# 100+ VoIP Calls on 802.11b: The Power of Combining Voice Frame Aggregation and Uplink-Downlink Bandwidth Control in Wireless LANs

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

***Abstract*— The bandwidth efficiency of Voice over IP (VoIP) traffic on the IEEE 802.11 WLAN is notoriously low. VoIP over 802.11 incurs high bandwidth cost for voice frame packetization and MAC/PHY framing, which is aggravated by channel access overhead. For instance, 10 calls with the G.729 codec can barely be supported on 802.11b with acceptable QoS – less than 2% efficiency. As WLANs and VoIP services become increasingly widespread, this inefficiency must be overcome. This paper proposes a solution that boosts the efficiency high enough to support a significantly larger number of calls than existing schemes, with fair call quality. The solution comes in two parts: adaptive frame aggregation and uplink/downlink bandwidth equalization. The former reduces the absolute number of MAC frames according to the link congestion level, and the latter balances the bandwidth usage between the access point (AP) and wireless stations. When used in combination, they yield superior performance, for instance, supporting more than 100 VoIP calls over a IEEE 802.11b link. We demonstrate the performance of the proposed approach through extensive simulation, and validate the simulation through analysis.**

## I. Introduction

**A**RGUABLY the two most popular trends lately are the widespread deployment of IEEE 802.11 wireless local area networks (WLANs) and the surge in Voice over IP (VoIP) services. IEEE 802.11 [1]–[5] has been popular for the Internet access in homes, schools, and businesses. At the same time, VoIP software is widespread, such as Skype [6], which has been downloaded more than 200 million times, and there are between 3.5 to 4 million simultaneous Skype users at any time [7]. Given these developments, it is necessary to explore and resolve the issues that arise when these technologies meet. In this paper, we attempt to address one of the most challenging issues when VoIP services are ran over IEEE 802.11 WLANs: *bandwidth inefficiency*.

We have known for some time that running VoIP over 802.11 LANs is extremely inefficient [8]–[11]. The table I exemplifies the timing overhead for sending a G.729 codec [12], and demonstrates that the efficiency for VoIP calls in the G.729 codec is bounded by 1.83%, even in the absence of competing traffic. This is the reason the maximum number of sustainable number of calls on the 802.11b link is referred to be only in the order of 10s [8], 20s [9], or even less than 10 [10] depending on the voice traffic model and the codec.

TABLE I
TYPICAL TIMING OVERHEAD OF A SINGLE VOICE FRAME TRANSMISSION OVER 802.11B LINK

| Delay component | Time ($\mu$s). | Reason |
|---|---|---|
| DCF Inter Frame Space (DIFS) | 50. | |
| Average channel access delay due to CA | 310. | 31/2 slots $\times$ 20 $\mu$s |
| Voice Frame (G.729) | 14.55 | 20/1.375(Msps) |
| RTP/UDP/IP encapsulation | 29.09 | 40/1.375 |
| LLC/SNAP encapsulation | 7.27 | 10/1.375 |
| MAC header and trailer | 20.36 | 28/1.375 |
| PLCP preamble and header | 192. | Long preamble |
| Short Inter Frame Space (SIFS) | 10. | |
| PLCP preamble and header (for ACK) | 192. | Long preamble |
| MAC header and trailer (for ACK) | 10.18 | 14/1.375 |
| Total | 835.45 | |

There are various approaches to improving the efficiency : header compression [13], [14], frame aggregation [8], [15], [16], and medium access control (MAC) protocol modification [17]–[19]. As to the header compression, a crucial observation from Table I is that the voice packetization overhead at and above the network layer (*i.e.*, Internet Protocol (IP)/ User Datagram Protocol (UDP)/ Real-time Transport Protocol (RTP)) is actually small. It accounts for only 3.66% of the total overhead, suggesting that from the header compression would generally be insignificant. The 802.11 MAC modification approach has been popular for service differentiation and real-time traffic Quality of Service (QoS) support. The Enhanced Distributed Channel Access (EDCA) in the 802.11e extension [20] also enables preferential treatment of real-time traffic through 802.11 parameter configuration. However, the EDCA itself is not a solution to the low-capacity problem, because it does not reduce the MAC framing overhead.

The fact that the bulk of the overhead lies in channel access, MAC/PHY framing, and 802.11 ACK transmission and inter-frame spaces (IFSs), suggests that reducing the sheer number of MAC transmissions through VoIP frame aggregation should be the most effective approach. For example, in wired networks, frame aggregation for VoIP traffic has been shown to yield bandwidth savings [15], [16]. In the 802.11 WLAN environment, this technique should be all the more effective, due to the existence of channel access overhead. Therefore, in this paper, we explore frame aggregation as a solution to the bandwidth inefficiency problem.

Manuscript received May 15, 2006; revised November 27, 2006.
Sangki Yun, Hyogon Kim, and Heejo Lee are with Korea University (e-mails: {lockdown99,hyogon,heejo}@korea.ac.kr).
Inhye Kang is with the University of Seoul (e-mail: inhye@uos.ac.kr).
Digital Object Identifier 10.1109/JSAC.2007.070506.

## II. Related work

In this section, we briefly summarize related work on frame aggregation, in addition to other techniques that efficiently carry the voice traffic over a single-hop 802.11 wireless LAN.

The voice capacity limitation issue in 802.11 networks has been raised in many studies [8]–[11], [17]. Garg and Kappes [10] notice that the framing overhead is one of the bottlenecks. They propose to increase the number of supportable calls by enlarging the RTP payload size. This approach has an advantage in terms of implementation complexity, since it can be implemented with the VoIP application, without requiring any cross-layer processing. A shortcoming of this approach is that the increased framing delay is always added to the mouth-to-ear delay, to create large voice payload. In addition, at the voice source, it is difficult to even know if the call would go over a wireless link, so this approach does not apply to the specific bandwidth inefficiency problem in VoIP over the 802.11 wireless LAN. Similar ideas were proposed by Medepalli *et al.* [9] and Hole and Tobagi [11] share the same pros and cons.

Wang *et al.* [8] proposes the M-M scheme, which uses MAC-level multicast and inter-call frame aggregation. It uses 802.11 multicast to deliver aggregated voice frames to originally intended receivers. The inter-call aggregation can result in high efficiency, with only small additional delay for aggregation if there are many concurrent calls, and the 802.11 multicast eliminates some overhead caused by the transmission of 802.11 acknowledgements (as shown in Table I), because the 802.11 multicast does not perform retransmission. However, for the very same reason, the MAC layer loss directly leads to multiple voice frame losses aggregated in the lost MAC frame. It can be undesirable on the 802.11 link where a modest number of competing nodes results in very high collision probability, or if the link is loss-prone due to bad channel conditions, as losses more severely affect call quality than delay [11].

Baldwin *et al.* [17] uses adaptive $CW_{min}$ control and deadline-based MAC queue management. The MAC queue drops packets that miss the preset deadline, transferring the bandwidth to in-time packets. The concern with this scheme is that it is difficult to know if a "stale" realtime packet is indeed useless at the receiver (in some cases, the delay requirement may even be changed by adaptively shifting the playout schedule [21]), so the dropping decision is awkward to enforce inside the network.

## III. Proposed ideas

In this section, we propose a novel approach that enables an 802.11b network to support over 100 VoIP calls. We take the example of the 802.11b network in this paper for expositional purposes, but the proposal is directly applicable to any IEEE 802.11 wireless network, to create maximal number of calls within the given bandwidth allocation. For convenience, throughout this paper, we will call our proposed approach "V100". V100 has two components: frame aggregation and $CW_{min}$ control. In this section, we discuss the two components of V100.

