Minimising *Perceived* Latency in Audio-Conferencing Systems over Application-Level Multicast

Nick Blundell and Laurent Mathy

Lancaster University, UK
{n.blundell, laurent}@comp.lancs.ac.uk

Abstract. In this paper, we propose a scalable and dynamically-adapting application-level multicast (ALM) routing protocol, designed specifically for audio-conferencing systems over the Internet.
Currently proposed ALM protocols try to optimise delay for the whole group of participating nodes during construction and maintenance of an overlay network which, when using standard packet flooding, can result in a number of the participants experiencing unacceptably-high latencies, unsuitable for real-time audio communication; whereas we propose to dynamically prioritise routing for those participants who are currently *in* conversation (i.e. those who require the lowest latencies in order to react to conversational cues) and allow higher latencies for participants who simply listen to the conversation without taking an active part in it in that particular moment in time.
Thus, we aim to provide low *perceived* latency for *all* of the audio-conference participants without any support from the underlying network.

1 Introduction

As a result of improvements in computer hardware, research into voice-over-IP (VoIP) technologies, and increased network bandwidth available to the home, point-to-point audio communication can more-or-less be achieved over the Internet, provided that network delay and packet loss do not exceed their tolerable thresholds (see section 2.1).

Group audio communication, on the other hand, has proven more challenging in the way of deployment, scalability, and of communication-channel quality: network-level multicast was proposed over a decade ago [6] as a solution for efficient, large-scale group communication over the Internet, but wide-scale deployment of the service has since been hampered due to various technical and administrative issues that surround it [4]. In response to the lack of a group-communication service in the network, various application-level techniques have been proposed [7].

In application-level multicast (ALM), many-to-many communication is achieved through using overlay trees in one of three ways: (1) a sender floods data to their

subtree through their children and also sends data to the tree root, who, in turn, floods the data to the rest of the group through its children; (2) a sender floods data to the group through their children and through their tree parent such that data flow is bi-directional on tree links [2]; or (3) several trees are built, rooted at each source, allowing data to be flooded to the whole group by each source as does a single source in one-to-many communication [4].

ALM has been proposed as a solution for audio conferencing [4], however, as an ALM group grows in size there is, inevitably, an increased imbalance in the degrees of latency (end-to-end delay) experienced by different, communicating node pairs within the group such that, and with regard to studies into user tolerance of latency in audio systems [9], a significant number of participant pairs will begin to experience difficulty in communicating with each other due to excessive latency in the audio channel.

In this paper, we consider a dynamic ALM-routing approach over standard flooding as a way to minimise the *perceived* latency experienced by audio-conference participants, drawing from patterns in conversation and from a user's perception of audio-channel quality.

The remainder of this paper is organised as follows: firstly, in section 2, we describe related work which has influenced our design rationale; next, in section 3, we present, in a preliminary study, our own observations of turn-taking in actual samples of conversation, before, in section 4, describing the proposed application-level network audio-conferencing routing protocol (ALNAC). In section 5, we present, through simulation, the effects of the proposed routing protocol on the group and the underlying network. Finally, in section 6 we give concluding remarks on the paper and describe future directions of the presented work.

2 Related Work

In this section, we describe two areas of particular relevance to the proposed work, namely: Internet packet-audio transmission and conversation analysis — the study of conversation.

2.1 Internet Packet-Audio Transmission

In packet-switched networks, audio transmissions are typically subjected to several latency components: sampling, packetisation, pre-processing (silence-suppression and compression), network transmission, network propagation, un-compression, and finally, playout buffering; with network-propagation delay being typically the least-predictable and most-dominant component for audio transmission over the Internet.

An abundance of studies into user tolerance of round-trip latency in audio-communication systems has been conducted and generally agrees upon the following levels of tolerance: excellent, 0–300 ms; good, 300–600 ms; poor, 600–700

ms; and quality becomes unacceptable for round-trip latencies in excess of 700 ms [9].

As latency increases, it is miss-interpreted by the user as extended pause in speech, causing confusion when they fail to get immediate responses from the other user(s); this, in turn, results in their eventual loss of synchronisation with the conversation.

