

Improving VoIP call capacity of multi-hop wireless networks through self-controlled frame aggregation

Sangki Yun*, Hyogon Kim*, Heejo Lee*, and Inhye Kang†

*Korea University, †University of Seoul

Abstract—In multi-hop wireless networks, the number of supportable VoIP calls can be surprisingly small due to the increased spatial interference. To mitigate the interference, voice frame aggregation can be used. In this paper, we depart from the traditional approaches that perform aggregation at the voice source, and propose a technique called the Self-Controlled Frame Aggregation (SCFA) that runs at wireless routers. The core idea of SCFA is to let the congestion itself control the degree of aggregation. Unlike existing frame aggregation approaches, SCFA does not incur fixed delay cost, since it is used only when and by exactly as much as it is needed. In this paper, we take the example of 802.11-based multi-hop network to show the impact of SCFA, since many emerging multi-hop networks are built on the 802.11 technology. The result shows that SCFA on 802.11-based multi-hop network can boost the number of calls approximately twofold, or extends the hop distance threefold for a given number of calls to carry.

I. INTRODUCTION

With the emerging multi-hop wireless networks such as mesh networks [1], the prospect of VoIP traffic flowing on multi-hop wireless path is fast becoming a reality. Unfortunately, VoIP traffic has been shown to be highly inefficient even in single-hop wireless networks, due to the small payload size. In particular, we can show that the bandwidth efficiency for VoIP on 802.11 wireless LAN (WLAN) is on the order of a few percent [2]. This problem is generally exacerbated in multi-hop networks mainly due to increased spatial interference. One way to mitigate the interference is to reduce the absolute number of outstanding MAC frames on the wireless links. Noticing that small voice frames are each carried by a MAC frame, many prior works propose to aggregate multiple voice frames into a MAC frame [3]–[5]. The biggest issue with these proposals, however, is that the aggregation is performed by the voice source. It would be asking too much from the VoIP application designers, since the application should know if there is a wireless link on the call path, and if the wireless congestion is enough to warrant the aggregation. Notice that blindly performing aggregation would incur higher end-to-end delay.

In this paper, we propose a novel scheme called Self-Controlled Frame Aggregation (SCFA) that can run on wireless routers to maximize the number of calls on a given call path without increasing the end-to-end delay of the calls. Compared with the existing voice frame aggregation works, the proposed idea has the following contributions.

This research is supported by the Ubiquitous Autonomic Computing and Network Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

- SCFA automatically determines the optimal degree of voice frame aggregation.
- SCFA does not incur additional delay in the process of voice frame aggregation.
- SCFA transforms VoIP over the multi-hop wireless network into a delay-bounded system from a capacity-bounded system.

This paper is organized as follows. In Section II we explain the problem using a simple example. We discuss our solution to the problem, the SCFA, in Section III. The performance of SCFA is compared with that in the unmodified system and existing voice frame aggregation approaches. The existing approaches are briefly discussed in Section V, and the paper is concluded in Section VI.

II. PROBLEM DESCRIPTION

In order to illustrate the problem addressed in this paper, let us consider the simple linear topology shown in Figure 1. Notice the topology can be easily adapted to a mesh network, by replacing the end nodes on either side by the backhaul gateway. The VoIP calls are assumed to use G.729 codec [6], and the employed wireless technology, 802.11b. The carrier sensing range has been set to twice the reception range that is equal to one hop distance. The end nodes on either side generate voice traffic, which for simplicity the internal nodes only relay. Also, no cross traffic is assumed. The traffic arrival rate at each end node is denoted by λ . In case of G.729 with two voice frames shipped together, the inter-packet gap is 20ms. So it follows that the average packet arrival rate with no silence suppression is $\lambda = 50$ pkts/s. In case there are 3 calls, the voice traffic arrival rate to an intermediate node such as node 2 is 6λ , *i.e.*, 3λ in each direction.