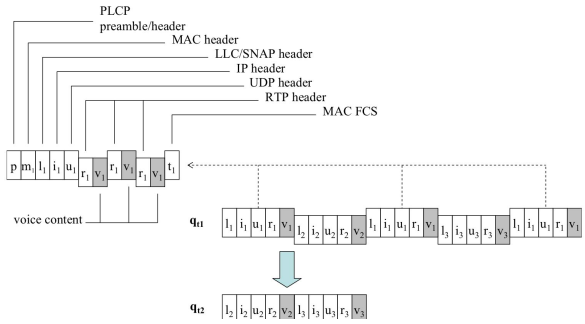
### A. Zero-delay frame aggregation

In the voice frame aggregation, we attempt to carry multiple voice frames in a single MAC frame. We could conceive of two aggregation approaches on 802.11b links. *Intra*-call aggregation refers to when voice frames from the same call are packed in the same MAC frame. In contrast, *Inter*-call aggregation packs the voice frames from different calls into the same MAC frame.

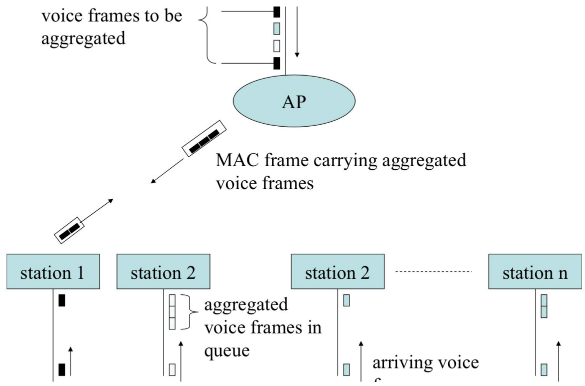
Wang *et al.* [8] proposes inter-call aggregation for the downlink traffic on the 802.11 WLAN. In order to deliver a MAC frame that contains voice frames to disparate stations, the 802.11 multicast is used, where a shim header is inserted to identify the aggregation. Despite the potential for high efficiency, however, this scheme may result in a high voice frame loss rate as we described earlier. Not only poor channel conditions can cause the losses, but so can modest number of contending stations. It is well known that the 802.11b MAC incurs high collision probabilities, *e.g.* 5% with 2 contending stations and 10% with 3 stations. The lack of acknowledgement (hence 802.11 retransmission) in such an environment can be fatal. The inter-call aggregation has this issue with the 802.11 multicast, therefore, we turn to the intra-call aggregation in this paper.

An obvious shortcoming of the intra-call aggregation, however, is the lower efficiency than in the inter-call aggregation [16]. In order to aggregate $k$ voice frames in a MAC frame, an additional delay of $(k-1) \cdot T_F$ is introduced in the mouth-to-ear delay, where $T_F$ is the voice framing interval. For instance, to collect $k$ incoming G.723.1 frames, we need to wait $(k-1) \cdot 30ms$. In considering the parsimonious delay budget allowed for the 802.11b link in VoIP applications, it would be difficult for the AP or wireless station to gather more than a few frames, before it violates the recommended delay requirement.

As to the shortcoming of the intra-call frame grouping proposed in previous works, we take a different view on its use: *we need to save bandwidth through frame aggregation only when there is bandwidth shortage* (*i.e.*, congestion). In essence, we need not trigger aggregation until multiple voice frames from the same call can accumulate in the MAC queue, which would incur the aforementioned high delay overhead. Therefore, we do the following in V100. Before the voice frame at the head of the interface queue is removed for the MAC layer transmission, V100 inspects the queue for other voice frames from the same call. If there is, V100 performs frame aggregation. We remark here that no additional delay is caused by doing this, since we do not delay the head-of-line voice frame for merging with the subsequently arriving voice frames. (So the method is called zero-delay frame aggregation (ZFA)). Instead, it aggregates frames from the same call if they are found residing in the queue concurrently. This is only right, because when there is no congestion, there is no concern of bandwidth efficiency. As congestion sets in, more and more voice frames in the same call will be accumulated on the MAC queue, which are altogether cleared when the foremost frame is removed for MAC transmission. Therefore, ZFA self-regulates to automatically adapt to the given load condition on the 802.11 WLAN.




(a) MAC frame format with voice frame aggregation and queue configuration before ($t = t_1$) and after ZFA ($t = t_2$). Subscripts are call numbers.




(b) ZFA can occur for both uplink and downlink.

Fig. 1. Zero-delay Frame Aggregation (ZFA).

Fig. 1 (a) depicts the operation of the ZFA scheme as well as the MAC frame format that transports multiple voice frames. The alphabet symbols represent the type of each component (*i.e.*, i: IP header, v: voice frame), where the subscripts represent the call that each component belongs (*i.e.*, $v_1$ is voice content for call "1", whereas $v_2$ and $v_3$ are two additional calls). At time $t_1$, the queue has 5 Logical Link Control/Subnetwork Access Protocol (LLC/SNAP) encapsulated voice frames ($q_{t_1}$). When the first voice frame is removed from the queue for a MAC transmission attempt, the voice frames from the same call ("1") are coalesced into the same MAC frame. As a consequence, the MAC frame carries out three voice frames from the queue, while the voice frames from other calls ("2" and "3") are left in the queue for later transmission ($q_{t_2}$). Notice that we need cross-layer inspection of the frame header in order to identify the call. In particular, the network (IP) and transport (UDP) layer headers need to be inspected. Fig. 1 (b) illustrates where the ZFA technique is applied. This is used where voice frames from the same call can accumulate, *i.e.*, both in the AP queue and in the queue of the station.

In the ZFA, as voice frames are aggregated, they share the same UDP/IP header. This means that some fields in these two headers need to be recomputed. First, UDP checksum and the length fields need to be recomputed. Second, 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. The LLC/SNAP header does not have any field affected by aggregation, accordingly, all but one LLC/SNAP header is dropped.

When the aggregated frame arrives at the receiver, it is de-aggregated in the MAC-level to be transferred to the upper layer. First, the UDP headers for each voice frame are reconstructed. The port pair is recovered from the UDP header of the aggregated frame, and the checksum and length field are recomputed. Likewise, the IP header is attached to each recovered UDP-encapsulated voice frame after the recovery of all fields, except the checksum and the length field, which should be recomputed. The LLC/SNAP headers are recovered from the aggregated frame and attached to each recovered IP datagram. Then, these recovered LLC/SNAP encapsulated frames are sent up to the LLC layer.

### B. $CW_{min}$ adaptation for up-down asymmetry and collision control

The second idea, to boost the number of calls, is to restore symmetry to uplink/downlink bandwidth distribution. As a result of the design feature of 802.11 MAC, all stations are entitled to an equal share of the channel bandwidth [22], and the AP is no exception to this rule. However, VoIP is a symmetric application that creates comparable traffic volume in either direction. In the face of active uplink traffic from wireless stations, the AP can be limited to only $1/k$ of channel bandwidth where $k$ is the number of actively transmitting stations including the AP. However, since AP represents $k-1$ calls, the AP can become the bottleneck if $k$ is large and/or if the bandwidth allotment for VoIP category (*e.g.*, the AC_VO in 802.11e [20]) is small. This asymmetry issue has recently been noticed for general data traffic [23], [24], but it is particularly acute for symmetric applications such as VoIP. Therefore, V100 adapts the bandwidth share of the AP so that the up/down balance is struck. For this, V100 uses $CW_{min}$ adaptation, which can precisely control the bandwidth distribution among stations. The bandwidth share ratio is inversely proportional to $CW_{min}$ ratio [25], [26]. The $CW_{min}$ for the AP is set to the $\left(\frac{1}{k-1}\right)^{th}$ of that for actively transmitting wireless stations. The new 802.11e [20] standard endorses the dynamic change of $CW_{min}$. For tracking the number of actively transmitting stations $k$ in the voice category, we discuss a novel scheme in Section IV-B in detail.