2.2 Conversation Analysis

Conversation analysis is the study of verbal communication between people, with an emphasis on how that communication is structured and on how it is affected by social or cultural settings [10].

An area of conversation analysis of particular relevance to this work is the study of *turn taking*: the basic form of organisation in conversation. In conversation, people naturally organise their spoken contributions (utterances) into turns, where each person silently waits, listening to the current speaker, for their turn to speak [10]. It is also worth noting that overlapping speech occurs rarely in conversation, since one person must remain silent to effectively listen to what another person is saying.

A person will typically wait for a duration of time after the current speaker becomes silent before taking their turn to speak: typically, the pause is one second for Anglo-Saxon English speakers [11].

A large part of verbal communication, among any number of participants, consists of turns that are somehow related to each other, known as *adjacency pairs* [10].

This organisation of conversation turns into adjacency pairs naturally leads to a degree of localisation, where, over a given interval of time, a small proportion of the participants present exchange turns with one another. Figure 1 gives an example of everyday conversation, illustrating localisation of interest through nested adjacency pairs. This localisation property forms the basis of our work.

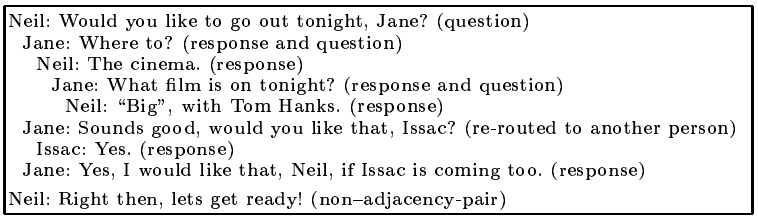


Fig. 1. An example of nested adjacency pairs, leading to localisation of interest in natural conversation.

Table 1. Accuracy of Next-Speaker Prediction when Considering a Backlog of n Distinct Speakers.

Back-Log Size	1	2	3	4	5	6	7
Accurate Predictions (%)	58	73	82	88	92	94	95

3 Preliminary Study of Next-Speaker Predictability

This section presents a preliminary study of patterns in samples of actual conversation, helping to support our design rationale in the following section.

To examine the extent of turn localisation (see section 2.2) that occurs in conversation, we performed a simple analysis of two audio-conference trace files and two public-meeting transcripts.

The trace files, which contained time stamps of talk spurts produced by participants, were logged from a locally developed audio-conference system using multiple-unicast transmission over a LAN. One audio-conference session included ten players of an online game which lasted for a duration of twenty minutes, and the other session was an informal discussion among eighteen people which lasted forty minutes.

The public meetings whose transcripts we analysed included 38 [5] and 42 [1] speaking participants respectively.

By stepping through the participant turns in each of the trace files and transcript files, we calculated the probability that the next speaker was amongst the set of n previous, distinct speakers, for various values of n. These probabilities, which represent the accuracy with which the next speaker can be predicted, are presented in table 1.

These results show two interesting properties that suggest turn localisation can be exploited in audio-conferencing systems: firstly, in more than half the cases, it is the previous speaker who answers the current speaker; secondly, the improvement in prediction accuracy quickly diminishes as more previous speakers are considered.

4 Dynamic Overlay-Routing Protocol Design

The proposed design for the application-level network audio-conferencing routing protocol (ALNAC) is derived from our understanding both of a user's perception of audio-channel quality (see section 2.1), and of turn-taking patterns in conversation (see section 2.2).

From this understanding, we hypothesize that, if a pair of overlay nodes are joined with minimal overlay distance when their respective participants are most sensitive to high latency (when engaged in conversation with each other) but are allowed to become further apart, with respect to overlay distance, when they are least sensitive to latency (when passively listening), then their perception will remain that of minimal network latency.

An ALM routing strategy that adapts in such a way will effectively reduce participants' sensitivity to scaling of the overlay network provided that, (1) the unicast distance between participant nodes is not excessively high, and (2) latency adaptation of the adaptive-routing protocol is sufficiently responsive to their changing state of interaction.