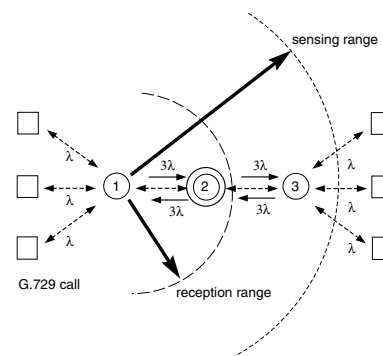


Fig. 1. Simulation topology (4-hop example).

Figure 2 shows the number of supportable calls versus the number of hops. As the number of hops that the VoIP calls have to span increases, we can support less and less calls within the required delay bound for telephony. It is 5 calls for 2 hops, 4 calls for 3 hops, and only 3 calls for 4 hops. This is because unlike in single-hop wireless LAN *the same packet can consume the wireless resource of a given node multiple times*. In Figure 1, node 1 and 3 have to either hear or transmit the same voice frame 4 times, whereas node 2 has to 5 times. If the nodes are placed more closely, the number could be larger. Notice such overhead would not be incurred without the spatial interference. The delay characteristic in

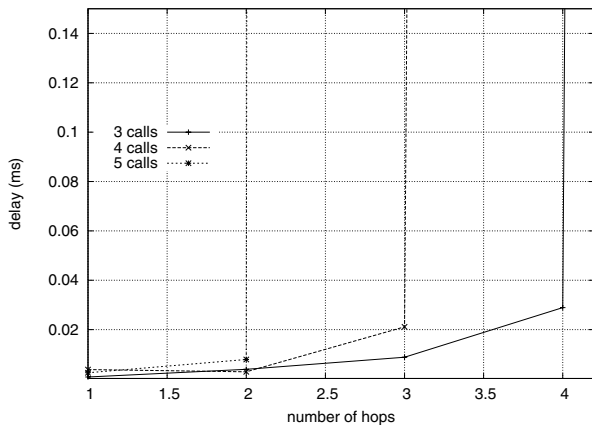


Fig. 2. Supportable calls vs. hop distance.

Figure 2 that surges over a certain hops can be explained in this vein. The average transmission time of a RTP/UDP/IP-encapsulated G.729 voice frame on a uncongested 802.11 link is around $835\mu s$ [2]. This is the average service time in queuing theoretic sense, so the service rate of the wireless nodes is bounded by $\mu = 1,200 \text{ pkts/s} = 24\lambda$. Now let us find the capacity at the bottleneck, *i.e.*, node 2. It senses transmissions from voice sources, receives from node 1 and 3, transmits the received traffic, overhears it from node 1 and 3. In the best case, the capacity at node 2 is used as shown in Table I. So this is why node 2 queue is barely stable in

TABLE I
CAPACITY USE AT BOTTLENECK NODE (BEST CASE).

Total capacity	$+24\lambda$	reason
Voice sources	-6λ	sensing
Node 1 + node 3	-6λ	reception
Node 2	-6λ	transmission
Node 1 + node 3	-6λ	overhearing
Remaining capacity	≈ 0	

Figure 2 with 3 calls at 4 hop call path given in Figure 1. But if the number of calls increases to 4, even the minimal required capacity becomes $28\lambda > \mu = 24\lambda$, and we cannot support them on 4-hop paths.

The previous example suggests that reducing the absolute number of MAC frames should be one approach to minimize the spatial interference. Voice frame aggregation can precisely cater to such needs [2]–[5]. We could conceive two