Note that the increase in the stations originating VoIP calls not only aggravates the asymmetry, but also increases the absolute traffic load. Therefore, by regulating the $CW_{min}$ for wireless stations, the adaptation also has to cope with the increased load. In the next section, we demonstrate that this $CW_{min}$ configuration approach indeed leads to optimal performance. For convenience, we will refer to this feature of V100 as Contention Window Adaptation (CWA). V100 is used to refer to the combination of ZFA and CWA in Section IV.

### C. Implementation issues

In this section, we briefly discuss various issues related to the implementation of V100.

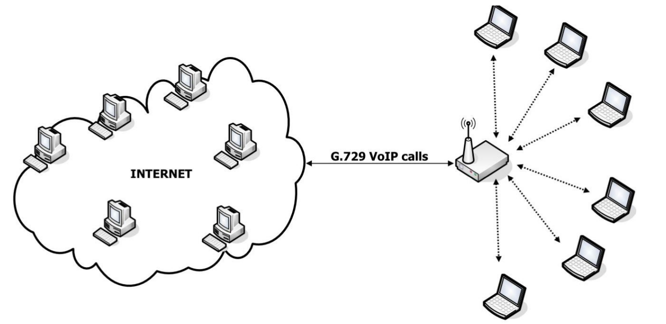



Fig. 2. Simulation topology.

Identifying VoIP frames in the packet stream is an issue in itself, which we simply assume is completed before V100 processes them, as other related works also assume. (Although the 802.11e extension specifies a separate queue for the AC_VO traffic class, the standard does not tell us how we can discern voice frames from other types of traffic, not to mention the original 802.11.) However, one way to classify VoIP packets is deep packet inspection. For instance, a heuristic method to classify VoIP packets was proposed in Kim *et al*. [27]. But a more practical and cheaper alternative for the AP is to use the frame size and the protocol (UDP) as the indicator of VoIP traffic [18], where small frames with UDP transport are considered voice frames. For simpler implementation of V100, we could use this quick-and-dirty type heuristic.

The frame aggregation/deaggregation processing in the interface between the MAC and LLC layer can be implemented at the device driver level, as shown by [28]. In the case IP Security (IPSEC) is used, the IP header reveals in its protocol number field whether IPSEC is being used, in which case the interface should be instructed to not perform the aggregation. In case the RTP/UDP/IP header compression such as Robust Header Compression (ROHC) [14] is used (even when the 802.11 is broadband), we believe it too should be detectable, and the frame aggregation can be disabled correspondingly. Skype, however, appears to use neither, although it uses AES to encrypt the voice payload [29].

In addition to the complexity of the implementation listed above, the peering requirement that both the AP and the wireless station should be able to understand the aggregated frame format on the MAC layer, is the practical difficulty of the proposed scheme. One solution is to use the shim header as in [8], [28], in a way to secure backward compatibility for the peers that do not understand the new format.

## IV. Analysis of the Proposed Scheme

In this section, the performance of ZFA and CWA is analyzed both separately and together. In order to validate the analytical results, we compare them with simulation, for which we use `ns-2` [30], with necessary modifications.

First, we lay the assumptions and approximations to make in the analysis. We assume that VoIP traffic is constant bit-rate (CBR) traffic with a 20ms interval, as in the G.729 codec. In the telephony jargon, the voice activity factor $\alpha$ is 1. Although simplistic, this straightforward model assists in simplifying the analysis and sheds light on notable properties of V100. In Section V, however, we will consider the realistic $\alpha$ value.

TABLE II

Parameters Used in the Performance Evaluation

| Parameter | Values |
|---|---|
| Network Architecture | Infrastructure Basic Service Set (BSS) |
| Simulation time | 300 seconds |
| Transmission Rate | 11Mbps |
| PLCP Header | 192 $\mu$s (Long Preamble) |
| Voice Codec | G.729 |
| Framing Interval | 20 ms |
| RTS-CTS | Disabled |

The performance of V100 is analyzed based on the *average* delay analysis. Although peak delay is the most practical metric to measure the number of possible calls, the average delay analysis is sufficient to reveal both the quantitative and qualitative differences of ZFA and CWA, as compared with either the unmodified 802.11 system or other frame aggregation approaches. The average delay analysis is also much simpler than the peak delay analysis. Again, we will eventually turn to the *peak*-delay in Section V, to estimate the number of supportable calls.

The wireless station and AP are dealt with differently. For the wireless station, the non-saturated condition is assumed, because, even with $\alpha = 1$, the stations may not always be backlogged with voice frames, due to the framing interval of the voice codec. On the other hand, due to the up-down asymmetry, the AP can easily be backlogged, even with only a few calls, in the absence of aggregation. Hence, we assume that the AP is saturated. However, we assume the MAC layer queue size to be infinite for both the AP and the stations.

Finally, for the simulation experiments that check the validity of the analysis, the default values defined in the 802.11b standard are used for the MAC parameters [2]. Other parameters used in the evaluation are summarized in Table II. The simulation focuses on the wireless link between the AP and stations, not on the wired path dynamics. Each experiment simulates 300 seconds of the system dynamics, of which the first 30 seconds are discarded as perturbation. For each case, 3 simulation instances are run with different seeds. The simulation topology is shown in Fig. 2, where multiple wireless stations (WSTAs) make calls with peers in the wired Internet through the AP.

### A. Average access delay under ZFA

When the MAC layer receives a packet from the application, the time until transmission is composed of queuing delay and channel access delay. The latter is time spent waiting for transmission at the head of the transmission queue, while the former is time spent proceeding to the head position. Under ZFA, however, we can view the uplink voice frames as experiencing zero queuing delay. Suppose that the head-of-line (HOL) voice frame is $v_j$ upon the arrival of another voice frame $v_k$ to the queue. The sojourn time of $v_k$ is bounded by the channel access delay of $v_j$, since as soon as $v_j$ is transmitted, $v_k$ is also transmitted . Hence, we can conveniently regard that only transmission delay exists for the aggregated voice frames $v_k \neq v_j$ in the uplink. Under this convention, the queuing delay cannot be positive, except when there are so many voice frames in the queue, so that a single MAC service data unit (MSDU) cannot carry them. With the maximum MSDU size of 2304 bytes, the channel access delay of $v_j$ is larger than $2304/20 \cdot 20ms = 2304ms$. Only an extremely overloaded network can cause such unacceptable channel access delay, and VoIP calls would be impossible due to the QoS problem. Accordingly, we exclude such case as impractical.

While the voice frames in the same call are aggregated, disparate calls cannot be transported in the same MAC frame (ours is an intra-call aggregation). So in the downlink, queuing delay exists. The downlink can be analyzed using the M/M/1//$M$ model [31], with the customer population (*i.e.*, number of calls) limited to $M$. Then, the expected AP queue length in the steady state is:

$$E[Q^d] = \sum_{k=0}^{M} \frac{k}{\sum_{l=0}^{M} \left(\frac{\lambda}{\mu}\right)^l \frac{l!}{(M-l)!}} \left(\frac{\lambda}{\mu}\right)^k \frac{k!}{(M-k)!}$$

Here, $\lambda$ denotes the arrival rate with which a station in the idle state attempts transmission of a voice frame. In terms of M/M/1//$M$, a customer transitions to "arriving" state as soon as service is completed. Therefore, $\lambda$ is the rate at which a customer finally arrives, exiting the arriving state. In ZFA, a call is serviced when its voice frames are shipped out in a MAC frame. The time between this instant and the next voice frame arrival constitutes the sojourn time at the arriving state (*i.e.*, idle state $(-1, 0)$). In accordance with the assumption above, $\lambda = 1/10ms$. The service rate $\mu$ is given by:

$$\mu = \frac{1}{E[T_i^d] + E[T_o^d] + E[T_s^d] + cE[T_c^d]}$$

where $E[T_i]$ is the time spent for backoffs, $E[T_o]$ is the channel time occupied by other stations, and $E[T_s]$ and $E[T_c]$ are time consumed in successful transmission and collision, respectively. $c = p/(1-p)$ is the average number of collisions per transmission attempt. The M/M/1//$M$ dynamics can be shown to be sufficiently close to the real system behavior [32].