Thus, the general strategy of ALNAC is to enable direct delivery (i.e. in unicast trip time) of audio packets from a transmitting node to the nodes hosting participants who will most likely take up the next conversation turn, whilst ensuring overlay scalability, potentially to a large number of audio-conference participants.

Our analysis of next-speaker prediction (see section 3) shows that the participants most likely to speak next are those who have already spoken in the recent past.

4.1 Adaptive Routing of Audio Packets

We describe ALNAC in terms of a shared-tree overlay network and choose, for the basic-protocol operation, to adopt uni-directional tree routing (i.e. where a transmitting node floods data packets through its children and supplies the rest of the tree via the tree root), since uni-directional routing ensures lower maximum hop counts on delivery paths than does bi-directional routing. It then follows, from describing shared-tree routing, that ALNAC can also be applied to per-source trees by considering the transmitting node to be the tree root.

An ALNAC node will obtain a set of child nodes and a parent node from the (non-specific) overlay-tree construction protocol used to build the shared tree, and, where available, to optimise routing, will make use of inter-node distance information collected and exposed by the overlay-construction protocol.

On transmitting an audio packet, an ALNAC node will send a copy of the packet to the tree's root node, but, rather than simply flooding the packet to its children — as performed in standard tree flooding — the transmitting node will choose to directly send the packet to a set of *target nodes* which will include a number of nodes hosting recently-speaking participants and a number (zero or more) of the transmitting node's children. Section 4.2 describes the process of selecting target nodes.

Note that, to avoid inter–talk-spurt jitter, we choose to update the target-node set between and not during talk spurts of the transmitting node: for example, if a participant is speaking, and at that same time begins to receive packets from a new speaker — not currently in the target-node set — the original speaker will defer inclusion of the newly-detected speaker in the target-node set until the start of the original speaker's next talk spurt.

Typically, resource constraints — network or application — impose a limit (a.k.a out-degree) on the number of nodes to whom an overlay node may simultaneously send a data packet. Therefore, by including recent speakers in its target-node set, the current speaker, of constrained out-degree, must deprive some, or all, of its children from receiving audio packets directly from itself;

we therefore propose that a node may temporarily delegate responsibility for supplying its deprived children to some target nodes.

To ensure that ALNAC is sufficiently responsive to the changing set of recent speakers, whilst ensuring consistency in tree routing state, the ALNAC header of transmitted audio packets may contain a *delegation chain*. The delegation chain is simply a list of delegated-node addresses, composed in such a way that it can be split up into smaller delegation chains for efficient re-delegation of delegated nodes when only a subset of the nodes may be supplied (see section 4.3).

On receiving an audio packet containing a delegation chain, a node will try to supply as many (deprived) nodes from the delegation chain as its out-degree limit will allow — splitting the chain as necessary — and may also deprive some of its own children in a similar manner to a speaking node when supplying recently-speaking participant nodes. Note that, in order to minimise the number of delegations per delegate, the receiving node will deprive no more than half of its own children to supply nodes in the incoming delegation chain.

To avoid loops and duplications in the overlay network, each audio-packet header will contain the address of nodes that were supplied on non-tree links by the transmitting node. Thus, the parent of a supplied node will choose simply not to forward an audio packet to a child that is indicated, in the header, to have been already supplied.

Figure 2 illustrates adaptive routing of ALNAC, where node S, with an out-degree limit of five, transmits an audio packet directly to three recently-speaking participant nodes, A_1, A_2, and A_3, and to the tree-root node, R. S also sends the audio packet to one of its children, D, nominating D as the most suitable delegate for S's three other children that have been deprived of receiving the packet directly from S. Consequently, D deprives two of its own children, allowing it to supply directly two of its siblings, and the third one, indirectly, through a 'chained' re-delegation (see section 4.3).