aggregation approaches on 802.11b link: *intra-call* and *inter-call* aggregation. The former packs the voice frames from the same call in the same MAC frame, whereas the latter collects the frames from multiple calls into the same MAC frame. While inter-call voice frame aggregation is shown to be highly efficient between VoIP gateways [7], it is difficult to apply it to general settings with arbitrary source and destination [2]. Since in intra-call approach the aggregation is applied to an individual packet flow, the VoIP call can be from any IP address to any IP address, and from any source port to any destination port. So in this paper we focus on the intra-call aggregation. While the intra-call aggregation is more flexible than inter-call aggregation, it suffers from low efficiency [8]. In order to aggregate n voice frames in a MAC frame, an additional delay of $(n - 1) \cdot T_s$ is introduced to the mouth-to-ear delay, where T_s is the sampling interval. Considering the parsimonious delay budget allowed for the voice call spanning multiple 802.11 links, it would be difficult for the AP or a wireless station to aggregate more than a few frames before it violates the recommended delay requirement. In fact, this is the biggest problem for most existing voice frame aggregation approaches [3]–[5]. Moreover, these proposals pay constant delay cost for gathering voice frames since the aggregation is done at the voice source, irrespective of the wireless link congestion condition [2]. For these limitations, we depart from the existing approaches and take a different approach. In particular, we take a novel view on the use of frame aggregation: we need to put squeeze on the bandwidth usage through frame aggregation *only when there is bandwidth shortage* (*i.e.*, *congestion*). We consider the aggregation unwarranted until the congestion sets in, which is signaled by multiple voice frames from the same call in the MAC queue. Moreover, exactly only those voice frames that concurrently reside in the queue are aggregated in one MAC frame, the degree of aggregation is adaptively controlled by the congestion level itself. The aggregation is thus neither overdone nor underdone, which can increase mouth-to-ear delay and lead to smaller voice capacity, respectively. Equally importantly, SCFA transforms the given multi-hop wireless system into a genuinely *delay-bounded* system. There is no delay explosion point for any load level, so as long as more delay budget is given, more calls can be supported. In sharp contrast, capacity-bounded systems have infinite delay over a certain load as shown in Figure 2. All existing voice frame aggregation approaches except SCFA share this undesirable property.

III. SELF-CONTROLLED FRAME AGGREGATION

The idea of SCFA is simple. Before the voice frame at the head of the queue is removed for the MAC layer transmission, it inspects the queue for other voice frames from the same call. If there are, SCFA packs them with the head-of-line frame. In stark contrast to existing aggregation approaches, we do not cause any additional delay since we do not wait for the subsequent voice frames. Instead, it aggregates only the frames from the same call residing in the queue concurrently. This is only right, because when there is no congestion, the voice frames from the same call will not be found together in the

queue, and we do not have to worry about the bandwidth efficiency. But as congestion sets in, more and more voice frames in the same call will find themselves concurrently cumulated in the MAC queue, which are cleared altogether when the foremost frame is removed for MAC transmission. Therefore, SCFA self-regulates and automatically adapts to the given load condition on the 802.11 WLAN. Figure 3 depicts the operation of the SCFA scheme.

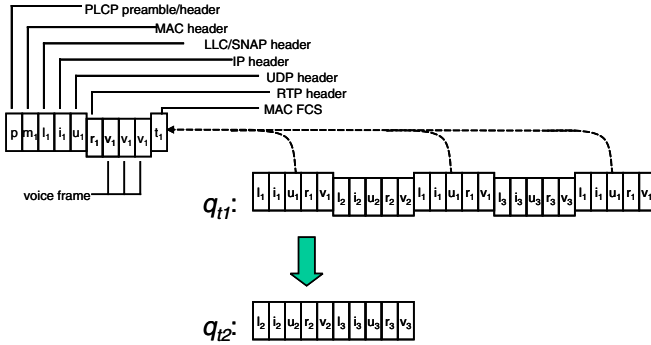


Fig. 3. SCFA operation example.

At time t_1 , the queue has 5 LLC/SNAP encapsulated voice frames (q_{t_1}). When the first voice frame is removed from the queue for MAC transmission attempt, the voice frames from call 1 are coalesced into the same MAC frame. As a consequence, a MAC frame carries out three voice frames from the queue, while the voice frames from calls 2 and 3 are left in the queue for later transmissions (q_{t_2}). Note that as voice frames are repacketized, they get to share the same RTP/UDP/IP header. So some fields in these headers need to be recomputed, while other headers (LLC/SNAP, MAC, PLCP) are unaffected except that all but one LLC/SNAP headers are dropped. First, the RTP header must correctly reflect the change in the voice sample size. Second, UDP checksum and the length fields need to be recomputed. Third, the IP total length and header checksum fields must be correspondingly updated. The IP header checksum needs re-computation because the total length field has changed.

