Boosting Video Capacity of IEEE 802.11n through Multiple Receiver Frame Aggregation

Kyungsso Lee, Sangki Yun and Hyogon Kim
Korea University

Abstract—The technology based on the IEEE 802.11 standard has been hugely successful, and is evolving towards even higher speeds and richer features. In particular, the 802.11n amendment of the standard aims to achieve the physical layer (PHY) rate of 600Mbps. Although the 802.11n offers sufficient bandwidth to support high-resolution video applications such as High Definition TV (HDTV), we find that the number of video streams that can be supported on the IEEE 802.11n networks depends heavily on how the frame aggregation is implemented. In addition to the frame aggregation scheme stipulated in the amendment, we explore a multiple-receiver frame aggregation scheme for video traffic. The comparative study reveals through extensive simulation that the proposed multiple-receiver aggregation scheme increases the number of supported video streams by a factor of 2 or higher. We also shed light on the qualitative difference in the dynamics of the two approaches. Whereas the aggregation efficiency worsens with traffic increase in the point-to-point aggregation (which is highly undesirable), the proposed multiple-receiver aggregation exhibits resiliency against congestion, by matching the aggregation efficiency to the traffic load.

I. INTRODUCTION

The IEEE 802.11 technology has been hugely successful in every aspect, and is evolving towards even higher speeds and richer features. In particular, the 802.11n amendment of the standard aims to achieve the physical layer (PHY) rate of 600Mbps by using MIMO and OFDM, and by combining multiple 20MHz channels. As of 2007, there are initial products in the market that claim to exceed 100Mbps.

Although the 802.11n offers sufficient bandwidth to support high-resolution video applications such as High Definition TV (HDTV), it remains to be seen how efficiently video streams can be carried over the 802.11 networks. Indeed, we find in our investigation that the number of video streams that can be supported on the IEEE 802.11n networks depends heavily on the manner the newly added frame aggregation feature of the 802.11n amendment is implemented.

As successive 802.11 standards stipulate higher speeds, the relatively slow control part of the transmission such as PHY preamble and header, MAC-layer acknowledgements, and contention time renders the efficiency of a single MPDU transmission increasingly lower. For this reason, the newly emerging IEEE 802.11n amendment includes frame aggregation as a standard feature to address the problem. It allows the aggregated MPDU (A-MPDU) size to be up to 65,535B (Draft 2.0 [1], 7.4a.1)\(^1\), which is 8 times the size of a single MPDU, thereby reducing the relative overhead.

\(1\)Compared with another format, Aggregated MSDU (A-MSDU), it is known to be more resilient to partial corruption, a highly desirable feature for aggregated frames.

An issue arises from the fact that there can be many ways to exploit the aggregated MPDU format to carry video frames. Our concern on the current draft of the 11n amendment is that the intended usage of the 802.11n A-MPDU is overly restrictive. Namely, the usage is effectively limited to unicast. Draft 2.0 stipulates, “All the MPDUs within an A-MPDU are addressed to the same receiver address” (\([\text{1, 7.4a.4}]\)). We believe that the restriction should be lifted for more flexible uses of this powerful optimization feature, our video frame aggregation case being one example.

Thus in this paper we stretch the A-MPDU format so that it can carry MPDUs addressed to different destinations. This is similar to the ‘TGn Sync Multiple Receiver Aggregate Multi-Poll (MMP) MPDU format [2]. But the difference is in how the aggregated frames are put into the A-MPDU. Instead of grouping and separately packaging the frames according to the destination (i.e., the “Unicast Group” in TGn Sync), we simply put the video frames (encapsulated as separate MPDUs) in the queue into the A-MPDU as such, limited only by the maximum A-MPDU size. There is no sorting and separate packaging. Albeit a small difference, it can make the aggregation much easier. When the frames are widely spaced in time (e.g. VoIP and video streaming), the per-destination framing complexity and encapsulation overhead can be avoided. And there is no need to wait for additional arrivals of the frames from the same flow.

