throughput from the use of aggregation. With the 802.11a physical layer, the throughput was more than doubled for the higher traffic loads. As the traffic load increases, the existing 802.11 throughput decreases, due to the increased contention for the medium. However, the aggregation throughput keeps increasing, as the mechanism is congestion triggered.

The throughputs of about 36 Mbps (802.11a) and almost 7 Mbps (802.11b) show how aggregation can achieve throughputs significantly higher than the current maximums. The simulations also show how the throughput of existing 802.11 levels off much lower – about 4 Mbps and 16 Mbps for 802.11b and 802.11a respectively – for smaller packet sizes.

3.5 Operational Details

Each packet aggregated into the frame payload requires its own control information. We have allocated 6 bytes per packet for this purpose as follows (see Figure 5). Each sub-header contains a 16 bit CRC packet check sequence (PCS), a 2 byte sequence control field and a length field. The PCS field allows each individual packet to be checked for errors. If only one packet out of several aggregated has an error, only that single packet needs to be re-transmitted.

2	2	2	
PCS	Sequence Control	Length	Packet

Figure 6. Proposed Aggregation Sub-Header

Each packet sent over an existing 802.11 wireless network has a sequence control field of 2 bytes. With aggregation, we have decided to keep the sequence control field for each packet. All operations that use the existing sequence control are therefore unaffected. Finally the length field only needs 12 out of the 16 bits left to represent all packet lengths (maximum allowable is 2132), with the last 4 bits reserved for future use.

Also, a subtype will be added to the Data type. This subtype will be called “Data (aggregated)”, and will take up one of the reserved Data subtype values. It is possible that future extensions to aggregation would also require separate subtypes – such as a combined acknowledgement, or selective repeat scheme.

3.6 Implementation Issues

As part of our on-going research, we will be implementing aggregation using a real networking environment, with computers running the Linux operating system. This implementation will utilise the NetFilter service built into the Linux network stack. This allows the user to control packet being sent across any network device. In our case, an aggregator program will intercept all packets being directed across the wireless network, and use CTA with load based queuing. We also aim to include optional priority queuing to enable the support of QoS services.

The main issue in this implementation will be the ‘pull’ mechanism of CTA. Traditionally packets are ‘pushed’ one by one over the wireless network. However, our proposed aggregation requires several packets to be ‘pulled’ at a time. The main problem is that a notification needs to be sent to the aggregator - the ‘pull’ request from the transmitter.

4 STORED FRAME HEADERS

4.1 Repetitive Transfers

As a further improvement to the MAC sub-layer operation, our research sought to utilise the fact that several frames may be sent between two stations that share very similar headers in each direction. For instance, a file transfer involves splitting the file into several small packets, which are sent one by one to be reassembled at the receiver. In this transfer, the same header will be used for each packet apart from the sequence number. If the header is remembered at the receiver from the first packet, it does not need to be resent.

Although the header is only a small fraction of the payload size, it is sent at a much slower rate than the actual data. This means reducing the header size can have a significant effect on throughput. Storing the frame header for a transferring a large file will have a significant affect, as the header size will be reduced a large number of times.

4.2 Complexity

The main issue with this scheme is complexity. The simplest form of storing frame headers is to keep only a single header. This is simple to implement and can be controlled with a single bit in the existing control fields. However, if more headers are stored, greater improvements may be possible, but may need a far more complicated control scheme, with the possibility of extra header fields and handshakes. At some point the extra overhead needed to support this scheme will outweigh the scheme’s benefits.

The implementation of this scheme will need to consider what level of complexity the system can be used for the best relative throughput improvement. For example, if N stations are all within reception of each other, should each station store one frame header per station? Obviously this scheme will not be suitable for large numbers of stations. Therefore an appropriate control scheme needs to be implemented to decide which frame headers are stored. However, if the main overheads can be confined simply to memory space, the current trends in memory size and cost means that storage should not be too much of a problem.

4.3 Benefits of Stored Headers

As part of the 802.11n standard extension, certain types of data transfer are designated as requiring a certain level of performance. For example, streaming video and audio requires low delay and strict ordering. To achieve this, 802.11n defines modes of operation that can deliver the required performance for several likely usage scenarios. It is proposed that some of these usage scenarios – such as streaming video – would benefit greatly from the repeated header scheme as they involve repeatedly sending similar data to a single receiver.

The common home applications that would benefit from this scheme are file transfers and streaming multimedia and games. Also, backhaul wireless links that uses 802.11 would benefit, as the number of receivers will be minimal for the static links of a backhaul link.

4.4 Operational Details

The stored frame header would still need to transmit the 2-byte frame control fields of the header. A new data subtypes "Stored Header" would be designated. Other than the frame control, only the sequence control field is needed, giving a header size of only 4 bytes (see Figure 7a), compared with 24 for the full header! If the packet is a different length but otherwise meets the criteria for a stored header, a different subtype would be used, and the duration field also included for a 6-byte header (Figure 7b).