IV. PERFORMANCE EVALUATION

In the multi-hop environment, analyzing the dynamics becomes prohibitively difficult. The topology can be complex, and so can be the interference between simultaneous traffic. Therefore, we turn to simulation to test the impact of SCFA in this paper. We regard the delay as the primary QoS metric since in the 802.11 WLAN environment, the frame loss on wireless channel is translated to delay by retransmissions up to 4 or 7 times [10]. So we will focus our attention to the voice frame losses due to delay requirement violation (a.k.a. “delay losses”) for the rest of the discussion. As to the delay requirement, the ITU-T recommendation G.108 specifies that for Internet telephony one-way mouth-to-ear delay be smaller than 250ms for G.729 and 200ms for a lower quality G.723.1 [11]. In this paper, we assume the G.729 codec is used by the voice sources, so the maximum acceptable mouth-to-ear

delay is set to be 250ms. Obviously, we have many delay components involved in the Internet telephony other than wireless link delay. There could be OS scheduling delay in case the terminal is a computer, hardware I/O (e.g., sound card) delay [12], codec frame delay, encoding/decoding delay, de-jittering (playout) buffer delay which is typically 2-3 packets long, transit delay across PSTN and IP network(s). For G.729 example that we take in this paper, the frame delay is 10 or 20ms, lookahead delay is 5ms, and de-jittering delay should be 60ms at the maximum. In the following discussions, therefore, we will assume that the sum of all these delays is at least 100ms, leaving the maximum of 150ms for our VoIP calls to cross the multi-hop network. Finally, we note that the voice traffic is usually modeled by ON-OFF model [13]. This is because telephone call traffic is an alternating series of talk spurts and silence periods. For instance, the voice activity factor is given by $\alpha = 0.43$ in Brady’s model [13]. For simplicity in qualitative presentation, however, we model the VoIP applications as generating CBR traffic in this paper, i.e., $\alpha = 1$.

A. Coverage and capacity

We start by looking at the impact of SCFA from two different angles: *coverage*, and *capacity*. The coverage shows us how many wireless links a VoIP call can span in the linear topology before it violates the given delay bound. On the other hand, the capacity metric measures the number of calls that can be supported given a fixed wireless hop distance.

Figure 4 shows the number of supportable calls versus the number of hops that the calls travel. Here we assume that the wireless router queue size is 50. In practice vendor implementations of the wireless router queue can wildly vary. In particular, the 50 slot setting roughly maps to 50ms of queuing under the maximum 802.11b speed of 11Mbps, which is about 20% of the rule-of-thumb guideline for wireline router configuration [14]. We set it to a small value since the VoIP traffic is sensitive to delay, and it must go over multiple wireless links in the multi-hop network. We notice in Figure

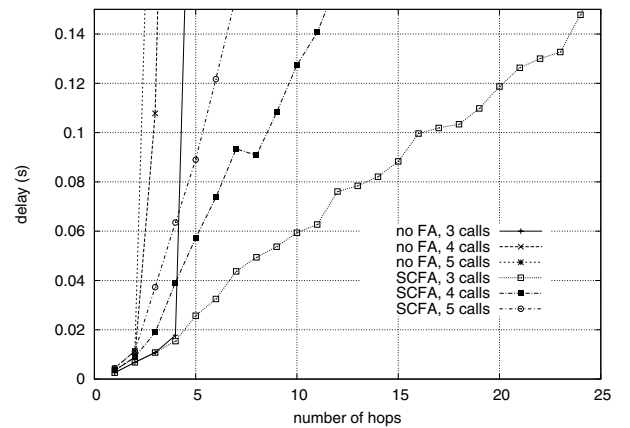


Fig. 4. Wireless delay vs. number of hops.