The focus of this paper, however, is not in performance difference, if any, caused by the aggregate format. Rather, it focuses on the exploration of the impact of multiple-receiver aggregate on the video carrying performance of the 802.11n link. In particular, we will show that the proposed scheme boosts the HDTV carrying capacity of the 802.11n link to more than twice the point-to-point\(^2\) aggregation stipulated in the current 11n draft. More importantly, we reveal a significant qualitative difference between the single-receiver and multiple-receiver aggregation approaches. Whereas the aggregation efficiency worsens with traffic increase in the point-to-point aggregation (which is highly undesirable), the proposed multiple-receiver aggregation exhibits resiliency against congestion, by matching the aggregation efficiency to the traffic load.

The rest of the paper is organized as follows. Section II describes the proposed scheme, the Instantaneous Multi-receiver Aggregation (IMA). The concept of congestion-triggered aggregation is introduced, and the frame format

\(2\)We use the terms “point-to-point aggregation” to mean “single-receiver aggregation”. Also, we use “point-to-multipoint aggregation” and “multiple-receiver aggregation” interchangeably in this paper.
Section V concludes the paper. Section III presents the performance comparison of the IMA and the unmodified 802.11n, in terms of throughput, delay, loss, and jitter. Section IV comparatively discusses related work. Finally, Section V concludes the paper.

II. INSTANTANEOUS MULTIPLE-RECEIVER FRAME AGGREGATION

The Instantaneous Multi-receiver Aggregation (IMA) scheme has two components: congestion-triggered frame aggregation and multiple-receiver transmission. The former addresses when and how many video frames are put into a single A-MPDU, whereas the latter, broadcast delivery of the aggregate frame to multiple destinations.

A. Congestion-triggered frame aggregation

Most existing frame aggregation works assume that we wait a fixed time duration to gather more frames, thus improving the aggregation efficiency further. However, we have the design philosophy that frame aggregation should be used parsimoniously. The reason is twofold. First, for some real-time applications such as video conferencing, it can degrade the perceived quality. Second, it is hard to justify when there is no shortage of bandwidth (i.e., no congestion). Therefore, we choose to perform the aggregation only when there is congestion. A natural choice would be to aggregate whatever video frames there are, queued at MAC, at the time of acquiring the transmission opportunity (TXOP). This is the central idea of the congestion-triggered frame aggregation (that was used in our previous work on VoIP aggregation [3]), which was also independently conceived in [4]. Since only the natural queueing delay is exploited in the aggregation, IMA incurs no additional aggregation delay.

B. Multi-receiver frame exchange

The difference from the congestion-triggered aggregation scheme from our previous work [3] is that it is applied across video streams in this paper. Namely, we perform inter-stream aggregation as opposed to intra-stream aggregation, as used in [1], [4]. So we rely on the frame format and the frame exchange sequence that are similar to those of the MMP in the TGn Sync proposal [2]. Fig. 1 shows how an aggregated frame is transmitted and acknowledged. In this example there are 3 neighbors to receive the video frame aggregated in the A-MPDU: a, b, and c. The MAC frame is broadcast, and each destination station responds with a Block ACK (BLACK). Between each BLACK, there is a SIFS. We will assume that the order of BLACK transmissions by the receivers is determined by the order they appear in the aggregated frame. And upon the transmission of the aggregated MAC frame, the transmitter reserves the channel for \( n \) ACKs and SIFSs using a long NAV, where \( n \) is the number of neighbors that are intended receivers of the video frame(s).

In case some ACKs are missing, the sender retransmits the MPDUs corresponding to those ACKs. We could either re-aggregate the retransmitted frames with the subsequent frames in the queue, or retransmit them in a smaller A-MPDU separately. In this paper, we choose the latter approach, under the assumption that the aggregation for the subsequent frames in the back of the queue is done simultaneously with the transmission of an A-MPDU in the front. In such situation, the re-aggregation would trigger the chain reaction, i.e., reassigning frames over a new set of A-MPDUs, which should be costly. On the other hand, the re-aggregation approach might have merit in terms of the transmission efficiency, but we defer the investigation of its potential to a future work.