The fact that ZFA has M/M/1//$M$-like queuing dynamics is a fundamental advantage of ZFA. A typical G/G/1 queue which can model the vanilla 802.11 suffers from the constraint that the service rate should be larger than the arrival rate for the existence of a steady-state (*i.e.*, the ergodicity condition). In other words, if average service time exceeds average inter-arrival time, the delay becomes unbounded. In contrast, in the M/M/1//$M$ queue, ergodicity is always assured, because longer service time discourages new arrival [31]. So ZFA does not suffer from the problem of delay explosion, where we call such a system *capacity-bounded*. Only the delay budget matters, where we call such system a *delay-bounded* system. This is the main reason that ZFA changes the 802.11 WLAN from capacity-bounded to delay-bounded.

Now, the uplink and the downlink delays can be given by

$$\begin{aligned} E[D^u] &= E[T_i^u] + E[T_o^u] + E[T_s^u] + cE[T_c^u] \\ E[D^d] &= \left(E[T_i^d] + E[T_o^d] + E[T_s^d] + cE[T_c^d]\right) \cdot E[Q^d], \end{aligned} \tag{1}$$

which can be obtained through a Markovian analysis [32].

Finally, the average delay given in (1) is the delay of the first voice frame $v_j$ in the MAC frame. The voice frames $v_k$ $(k > j)$aggregated together with $v_j$, experience less delay

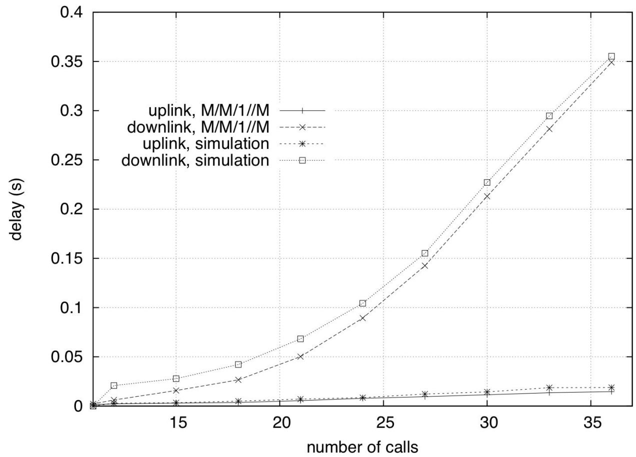



Fig. 3. Calls vs. 802.11 link delays under ZFA.

since they arrived later. Specifically, they experience $(j-i) \cdot T_F$ less delay. Accordingly, we must modify (1) as follows:

$$E[D'] = \sum_{i=0}^{\lfloor E[V] \rfloor} \frac{E[D] - i \cdot T_F}{E[V]}, \tag{2}$$

where $E[V]$ is the average number of voice frames aggregated in a MAC frame, and is given by $E[V] = E[D]/T_F + 1$.

Fig. 3 shows the uplink and downlink delays against the number of calls, given by the analysis in (2). If the average delay budget for the 802.11b link $\bar{d}_{req}^{802.11}$ is 100ms, for instance, the number of sustainable calls is approximately 24, as the downlink hits the limit first. This result with ZFA is approximately a 2-fold increase in the average number of calls, compared to the vanilla 802.11b, as we will show below. This is still far less than we would expect out of the 802.11b capacity (G.729 codec rate is only 8kbps). To make the improvement even greater, we introduce the second feature of V100: CWA. In the next section, we demonstrate that when ZFA is combined with CWA, significant improvement is achieved.

### *B. CWA: capacity boosting through equalizing delay*

In CWA, we attempt to provide more share of the 802.11 bandwidth to the AP than the 802.11 protocol, with the goal of resolving the huge delay difference observed in Fig. 3, which would allow us to accommodate large numbers of VoIP calls. This objective could be achieved in two ways: either provide smaller $CW_{min}$ to the AP or provide larger $CW_{min}$ to wireless stations. In this paper, we take the latter approach, *i.e.*, fix the $CW_{min}$ for AP while scaling it for wireless stations in proportion to the VoIP call intensity. In terms of the implementation, the former is simpler. The AP can just adjust its $CW_{min}$ internally. For the latter, the AP must broadcast the $CW_{min}$ value, that the wireless stations should use in the beacon, which the 802.11e standard has as a feature [20]. The reason we take the latter approach is to control congestion. As we mentioned in Section III-B, the increase in the VoIP call volume increases both asymmetry and traffic load. Thus CWA should serve a double purpose, to adjust the bandwidth distribution between uplink and downlink, and to relieve the 802.11 link of the excessive collision probability arising

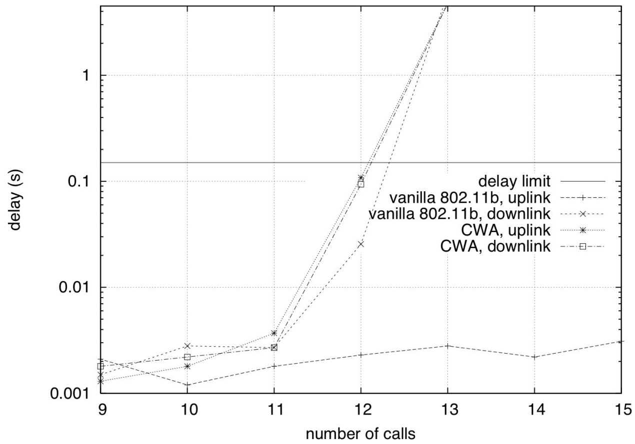



Fig. 4. Delay with CWA but without ZFA.

from the increased VoIP traffic. By adaptively configuring the $CW_{min}$ for the majority (*i.e.*, wireless stations) instead of a single station (*i.e.*, the AP) should have a more immediate impact on the contention level.

In our previous work, we demonstrated that the ratio of throughput can be controlled in proportion to the reciprocal of the given $CW_{min}$ ratio [25], [26]. In order to achieve up-down symmetry, therefore, we should set the $CW_{min}$ of the wireless stations as follows:

$$CW_{min}^{(W)} = CW_{min}^{def} \cdot \gamma \tag{3}$$

where $CW_{min}^{def}$ is the default minimum contention window size (*i.e.*, 31 in 802.11b) and $\gamma$ is the effective number of contending stations for uplink transmission. We use $\gamma$ instead of $n$, since the system is not saturated.