4 that with the same number of VoIP calls contending for the wireless bandwidth, SCFA extends the span of the QoS-satisfying VoIP call radius at least three-fold. For instance, the

maximum number of reachable hops subject to the peak delay requirement of $\hat{d}_w = 150\text{ms}$ with no frame aggregation and SCFA is as follows. With 3 calls it is 4 versus more than 20 calls, with 4 calls it is 3 versus more than 10, with 5 calls it is 2 versus 6 hops, respectively.

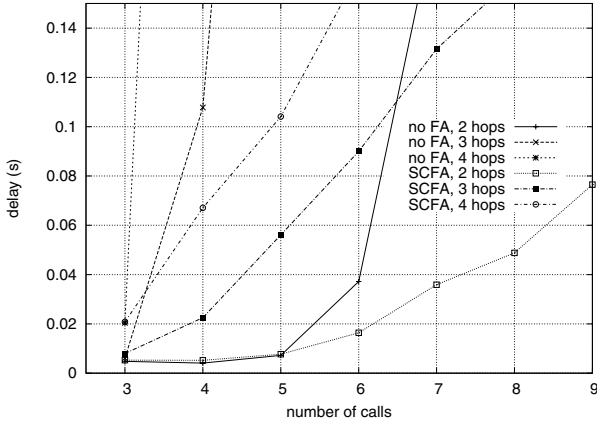


Fig. 5. Wireless delay vs. number of calls.

Figure 5 shows the SCFA performance from the capacity angle. Here we estimate the number of supportable calls when the wireless span of the calls are fixed to 2, 3, or 4. Again, we see the SCFA is more effective whether or not the wireless delay is less than $\hat{d}_w < 150\text{ms}$. We observe from the figure that the capacity under SCFA is nearly doubled for each case. Specifically, with 2 hops it is 6 versus more than 10 calls, with 3 hops it is 4 versus 7, with 4 hops it is 3 versus 5 calls, respectively.

B. Delay-boundedness

By now, one might wonder if there is any difference between the existing source-based frame aggregation [3]–[5] and the SCFA. In this section, we show that there is a qualitative difference (*i.e.*, queuing dynamics) that leads to the difference in performance: incurred delay cost and the robustness to congestion. In essence, source-based aggregation is *capacity-bounded* just like an unmodified system. In stark contrast, SCFA is genuinely *delay-bounded*.

In source-based frame aggregation, the voice traffic source packs multiple voice samples or a single enlarged voice sample in the same RTP/UDP/IP packet. Not knowing how much congestion is occurring, or even whether the call goes through a wireless link(s) down the call path for that matter, the voice source has to blindly determine the amount of voice sample to be put into the same packet. In contrast, the SCFA automatically reflects the wireless link congestion to the degree of aggregation. The fact that the SCFA dynamics is well modeled by the M/M/1/M [2], which has strongly self-regulating property [16], testifies to the fact. Notice this adaptive property is unique to the SCFA. Figure 6 illuminates the difference in dynamics and efficiency between the two approaches. Here we compare the SCFA with 20ms payload size against the source-based frame aggregation with 50ms voice payload size (we intentionally give favor to the source-based aggregation). We notice the following. First, the source-

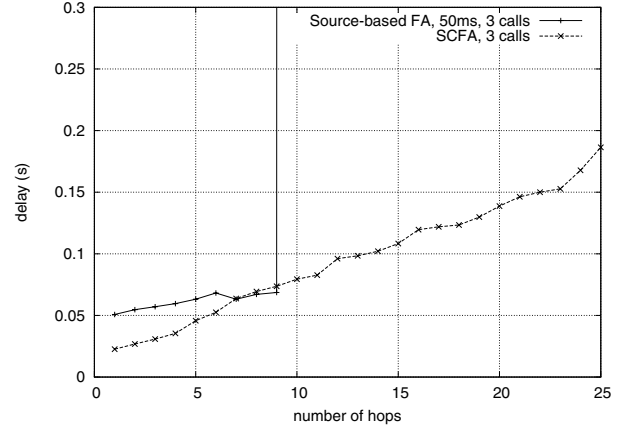


Fig. 6. Wireless delay for VoIP of SCFA vs. source-based aggregation.