C. 802.11n frame aggregation

The 802.11n draft does not specify exactly how frame aggregation should be done. But according to [1], 7.4a.4, we know at least that the aggregation is usable only for point-to-point communication. Namely, the aggregation is done over the frames in the same stream (going to the same destination station). There could be more than one way to implement even this point-to-point frame aggregation, but in this paper we assume for both the 802.11n aggregation and IMA that the aggregation algorithm does neither collect non-contiguous but concurrently queued frames from the same video stream nor wait for additional delay to achieve higher aggregation efficiency. In other words, the 802.11n frame aggregation can aggregate only the back-to-back queued video frames that belong to the same video stream, but not those intervened by the frames destined to other receivers. This assumption may be relaxed, but at higher queue management complexity, it will still be less efficient than inter-stream aggregation of IMA. The latter assumption, by the way, is for fair comparison with IMA, which does not use any additional delay.

III. PERFORMANCE EVALUATION

A. Experiment settings

In our simulation, we use the 802.11 module in the Qualnet simulator [5], with necessary modification for the 802.11n aggregation and the IMA. For wireless links, we use the two-ray model to model path loss, and we assume a standard deviation of 4dB for shadowing effect. The nodes are within mutual interference range, hence RTS/CTS is not used. We assume that the links operate at 135Mbps, with 40MHz channel and the guard interval (GI) of 800ns.

Currently, most HDTV broadcasts use MPEG-2, but it is expected that more broadcasters will use MPEG-4 AVC in...
future in order to save bandwidth. MPEG-2 requires around 18Mbps for HDTV, whereas MPEG-4 AVC will use 12Mbps in early coders and 8Mbps in mature codes [6]. So in our experiments, we set the video sources to generate either 12Mbps streams or 18Mbps streams. Although the real-life MPEG video streams will have highly variable frame sizes, we use equal sized frames in the simulations for simplicity.

As in the application requirement of the 11n proposal [7], and due to the fact that the most popular MTU size is that of the Ethernet in the Internet, we assume that HDTV traffic has 1500B constant packet size. The requirement also specifies 200ms peak delay requirement, but in this paper we apply a more stringent delay limit of 100ms, considering the possibility of interactivity such as channel zapping or video conferencing. All data frames delivered beyond this delay limit are considered delay losses, and are discarded.

We use the topology shown in Fig. 2 to investigate the potential of the IMA. Specifically, we compare the video stream throughput, delay, jitter, and loss under the IMA scheme with those of the 802.11n aggregation. In the wireless HDTV distribution topology, a number of HDTV video streams are placed between hosts in the wired network (e.g., Internet) to wireless stations via the access point (AP). For instance, in the figure the wired node S1 flows a HDTV video stream to D1. In this experiment, we incrementally add the number of video streams by adding a pair of wired-wireless nodes between which the video stream flows.

![Wireless HDTV distribution topology.](image)

**B. Results**

In the given topology, the video streams converge to the AP before they are delivered to the wireless stations over the 802.11 link (Fig. 2). Thus the aggregation is performed only at the AP. Once aggregated, the video frames from multiple streams are broadcast to the receiving wireless stations. The aggregation reduces the absolute number of MAC frames, and so does the MAC transmission overhead. As a consequence, the video frame queue at the AP becomes shorter. In IMA, the aggregation is adaptive. The number of video frames shipped in an A-MPDU is automatically determined by how frequently the AP acquires the TXOP. Namely, the arriving video frames will keep being aggregated until eventually the node obtains the TXOP through contention. Again, IMA uses natural queueing delay for aggregation, thus does not add to the mouth-to-ear delay in the process of aggregation.

Fig. 3 shows the results for 12Mbps HDTV video streams. In Fig. 3(a), the throughput of each stream for the 802.11n is shown to fall drastically from 4 streams and on. It is because in this situation most video frames suffer delays larger than 100ms (Fig. 3(b)). Because the aggregation is not sufficient, more A-MPDU transmissions are incurred, and the MAC efficiency is not enough to accommodate more video streams. This situation is reflected in the loss statistics, where the delay losses dominate the 802.11n data. In contrast, the capacity overrun for IMA occurs at 8 streams. So the number of supportable streams is 3 vs. 7 on a 135Mbps PHY. Similar qualitative results are obtained for 18Mbps video streams (Fig. 4), except that the number is 2 vs. 5. So in both cases, IMA supports more than twice the streams that the 802.11n does with its native aggregation scheme. In terms of the maximum MAC efficiency of the IMA, it is over 60% in both cases.