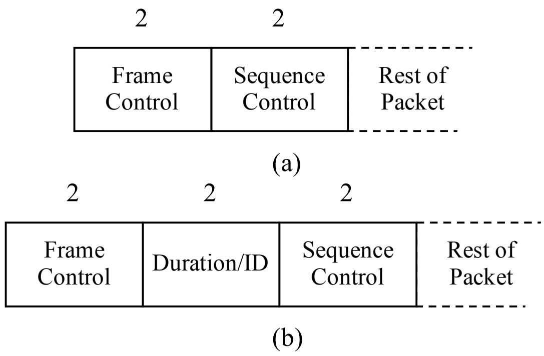

Figure 7. Reduced Header Sizes of Stored Header System

5 COMBINED ACKNOWLEDGMENTS

5.1 802.11e Block Acknowledgment

Currently, the default operation of 802.11 is to acknowledge every frame received. This requires an entire transaction, including reserving the wireless medium, and no data transmission can occur in this period. While the 802.11e extension added a Block Acknowledgment control frame, it is optional, requires extra negotiation, and requires the network to be operating in the contention free HCF mode. This means that it cannot be used for ad-hoc networks, or with station without 802.11e capability.

However, the block acknowledgement scheme is very useful, as it lowers the number of frames sent that take up the medium without sending actual data. Therefore it will be beneficial to have a similar scheme available in all 802.11 wireless networks.

5.2 Combined Acknowledgements

The authors propose a 'combined' acknowledgement system. The same acknowledgement frame header as existing 802.11 will still be used, but will be followed by the sequence control fields of each frame that is being acknowledged. As the sequence number of each frame is independent of whether the frame was aggregated, was sent alone, or what order it was sent, it can be used to uniquely identify the frame.

This system will not require the station to be operating in any specific control mode, and could be used for both aggregated packets and packets sent individually. Block acknowledgements work over a certain time duration in which multiple frames are sent. A similar system could be used where a combined acknowledgement is only sent either after every N frames or after every N ms, which ever occurs first. This would be defined as part of the local network settings, and would be consistent across all stations in the network.

5.3 Extensions

A possible extension to the acknowledgement scheme would be to add some form of link quality measurement. This could be implemented directly by adding a link quality field to the acknowledgement header, or indirectly by assessing the proportion of successful transmission over a specified time period.

This technique would be very useful for frames that failed to transmit – as the transmitter of these frames could use information about the link quality in a control scheme e.g. a payload size backoff scheme for aggregation, to improve the chances of correct reception when it the frames are resent.

6 CONCLUSIONS

In conclusion, these three improvements will considerably increase the throughput for 802.11 wireless networks. These improvements could definitely be incorporated into the 802.11n standard as part of the overall goal of providing very high throughputs in excess of 100 Mbps.

Aggregation provides a very significant increase in throughput to 802.11 wireless networks. This is achieved by increasing the average packet size, which in turn lowers the proportion of overheads per-packet. Combined acknowledgements and stored headers will also provide throughput improvements, but not on the same scale as aggregation. However, they can be implemented with little or no increase in overheads and system complexity.

The authors consider that some or all of these three improvements (or parts thereof) could be used as part of future 802.11 standards to improvement the performance of wireless networks. With some development, the improvements may possibly be deployed using existing 802.11 wireless networking hardware.

7 FUTURE WORK

There are two main sections of future work for this project. The first is the implementation of aggregation under Linux, using 802.11b hardware. This will prove that aggregation can be used in real-world networks, and will provide experimental result on the effect of aggregations on throughput.

Second is the simulation of the stored frame header scheme and combined acknowledgements scheme. This requires some significant developments to the simulation script – including the different traffic characteristics per station needed to assess the result of the stored header system, and the development of a stored header control scheme.

Also, the benefits of the stored frame header system may possibly be applied to the physical layer control systems. One possibility is that information may be repeated in the preamble and PLCP during each 'transaction' i.e. data frame plus a return acknowledgement. By storing this information, the return time of the acknowledgement frame could be markedly reduced. Similarly, aggregation could be implemented at the physical layer, with multiple frames included in a single transmission to multiple receivers. All stations addressed would receive the frame, and just process the one frame that was addressed to them. In this way, multiple rates over a single transmission could also be possible – solving backwards compatibility.

8 REFERENCES

1 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. IEEE Std 802.11, 1999.

2 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band. IEE Std 802.11b-1999, 1999.

3 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer Extension in the 5 GHz Band. IEE Std 802.11a-1999, 1999.

4 Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band. IEE Std 802.11g-2003, 1999.

5 P. Garg, R. Doshi, R. Greene, M. Baker, M. Malek and X. Cheng. "Using IEEE 802.11e MAC for QoS over Wireless". Proceedings of the IEEE Performance, Computing and Communications Conference, Phoenix, Arizona, USA, pp. 537-542, 2003.

6 J. Rosdahl. "Project Authorisation Request for High Throughput Study Group". IEEE Working Group Paper 802.11-02/0798r7, March 2003.

7 M. Baker and D. Tang. "Analysis of a Local Area Wireless Network". Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, Atlanta, Georgia, USA, pp. 107-118, September 2002.

8 A. R. Dias, S. Rouquette-Leveil and S. Simeons. "Multiple Antenna OFDM Solutions for Enhanced PHY". IEEE Working Group Paper 802.11-04/0229r0, March 2004.

9 A. Jain, M. Gruteser, M. Neufeld and D. Grunwald. "Benefits of Packet Aggregation in Ad-Hoc Wireless Networks". Technical Report CU-CS-960-03, Department of Computer Science, University of Colorado, Boulder, Colorado, USA. August 2003.