based aggregation pays the fixed delay cost from the start. The 50ms aggregation delay is a fixed cost that is always added to the mouth-to-ear delay. Second, the delay for the source-based aggregation surges at around the point where its delay curve meets that of the SCFA. Since the SCFA is optimal, only around this point the source-based aggregation is optimal. However, beyond this point, the source-based aggregation cannot sustain the load using the fixed (50ms) aggregation. So it is capacity-bounded in queuing theoretic sense. As a result, it diverges whereas the SCFA keeps stabilizing the system by using increased aggregation. In essence, SCFA is delay-bounded. If only larger delay budget is available, still more calls can be accommodated. And the SCFA can always outperform the source-based aggregation.

C. Performance under loss

So far, we have assumed that there is no frame loss on the wireless path, which in practice is most probably not true. And the perceptible call quality is indeed sensitive to voice frame losses. Since frame aggregation enlarges the voice payload per packet, it proportionately increases the frame error rate. In this section, we briefly consider the impact of wireless loss on the performance of SCFA.

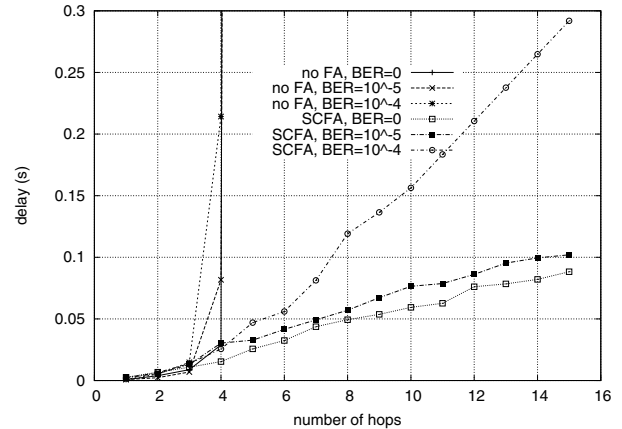


Fig. 7. Wireless delay for SCFA under loss, 3 calls.

Recollect that the 802.11 MAC copes with losses with retransmissions, so most losses are transformed to delays. For

instance, if the MAC frame error rate (FER) is 20%, with *LongRetryLimit* (=4) retransmissions the probability of MAC frame loss is 3.5 out of 10,000. Since less than 0.19% voice loss rate does not affect the MOS rating for G.729 [5], even 20% FER *per se* cannot be a problem. But retransmissions increase the jitter, which does decrease the number of voice frames arriving within \hat{d}_w . Figure 7 shows the wireless delays for the 3-call scenario, under bit error rates (BERs) ranging from 0 to 10^{-4} . Note that the BER of 10^{-4} is a high value. Given

$$e_f = 1 - (1 - e_b)^N \approx e_b \cdot N, \quad (1)$$

where N is the number of bits in a frame, e_f is the frame error rate and e_b is the BER, it would give us the frame error rate of 1 for 1500B, the most popular packet size in the Internet. We notice in Figure 7 that the number of hops has not changed for the no frame aggregation case. It is still limited to 4. But for the SCFA, the number of hops has been reduced significantly. The coverage under $\hat{d}_w = 150\text{ms}$ is now 9 hops for BER $e_b = 10^{-4}$, compared to more than 20 hops without loss (see Figure 4). However, the number of hops that can be covered under the SCFA is still quite large. SCFA can support 3 calls up to 9 hops, but a 9-hop deep multi-hop network will be quite rare. For small number of hops, the SCFA performance is comparable to the no aggregation case. It suggests that the impact of larger frame size with frame aggregation is not as extensive as to cancel out the gain obtained by reducing the number of MAC frames and hence the interference.