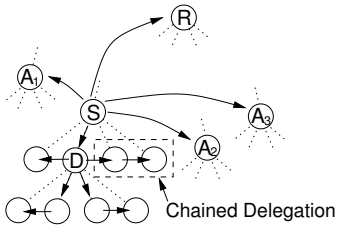


Fig. 2. Adaptive routing of ALNAC to three recently-speaking participant nodes.

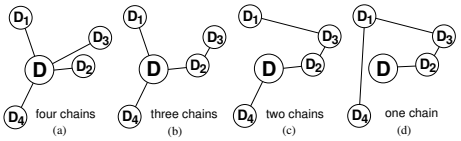


Fig. 3. Construction of an efficient delegation chain through all of the delegated nodes, D_1, D_2, D_3, and D_4, assigned to the delegate node, D.

4.2 Target-Node Selection

The number of selected target nodes is ultimately constrained by the out-degree of the transmitting node. There is, therefore, a trade-off between the number of recently-speaking participants who can be directly sent the audio frame and the extent of tree-route disruption caused by child-node deprivation and consequential delegation.

We propose a simple optimisation control for this trade-off by enforcing that a minimum number, n, of the transmitting node's children are always included in the target-node set; where the value for n can be chosen between zero and the transmitting-node's maximum fanout ($out\text{-}degree-1$) — at which value, routing will become standard tree flooding.

Where target-node selection allows the inclusion of a subset of children, the central-most children will be selected as approximations to the most suitable delegates.

4.3 Delegation

A transmitting node will nominate a subset of its target-node set as delegate nodes: if distance information is available among a subset of the target nodes, they will be nominated as delegates; otherwise, where no distance information is available, *all* of the target nodes will be nominated as delegates in favor of a balanced number of delegations per delegate.

Once a transmitting node has determined its sets of target nodes, delegate nodes, and deprived children, the transmitting node will attempt to assign each deprived child to its closest nominated delegate node, whilst ensuring a balanced distribution of delegated nodes per delegate node. If no distance information is available, the assignment will be made arbitrarily, based on the difference between network addresses.

As an optimisation to the delegation process, where inter-node distance information is available between a delegate node and each of its assigned delegated nodes, a transmitting node will give each delegate sufficient information in its delegation chain to allow efficient re-delegation should a delegate be assigned more nodes than it is able to supply.

The transmitting node will calculate delegation chains for a delegate node using the following algorithm:

- start with optimal chains (paths) from the delegate to each of its assigned delegated nodes (see figure 3(a))
- merge the two chains that produce the shortest path, producing $n-1$ chains (see figure 3(b))
- store the address of the first node in the chain that was joined to the appended chain, such that the chain can later be separated again at that point.
- continue, in the same manner, to merge the chains and to store the address of the joining node, until a single chain remains that spans all of the delegated nodes (see figure 3(c)–(d))

The result of this process is that we have a single, ordered chain from the delegate node through each delegated node. Thus, when a delegate node receives this information in the audio packet, it is able to break the chain into as many chain fragments as it is able to directly serve, efficiently re-delegating any remaining nodes. The reader should note that target nodes receive different delegation chains, comprising the deprived children assigned to them by the transmitting node.

4.4 Extended Protocol Operation

We have described, for the basic protocol operation, how we can avoid duplication in the overlay network when adopting a uni-directional routing approach, however, since a transmitting node may send a data packet directly to any node in the group — a recently-speaking participant node — there is an opportunity to achieve more uniform overall overlay latency by allowing a receiving node to forward a packet up to their parent.

We therefore define a protocol parameter to toggle a receiving node's ability to forward a packet up the tree if the packet is not from, or indicated (in the packet header) to have been already supplied to, its parent node. The effectiveness of such a bi-directional routing technique for reducing overall latency has been shown through random packet jumping in the context of resilient multicast over ALM [3].

5 Simulation Experiments

In this section, we analyse, through simulation, the performance of the ALNAC routing protocol in comparison to non-adaptive overlay-tree flooding typically used in ALM group communications.

5.1 Performance Measurements

Since ALNAC is designed to adapt latency in the audio-communication channel to the conversational patterns of audio-conference participants we choose to measure delay of received audio packets against a participant node's elapsed time since audio-packet transmission.