In this paper, we use a novel method using the AP's downlink queue length in order to estimate $\gamma$. Under the symmetric traffic and ZFA, we argue that $\gamma$ in the uplink should be roughly on the order of the number of calls to register voice frames in the AP's downlink queue. Under the same $CW_{min}$, $T_i$, $T_c$ and $T_s$ in (1) are all equal in uplink and downlink directions. by approximating the uplink delay and the downlink delay without the queue length term $E[Q^d]$, we only need to inflate the uplink terms (i.e., $T_i^u$, $T_s^u$, $T_c^u$, $T_o^u$) by a factor of $E[Q^d]$ to balance delays [32]. In other words, we set $\gamma = E[Q^d]$ where $E[Q^d]$ is directly observable inside the AP. In [32] we confirm that $\gamma$ is estimated this way and the AP downlink queue sizes given by the M/M/1//M model (and also simulation) are within 2.3% of each other through $N = 90$.

Using the $\gamma$ value obtained from the AP queue, as discussed above, we first apply CWA to the vanilla 802.11 queue, *without ZFA*. Fig. 4 shows the result. Surprisingly, CWA, used alone, hardly helps. In fact, there is no improvement whatsoever. The number of calls subject to $\bar{d}_{req}^{802.11} = 100ms$ is 12 with the vanilla 802.11b, and so it is with CWA.

This is an interesting result. It means that the asymmetry is not causing the bottleneck for the AP in the vanilla 802.11b link. With $N = 12$ calls, a voice frame transmission time of 835$\mu$s (Table I), and 20ms G.729 voice frames, we notice the 12 wireless stations collectively produce 600 voice frames per second in the uplink. Considering the matching downlink voice traffic from the AP, this is absolutely over the capacity of the 802.11 link. Namely, the utilization is

$$\rho = \frac{\lambda}{\mu} = 1200 \cdot 835.45 \times 10^{-6} = 1.0025.$$

Therefore, the 802.11b link delay begins to diverge before the asymmetry begins to cause problems (Note that the delay axis is in log scale.)

Recollect that what CWA does in this situation is to redistribute the delay between the uplink and downlink. However, in this case it is ineffective since the downlink delay goes unbounded, any redistribution attempt would only place both the uplink and downlink delays at practically unbounded values. In fact, Fig. 4 confirms that in CWA, the uplink delay indeed soars to a level comparable to that of the downlink delay. *This implies that the uplink-downlink delay redistribution with the capacity-bounded systems would not increase the number of acceptable calls.*

### C. Synergy between ZFA and CWA

In previous sections, we saw that in ZFA, the queuing delay for the downlink traffic is the performance-limiting factor. The resulting call capacity increase is limited to only two-fold, given $\bar{d}_{req}^{802.11} = 100ms$. Meanwhile, CWA alone does not improve call capacity at all, the vanilla 802.11 WLAN being the capacity-bounded system. In this section, it is demonstrated that combining these two components of V100 has a synergistic effect and achieves significantly higher increase in the call capacity given the same delay requirement. This improvement would not be possible when any scheme is used alone.

The synergy arises by allowing ZFA and CWA solve each other's problems, *i.e.*, *CWA curing the delay asymmetry of ZFA, and ZFA solving the capacity problem of CWA*. A particularly good feature of ZFA is that it not only makes the system delay-bounded, but also decreases the link utilization $\rho$, by reducing the number of MAC frames on the link. It gives CWA a lower $\rho$, so that CWA can effectively redistribute delay for ZFA, between the uplink and downlink, unlike in the capacity-bounded system, such as the vanilla 802.11 WLAN. In the act of reducing the delay of the downlink traffic significantly, in exchange for the slight increase in the uplink traffic delay, CWA can make the ZFA have an increased number of calls within the delay-bound.

Fig. 5 shows both by analysis and simulation the result of applying CWA (3), in addition to ZFA. From the figure, it is confirmed that CWA successfully equalizes the bandwidth in the uplink and downlink, hence the comparable delay.

In comparison with the ZFA alone, we notice that the downlink delay has been drastically reduced in ZFA+CWA, at the cost of increased delay in the uplink. According to the simulation, $\bar{d}_{req}^{802.11} = 100ms$ allows 90 calls, while the analysis overestimates it by a few calls. This is a significant jump from the ZFA result in Fig. 3, not to mention the 7-8 fold increase from the vanilla 802.11b system.

### D. Comparison with source-based frame aggregation

Unlike V100, most voice frame aggregation proposals for wireless links [9]–[11] do not reflect congestion condition

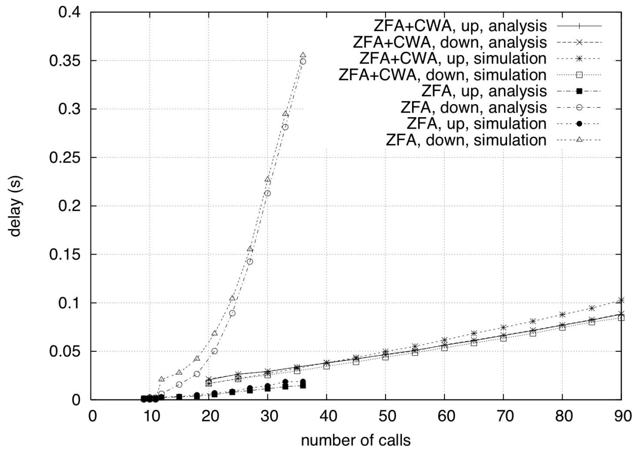



Fig. 5. Uplink/downlink delay under ZFA and ZFA+CWA.

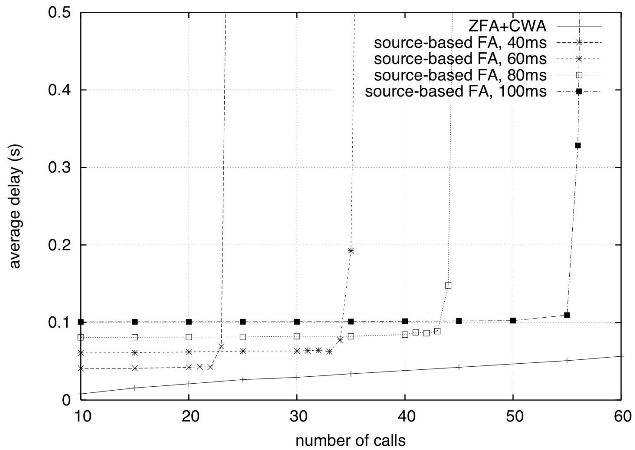



Fig. 6. Downlink delay, with 20ms voice payload for ZFA+CWA and 30, 50, and 80ms voice payload for source-based FA

to the level of aggregation. Rather, the voice samples are coalesced to a fixed sized length, to be shipped in the same RTP payload. This RTP payload elongation is usually achieved at the voice source, and is called the source-based frame aggregation (SFA), in this paper. The delay characteristics of ZFA+CWA are compared with that of these traditional SFA. Fig. 6 compares these as a function of the number of calls. We assume 20ms voice payload for ZFA+CWA and 40, 60, 80 and 100ms voice payload for the SFA.

As observed, the downlink delay dynamics of the SFA is completely different from ZFA+CWA, but quite similar to that seen in Fig. 4 for an unmodified 802.11 system. The only difference from the unmodified system is that the delay explosion points come at 23, 34, 45 and 55 calls for 40, 60, 80 and 100ms aggregation delay, respectively, instead of 12 calls. These thresholds are explained in exactly the same way as in Fig. 4.

Fig. 6 imparts a clear message. Without adaptation, the SFA becomes suboptimal in the majority of load regimes. When the contention level at the WLAN is low, it overshoots and the mouth-to-ear delay is unnecessarily large. In contrast, when the contention level is high, it can undershoot the required level of aggregation and cause the WLAN to lose stability, hence the unbounded delay. Although SFA could employ RTCP receiver reports (RRs) to find the delay and loss characteristics of the end-to-end conversation [33], the reports are not specific to the wireless link. In fact, they do not inform whether there is a wireless link in the end-to-end path, or if there is, what kind of wireless link. It is not clear how the RTCP-based adaptation for the SFA schemes can help resolve the bandwidth inefficiency issue addressed in this paper. In contrast, the V100 operation is directly linked with the congestion situation at the WLAN, and automatically determines the level of frame aggregation. This results in the resilience of the V100 against the load fluctuation in Fig. 6.