As to the aggregation performance, Fig. 5 reveals an important qualitative difference of the two aggregation schemes. The figure plots the fraction of the video frames that are aggregated with other frames in each scheme, as a function of the traffic intensity. We notice that as the number of video streams increases, the two aggregation schemes exhibit opposite trends. The 802.11n frame aggregation progressively loses aggregation efficiency. This is because the adjacent video frames in the same stream are increasingly more likely separated by intervening frames from other streams. It is never a desirable result, since the aggregation efficiency should match the increasing traffic intensity. But note that when multiple video frames converge from scattered sources (e.g. on the Internet) to the 802.11n aggregation point, such shuffling is almost inevitable. In contrast, the IMA reacts to the heightened level of congestion due to the traffic increase by upping the degree of aggregation. This led to the higher performance demonstrated in Fig. 3 and 4, and it clearly shows that unlike the standard point-to-point frame aggregation the point-to-multipoint aggregation is resilient to congestion.

Another observation from the above experiment is that lack of isolation between flows is highly undesirable in frame aggregation. Fig. 3 and 4 show that after capacity overrun, all video streams is plummeted in terms of the per-stream throughput due to the high delay loss rates. Therefore, the aggregating AP should be aware of its video supporting capacity and exercise some admission control to prevent the quality collapse of all video streams. From the perspective of the aggregation, the capacity overrun can be signaled by unbounded growth of the MAC queue with the aggregation stretched to the maximum A-MPDU size.

**IV. RELATED WORK**

Although the current 802.11n draft has only the one-to-one aggregation, there have been proposals for one-to-many aggregation. The rationale is that there may not be enough traffic between any single pair of stations for aggregation to
Fig. 3. Results in the wireless distribution topology, 12Mbps MPEG video.

Fig. 4. Results in the wireless distribution topology, 18Mbps MPEG video.
be beneficial. The TGn Sync proposal [2] defines Multiple Receiver Aggregate Multi-Poll (MMP), which is similar to IMA. Being a standard proposal, however, it does not explore how such approach performs quantitatively and qualitatively against various traffic types, such as video.

Li et al. [8] considers an improvement of MMP. In the scheme, the inefficiency due to sequential acknowledgements are removed using multi-packet reception (MPR) enabled by signal processing and antenna array. It is shown that visible performance improvement is achieved over non-MPR schemes. In this paper, we do not consider such hardware-assisted capability for aggregation. Also, this work does not consider particular traffic characteristics, but uses generic traffic model in performance evaluation.

As to the impact of the aggregation on different types of traffic, Li et al. [4] shows that even one-to-one aggregation with partial recovery (similar to the A-MPDU partial recovery) can yield significantly higher performance than the 802.11 DCF. Also it proposes zero-waiting, a notion similar to the congestion-triggered frame aggregation. With 432Mbps PHY data rate, this work shows that the 802.11 DCF can support only 2 video streams while the proposed scheme can support up to 9 or 10 streams under 200ms delay requirement. The comparatively weak performance of this scheme is due to the fact that it uses point-to-point aggregation, while IMA uses point-to-multipoint aggregation. The proposed scheme is shown to boost the TCP and the VoIP performance as well.

Speaking of aggregation for VoIP, we show in our recent work [9] that IMA can boost the performance of the VoIP over multi-hop 802.11 networks by a factor of 7, without sacrificing the perceptual call quality.

V. CONCLUSION

The IEEE 802.11n amendment paves the way to the PHY rate which is as high as 600Mbps. Although this is sufficient bandwidth to support high-resolution video applications such as High Definition TV (HDTV), we find in this paper that the number of video streams that can be supported on the IEEE 802.11n networks depends heavily on how the frame aggregation feature is implemented. In addition to the frame aggregation scheme stipulated in the 802.11n amendment, we explore a multiple-receiver frame aggregation scheme for video traffic. The comparative study reveals through extensive simulation that the proposed multiple-receiver aggregation scheme increases the number of supported video streams by a factor of 2 or higher. We also shed light on the qualitative difference in the dynamics of the two approaches. Whereas the aggregation efficiency worsens with traffic increase in the point-to-point aggregation (which is highly undesirable), the proposed multiple-receiver aggregation exhibits resiliency against congestion, by matching the aggregation efficiency to the traffic load.

REFERENCES