V. RELATED WORK

Most of the voice frame aggregation works so far are for the single-hop wireless environment. Wang *et al.* [15] uses the inter-call aggregation in the single-hop wireless LAN context. In order to guarantee simultaneous delivery to multiple wireless stations, it utilizes IP multicast over link-level (i.e., 802.11) multicast. The scheme shows that the allowable number of calls can be increased nearly twofold. However, it requires new frame format and 802.11-level multicast [2], which SCFA does not require. Garg and Kappes [4], Hole and Tobagi [5], and Medepalli *et al.* [3] all show that more VoIP calls can be supported with source-based aggregation. But as we discussed in Section IV-B, these source-based schemes are capacity-bounded, and results in suboptimal performance. Ng *et al.* [17] applies the frame aggregation idea to the multi-hop wireless network. It uses the same inter-call frame aggregation method as in [15], over the calls sharing the exactly same call path in the multi-hop topology. However, it would require a significant number of VoIP calls to have those that span the same wireless path in order to obtain the desired efficiency. Moreover, it also requires IP and 802.11 multicast support.

VI. CONCLUSION

Frame aggregation is arguably the most effective measure against the low efficiency in carrying VoIP calls across the multi-hop network, where the calls may cross multiple wireless links. In this paper, we propose self-controlled frame aggregation at the wireless routers in the multi-hop network. Its simple idea is to aggregate the voice frames from the same call when

the frames concurrently reside in the same router queue. The SCFA algorithm is optimal in the sense that it does not incur any additional delay cost for aggregation, since the aggregated voice frames depart with the first frame in the queue. It is genuinely delay-bounded and has excellent resilience against VoIP traffic load increase. The SCFA is shown to increase the call capacity of the multi-hop network almost two-fold. In terms of the coverage, the SCFA increases the number of hops that a VoIP call covers three-fold.

REFERENCES

- [1] R. Bruno, M. Conti, and E. Gregori, "Mesh Networks: Commodity Multihop Ad Hoc Networks," *IEEE Communications Magazine*, March 2005.
- [2] S. Yun, H. Kim, "Squeezing 100+ VoIP Calls out of 802.11b Networks," submitted for publication. Available at <http://widen.korea.ac.kr/V100.pdf>.
- [3] K. Medepalli, P. Gopalakrishnan, D. Famolari, T. Kodama, "Voice capacity of IEEE 802.11b, 802.11a, and 802.11g wireless LANs," Proceedings of IEEE Globecom 2004.
- [4] S. Garg, M. Kappes, "Can I add a VoIP call?," Proceedings of IEEE ICC, 2003.
- [5] D. Hole, F. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," Proceedings of IEEE ICC 2004.
- [6] *Coding of speech at 8kbit/s using conjugate-structure algebraic-ode-excited-linear-prediction*, ITU-T Recommendation G.729, Mar. 1996.
- [7] H. Kim, I. Kang, E. Hwang, "Measurement-based multi-call voice frame grouping in Internet telephony," *IEEE Communications Letters*, 6(5), May 2002.
- [8] H. Kim, I. Kang, "Measurement-based Frame Grouping in Internet Telephony," *IEE Electronics Letters*, 37(1), Jan. 2001.
- [9] *Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s*, ITU-T Recommendation G.723.1, Mar. 1996.
- [10] ANSI/IEEE, "802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1997.
- [11] *Application of the E-model: A planning guide*, ITU-T Recommendation G.108, Sept. 1999.
- [12] T. J. Kostas, M. S. Borella, I. Sidhu, G. M. Schuster, J. Grabiec, and J. Mahler, "Real-time voice over packet-switched networks," *IEEE Network Magazine*, 12(1), Jan. 1998.
- [13] P. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation," *Bell Syst. Tech. Journal*, 48(7), Sept. 1969.
- [14] R. Bush and D. Meyer, *Some Internet Architectural Guidelines and Philosophy*, RFC 3439.
- [15] W. Wang, S. C. Liew, V. O. K. Li, "Solutions to performance problems in VoIP over a 802.11 wireless LAN," *IEEE Transactions on Vehicular Technology*, 54(1), Jan. 2005.
- [16] L. Kleinrock, *Queueing Theory, Vol. 1*, Wiley, 1976.
- [17] P. C. Ng, S. C. Liew, W. Wang, "Voice over wireless LAN via wireless distribution system," Proceedings of IEEE Vehicular Technology Conference (VTC), 2004.