TABLE III
R-SCORE TO MOS MAPPING [34]

| R-score | Quality of voice rating | MOS |
|---|---|---|
| $90 < R < 100$ | Best | 4.34 – 4.5 |
| $80 < R < 90$ | High | 4.03 – 4.34 |
| $70 < R < 80$ | Medium | 3.60 – 4.03 |
| $60 < R < 70$ | Low | 3.10 – 3.60 |
| $50 < R < 60$ | Poor | 2.58 – 3.10 |

SFA does not specifically solve the bandwidth efficiency problem of VoIP traffic over the 802.11 link, however, it has an advantage over V100 in terms of implementation complexity, since it can be implemented with VoIP application, without requiring cross-layer processing or modification of the existing protocol implementations. Finally, we note that SFA is orthogonal to V100, and therefore they can be used together, subject to the combined mouth-to-ear delay budget.

## V. NUMBER AND QUALITY OF CALLS UNDER REALISTIC ASSUMPTIONS

This section evaluates the number of VoIP calls that V100 can accommodate on a 802.11b link. In addition to the *number* of the calls, we also consider the *quality* of the accommodated calls. First, we briefly discuss the call quality assessment methods.

### A. Call quality assessment

Various methods assessing speech quality are proposed [29], [34]–[36] by the ITU-T. P.563 [37] and P.862 (PESQ) [35] are analog-based, thus not immediately related with VoIP. Meanwhile, P.564 (P.VTQ) [38], completed in June 2006, is a new recommendation, specifically targeted for VoIP. However, what is the most commonly used assessment method for quantifying VoIP call quality in the literature is the E-model. The E-model defined in ITU-T G.107 recommendation [36] provides a way for assessing the Mean Opinion Score (MOS), which represents the satisfaction of the VoIP user from network behavior. The R-score derived from the E-model allows quantification of the voice quality, through network parameters such as the delay and loss [34] values. The function for measuring the voice quality is given by

$$
\begin{aligned}
R &= 94.2 - 0.024D - 0.11(D - 177.3)H(D - 177.3) \\
&\quad -11 - 40log(1 + 10L),
\end{aligned} \tag{4}
$$

where $D$ is the mouth-to-ear delay, $L$ is the total loss including the network and "delay loss," that is the arrival of the voice frame too late to meet the playout schedule, and $H$ is the Heaviside function. The R-score to MOS mapping is given in Table III. Although the new P.564 may be used more frequently than the E-model sometime in future, the E-model

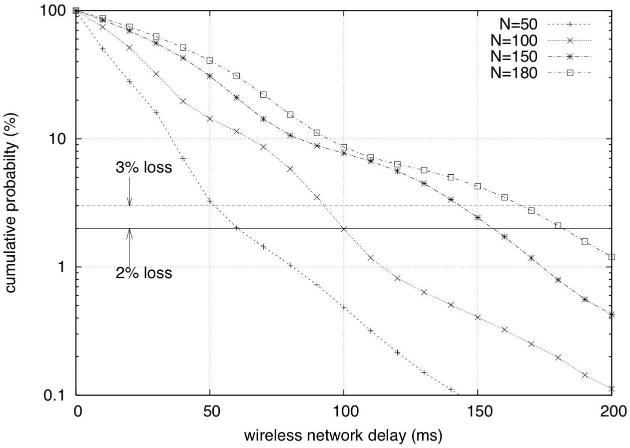



Fig. 7. Complementary CDF (CCDF) for aggregated G.729 frames under V100, $\alpha = 0.43$.

is selected because its popularity in the literature makes the evaluated quality numbers in our work easily comparable and understandable, with the exposure to similar works.

Outside the ITU-T standard context, K. Chen $et\ al.$ [29] proposes a novel index that quantifies the degree of VoIP user satisfaction. The proposed User Satisfaction Index (USI) is based on network QoS metrics, such as the bit rate, jitter and the round-trip time (RTT), and accuracy is validated through extensive Skype traffic experiments. In addition, [29] shows that users are relatively more tolerable to delay degradation than bandwidth degradation, and the delay bound in the ITU guideline is too strict. This result may justify our scheme better, since it is delay-bounded. However, in this study, more strict ITU G.108 requirement of maximum 250ms end-to-end delay [39] is followed, because using a large delay bound can have an unintended bias toward our scheme. Below, we assume that the sum of framing, lookahead, dejittering for G.729 codec and all other delays is bounded by 150ms, leaving at least 100ms for crossing the 802.11b link.

### B. Call capacity and quality evaluation

Prior to this section, we have maintained two unrealistic assumptions as to ZFA+CWA, mainly for ease of analysis and exposition. First, we considered the average delay requirement $\bar{d}_{req}^{802.11}$. Then, we assumed the voice activity factor $\alpha = 1$. In this section, we modify these assumptions to be more realistic, in order to estimate the number of possible calls with the ZFA+CWA on 802.11 networks, subject to the delay requirement set by the ITU-T recommendations.

First, telephone call traffic is an alternating series of talk spurts and silent periods. With silence suppression, or Conversely, voice activity detection (VAD), vocoders generate voice frames only in talk spurts. In particular, G.729 Annex-B and G.723.1 Annex-A include an integrated VAD function. Frequently, the voice traffic is modeled by Brady's ON-OFF model [40], where $\alpha = 0.43$. Accordingly, $\alpha = 0.43$ in this section.[1]

Second, although many prior works also compute the capacity based on the average delay [10], [11], what determines the maximum number of calls is the peak delay $\hat{d}_{req}^{802.11}$. Due to the variability caused by queuing and channel access time on the 802.11 link, as well as the statistical nature of the talk spurt occurrences from different calls, the peak delay could be much higher than average delay $\bar{d}_{req}^{802.11}$. Although for such variability, there is the de-jittering buffer at the receiver, if the jitter of a voice frame is so large that even the buffer cannot handle it, the frame cannot be used to reproduce voice and is simply thrown away. Such "delay losses" affect the call quality, therefore, we take account of such delayed frames in estimation of the number of sustainable calls.

[1]There are some recent systems, most notably Skype [6], that do not use VAD and thus $\alpha = 1$.

TABLE IV
QUALITY OF CALLS UNDER V100

| Calls | No FA | V100 | |
|---|---|---|---|
| | vocoder=25ms $\hat{d}_{req}^{802.11} = 100ms$ no jitter buffer | vocoder=25ms $\hat{d}_{req}^{802.11} = 100ms$ no jitter buffer | vocoder=25ms $\hat{d}_{req}^{802.11} = 100ms$ jitter buffer = 60ms |
| $N = 20$ | 80.0 | 80.9 | 79.5 |
| $N = 40$ | 5.7 | 80.0 | 77.7 |
| $N = 50$ | - | 78.7 | 77.6 |
| $N = 100$ | - | 71.6 | 76.4 |
| $N = 150$ | - | 57.1 | 70.6 |
| $N = 180$ | - | 53.1 | 64.1 |

Fig. 7 shows the Complementary Cumulative Distribution Function (CCDF) of the 802.11b delay experienced by the aggregated G.729 voice frames with a different number of calls under V100. According to an European Telecommunication Standards Institute (ETSI) experiment [41], the MOS rating for a G.729 call stays above 3.0 ("fair") if the loss rate is maintained below 2 to 3%. In the figure, we observe that with delay loss $L_d \leq 3\%$ and $\hat{d}_{req}^{802.11} \leq 100ms$, $N = 105$ calls can be supported, owing to the lowered $\alpha$. Namely, if we require 100ms peak delay bound, up to 3% of the voice frames are lost due to excessive jitter under 100 calls. With a more lenient requirement of $\hat{d}_{req}^{802.11} \leq 150ms$, however, the sustainable number surges to $N$ =153.

Based on the data from the experiment of Fig. 7, we show the R-scores in Table IV, with the delay requirement of $\hat{d}_{req}^{802.11} = 100ms$. In the table, we compare the vanilla 802.11 WLAN with V100, with two different configurations employed in the latter. The first is without a jitter buffer, and the second is with a 60ms jitter buffer. Notice that the use of the jitter buffer increases the mouth-to-ear delay, and has a negative impact on the R-score by increasing the $D$ term in (4). However, it helps salvage those voice frames with $100ms < D \leq 160ms$, that are otherwise abandoned as delay loss, and so it improves on $L$.

In Table IV, we notice that the vanilla 802.11b fails beyond 20 calls, whereas V100 maintains "Medium" call quality, up to 100 calls (without jitter buffer) or 150 calls (with jitter buffer). When the system is under low load, the use of the jitter buffer leads to the lower R-score, due to the increased mouth-to-ear delay. However, as the load increases (see $N = 150$ and $N = 180$) it helps maintain the R-score at the medium level, by salvaging voice frames with $100ms < D \leq 160ms$, before it degrades to "Low" quality at $N = 180$. In essence, *V100 not only increases the call capacity to over 100 on the 802.11b WLANs, but also maintains an acceptable level of*

TABLE V
PARAMETERS FOR COMPARISON

| Parameter | Scheme | | |
|---|---|---|---|
| | Wang *et al.* | Baldwin *et al.* | Medepalli *et al.* |
| Codec | G.729 | G.711 | G.711 |
| $\alpha$ | 1.0 | 0.43 | 0.48 |
| $L_d$ (%) | 1 | 10 | 2 |
| $\hat{d}_{req}$ (ms) | 30 | 100 | 100 |
| Source aggregation delay (ms) | N/A | N/A | 50 |

*call quality*.[2]

Fig. 8 compares the number of calls under V100 with those under Wang *et al.* [8], Medepalli *et al.* [9], and Baldwin *et al.* [17]. These schemes were picked from the category of inter-call FA, SFA, and MAC modification, respectively. Since they used different parameters (Table V), we change the parameters for V100 accordingly in each comparison. For instance, for comparison with Baldwin *et al.*, we changed the V100 to use the G.711 codec, $\alpha = 0.43$, $L_d = 10\%$, and $\hat{d}_{req}^{802.11} = 100ms$. Notice in Fig. 8 that the number of calls under V100 is over 100, even with G.711 due to the magnitude of tolerated loss $L_d = 10\%$. Medepalli *et al.* uses an extra 50ms for RTP payload increase at the source, its performance is compared with V100, with only 100ms peak delay requirement, which is more stringent.

Like the SFA schemes, the inter-call frame aggregation is also *capacity-bounded*. Accordingly, the Wang *et al.* [8] does not scale simply because a larger delay budget $\hat{d}_{req}^{802.11}$ is given. The number of G.729 calls that can be maximally supported under this scheme with $\alpha = 1$, $\hat{d}_{req}^{802.11} = 30ms$ and $L_d = 0.01$ is $N = 21.7$ [8]. It is natural that the scheme is bounded at 21.7, which is roughly twice the bound, 12, for non-aggregated voice traffic (see Fig. 4). With one-way (downlink) aggregation, the improvement is roughly $\frac{2k}{k+1} \approx 2$ since $2k$ packets are reduced to $k + 1$ packets, *i.e.*, $k$ uplink packets and 1 aggregated downlink packet. When compared with this scheme, V100 supports only 15 calls with 30ms delay budget, which is 6 calls below. However, V100 is *delay-bounded* and scales with the delay budget. It can therefore outperform Wang *et al.* with larger delay budget, as shown in Fig. 8. Finally, the reason that V100 falls below 100 calls is due to the increased voice activity factor of $\alpha = 1$ in the comparison setting.

## VI. CONCLUSION

In this paper, we investigate a scheme called V100, to overcome the poor bandwidth efficiency that VoIP calls suffer from on the IEEE 802.11 wireless LANs. The V100 scheme performs adaptive intra-call frame aggregation in the wireless interface queue, and controls the contention window size ratio between the AP and the wireless stations. The former reduces the absolute number of MAC frames on the link, without adversely affecting the end-to-end delay of the call, while the latter allows the AP to accommodate more calls. We show that the combined use of the two mechanisms can easily increase call capacity to over 100 calls, under 100ms peak delay budget, although the proposed system has the delay-bounded property that it can accommodate a greater number of calls, given a larger call delay budget. These mechanisms require little or no changes in incumbent protocols, such as RTP, UDP, and IP. The changes are required in the MAC-layer implementation of the interface queue, and the dynamic control of the contention window, which is supported in the 802.11e extension. In terms of the quality, V100 has an R-score greater than 70, showing that it not only boosts the call capacity of the IEEE 802.11 wireless LANs, but also maintains the quality of the accommodated calls. Although explained with 802.11b, the proposed idea can be applied to any 802.11 network that needs to achieve a maximal number of VoIP calls within the given bandwidth allocation.

[2]The R-score does not reflect other delay components such as wireline network delay, so the real R-score can be lower than that shown in Table IV. However, we exclude those extra delay components from computation, mainly for fairness, as the vast majority of other related work does not consider the non-wireless delays.

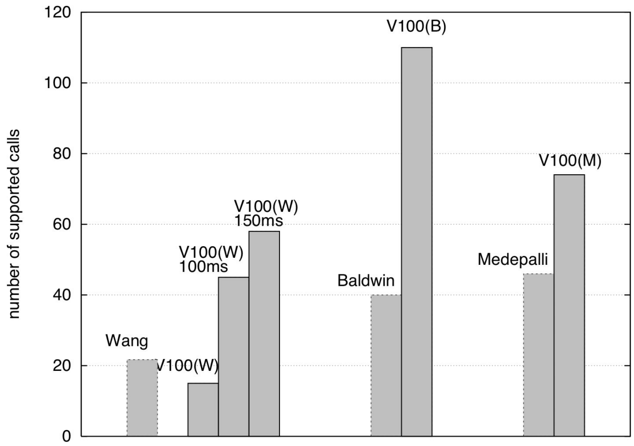



Fig. 8. Call capacity comparison of different approaches.

## ACKNOWLEDGEMENT

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. R01-2006-000-10510-0) and the Ubiquitous Autonomic Computing and Network Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

## REFERENCES

[1] ANSI/IEEE, "802.11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," June 1997.

[2] ANSI/IEEE, "IEEE 802.11b-1999: Supplement to 802.11-1999,Wireless LAN MAC and PHY specifications: Higher speed Physical Layer (PHY) extension in the 2.4 GHz band," Sept. 1999.

[3] ANSI/IEEE, "IEEE 802.11a-1999: Amendment 1: High-speed physical layer in the 5 GHz band," Nov. 1999.

[4] ANSI/IEEE, "IEEE Std 802.11g-2003: Amendment 4: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band," June 2003.

[5] IEEE P802.11n/D1.0, "Amendment: Medium Access Control (MAC) and Physical Layer (PHY) specifications, enhancement for higher throughput," March 2006.

[6] Skype, http://www.skype.com.

[7] http://www.skypejournal.com/blog/archives/2005/05/3_million_skype_1.php

[8] 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*, vol. 54, no. 1, pp. 366 - 384, Jan. 2005.

[9] K. Medepalli, P. Gopalakrishnan, D. Famolari and T. Kodama, "Voice capacity of IEEE 802.11b, 802.11a, and 802.11g wireless LANs," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, pp. 1549 - 1553, Nov. 2004.

[10] S. Garg and M. Kappes, "Can I add a VoIP call?," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 779 - 783, May 2003.

[11] D. Hole and F. Tobagi, "Capacity of an IEEE 802.11b wireless LAN supporting VoIP," in *Proc. IEEE International Conference on Communications (ICC)*, pp. 196 - 201, June 2004.

[12] Coding of speech at 8kbit/s using conjugate-structure algebraic-odeexcited-linear-prediction, ITU-T Recommendation G.729, Mar. 1996.

[13] S. Casner and V. Jacobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links", RFC 2508, Feb. 1999.

[14] L.-E. Johnson and G. Pelletier, "Robust Header Compression (ROHC): A Link Layer Assisted Profile for IP/UDP/RTP," RFC 3242, Apr. 2002.

[15] H. Kim and I. Kang, "Measurement-based Frame Grouping in Internet Telephony, *IEE Electronics Letters*, vol. 37, no. 1, pp. 71 - 72, Jan. 2001.

[16] H. Kim, I. Kang, and E. Hwang, "Measurement-based multi-call voice frame grouping in Internet telephony," *IEEE Communications Letters*, vol 6, no. 5, pp. 199 - 201, May 2002.

[17] R. Baldwin, N. Davis IV, S. Midkiff, and R. Raines, "Packetized voice transmission using RT-MAC, a wireless real-time medium access control protocol," *Mobile Computing and Communications Review*, vol. 5, no. 3, pp. 11 - 25, July 2001.

[18] T. Hiraguri, T. Ichikawa, M. Iizuka, and M. Morikura, "Novel multiple access protocol for voice over IP in wireless LAN," *IEICE Transactions on Communications*, vol. 85, no. 10, pp. 517- 523, Oct. 2002.

[19] A. Banchs, X. Perez, M. Radimirsch, and H. Stuttgen, "Service differentiation extensions for elastic and real-time traffic in 802.11 wireless LAN," in *Proc. IEEE Workshop of High Performance Switching and Routing*, pp. 245 - 249, May 2001.

[20] ANSI/IEEE, "802.11e: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Medium Access Control (MAC) Enhancements for Quality of Serivce (QoS)," Nov. 2002.

[21] R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, "Adaptive playout mechanisms for packetized audio applications in wide-area networks," in *Proc. Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pp. 680 - 688, June 1994.

[22] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Dudda, "Performance Anomaly of 802.11," in *Proc. Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pp. 836 - 843, March 2003.

[23] D. Leith, "Experimental evaluation of TCP performance and fairness in an 802.11e testbed," in *Proc. the 2005 ACM SIGCOMM workshop on Experimental approaches to wireless network design and analysis*, pp. 17 - 22, Aug. 2005.

[24] A. Grilo, and M. Nunes, "Performance evaluation of IEEE 802.11e," in *Proc. IEEE International Symposium of Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 511 - 517, Sept. 2002.

[25] Y. Yang, J. Wang, and R. Kravets, "Distributed optimal contention window control for elastic traffic in wireless LANs," in *Proc. Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, pp. 35 - 46, March 2005.

[26] H. Kim, S. Yun, and I. Kang, "Resovling 802.11 performance anomaly through QoS differentiation," *IEEE Communications Letters*, vol. 9, no. 7, pp. 655 - 657, July 2005.

[27] H. Kim, B. Roh, S. Yoo, "Online RTP packet classificantion for real-time multimedia traffic management in the Internet," in *Proc. International Conference on Information Systems (ICIS)*, pp. 371-378, July 2002.

[28] Y. Kim, S. Choi, K. Jang, and H. Hwang, "Throughtput enhancement of IEEE 802.11 WLAN via frame aggregation," in *Proc. of IEEE Vehicular Technology Conference (VTC)-Fall*, pp. 3030 - 3034, Sept. 2004.

[29] Kuan-Ta Chen, Chun-Ying Huang, Polly Huang and Chin-Laung Lei, "Quantifying Skype User Satisfaction,"in *Proc. ACM SIGCOMM*, pp. 399 - 410, Sept. 2006.

[30] ns-2, http://www.isi.edu/nsnam/ns.

[31] L. Kleinrock, *Queueing Systems, V.1: Theory*, pp. 106-107, 1975.

[32] S. Yun and H. Kim, "100+ VoIP calls on 802.11b: the power of synergistic voice frame aggregation in wireless LANs," technical report, Korea University, Available at http://widen.korea.ac.kr/V100.pdf, Jan. 2006.

[33] Xiaokun Yu. (2006). "Adaptive Wireless Multimedia Services, ", Master's Thesis, KTH Information and Communication Technology, Stockholm, Sweden, May 2006.

[34] R. G. Cole and J. H. Rosenbluth, "Voice over IP performance monitoring," *ACM Computer Communication Review*, vol. 31, pp. 9 - 24, April 2001.

[35] An objective method of end-to-end speech quality assessment of narrow-band telephone networks and speech codes, ITU-T recommendation P.862, Jan. 2001.

[36] The E-model, a computational model for use in transmission planning, ITU-T Recommendation G.107, Dec. 1998.

[37] Single-ended method for objective speech quality assessment in narrow-band telephony applications, ITU-T Recommendation P.563, May 2004.

[38] Conformance Testing for Narrowband Voice over IP Quality Assessment Models, ITU-T Recommendation P.564, June 2006.

[39] Application of the E-model: A planning guide, ITU-T Recommendation G.108, Sept. 1999.

[40] P. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation," *Bell Syst. Tech. Journal*, vol. 48, no. 7, pp. 2445-2472, Sept. 1969.

[41] *Actual measurements of network and station characteristics and performance parameters in TIPHON networks and their influence on voice quality*, ETSI TR 101 329-6 V2.1.1, Feb. 2002.


**Sangki Yun** received his B.S. degree from Kwangwoon University in 2004, and the M.S. degree from Korea University in 2006. Currently he is a Ph. D. student in Korea University. His research interests include wireless networking, 3G/4G networks and peer-to-peer network.

**Hyogon Kim** is an associate professor at the Korea University. He got his Ph.D. from the University of Pennsylvania in 1995. Prior to joining the Korea University, he was a research scientist at Bell Communications Research (Bellcore). His research interests include Internet protocols and applications, wireless networking, network security, and computational neuroscience.

**Heejo Lee** received his B.S., M.S., Ph.D. in Computer Science and Engineering from Pohang University of Science and Technology (POSTECH), Korea in 1993, 1995 and 2000, respectively. Since 2004, he has been an assistant professor at the Department of Computer Science and Engineering, Korea University, Seoul, Korea. From 2001 to 2003, he was at Ahnlab, Inc. as Chief Technology Officer (CTO) and Director of Technology Planning Department. From 2000 to 2001, he was a Post Doctoral Research Associate at the Network Systems Lab of the Department of Computer Sciences, and at the Center for Education and Research in Information Assurance and Security (CERIAS), Purdue University. His research interest includes computer and communication security, parallel scientific computing, and fault-tolerant computing.

**Inhye Kang** is an associate professor at the University of Seoul. She got her Ph.D. from the University of Pennsylvania in 1996. She joined the University of Seoul in 1999, before which she was a senior researcher at Samsung SECUi.com. Her research interests include formal methods, system verification, and computer networks.