In addition to examining adaptation, we examine the impact that such adaptation has on the overlay network as a whole and on the underlying network by observing group delay characteristics and network stress (the extent of packet replication on network links).

5.2 ALNAC Protocol Parameters

In the proposed ALNAC design we left the following two parameters open for experimentation: the minimum number of children (MC) that a transmitting node is forced to include in the target-node set; and a flag (UP), enabling a node to forward a received packet up the tree if its parent has not been already supplied.

5.3 Simulation Set-Up

We simulated the ALM protocols using a locally-developed, packet-level network simulator, which implemented shortest-path routing over a 600-node, Internet-like GT-ITM[13]–generated core-network topology.

To simulate the presence of audio-conference client nodes in different network domains, we connected, with one-millisecond links, a further 100 nodes to random nodes at the core-network's edge; upon these client nodes, we ran the various ALM routing protocols.

As a benchmark for the best-attainable delay for all client nodes, akin to the delay attainable through network-level multicast, we implemented a naïve multiple-unicast protocol client.

For an overlay-tree construction protocol we implemented the Tree Building Control Protocol [8] (TBCP): a scalable, tree-first ALM protocol for building efficient, cost-constrained overlay trees. Note that, to further optimise TBCP for latency, we used the TBCP score function proposed in [12]. In the experiments, we fixed the maximum fanout of the overlay tree at five in consideration of turn-prediction backlog sizes (see section 3).

Over the overlay tree we ran both standard uni-directional flooding, representative of the non-adaptive routing typically used in ALM, and ALNAC using various parameter values.

To test adaptation of ALNAC under realistic conditions we instructed simulation clients to reproduce traffic from packet-trace files which were generated by a simple, unicast-based audio conference with 18 speaking participants that ran on the university LAN; also, for testing the impact of ALNAC whilst heavily exercised, we synthesised audio-conference traffic to produce random client interaction (i.e. periodically a client node, picked at random, was instructed to transmit an audio packet).

Thus, each simulation ran as follows: initially, 100 ALM clients were instructed to join the overlay network over a period of ten minutes in a random, uniformly-distributed fashion; then, the clients proceeded to transmit traffic in their designated patterns.

5.4 Simulation Results

The results in this section, presented for each protocol, are averaged over ten simulations using different client-node topology placement seeds.

Figures 4(a) and 4(b) show the average and maximum delay experienced by client nodes against elapsed time since they last spoke.

As expected, given the adaptive nature of ALNAC, figure 4(a) shows, clearly, a reduction in delay for recently-speaking nodes over that experienced when using standard tree flooding. We see, with one exception, a decreased delay when ALNAC enforced fewer child targets on nodes, since more of the recently-speaking nodes could be directly supplied; the exception occurred when no child targets were enforced (MC = 0) and when upward routing was disabled (UP

= 0), since nodes were not being forced to use any efficiently-constructed tree routes when transmitting packets.

We see from figure 4(b) how ALNAC maintained unicast delay for a minimum elapsed time-since-transmission of 7 seconds for all client nodes that had recently spoken; this time increased to 11 seconds when fewer child targets were enforced by ALNAC since more recently-speaking nodes could be accommodated by transmitting nodes. The best maximum delay achieved for recently-speaking nodes occurred when no child targets were enforced and upward routing was enabled since audio packets were being broadly disseminated over the tree by transmitting nodes.

Figures 5(a) and 5(b) show the cumulative distribution functions (CDFs) of the average and maximum delay experienced by client nodes. The simulation traffic was synthesised so as to fully exercise the ALNAC protocol with frequent and random client interaction.

We see from figure 5(a) that, in general, ALNAC improved average delay of the group over standard flooding; this occurred as a result of both packet flow being no longer bound only to overlay-tree paths and leaf nodes of the tree becoming more useful for disseminating packets to the group; the improvement was even greater when bi-directional routing was enabled.

Interestingly, the enforcement of child targets on transmitting nodes had little effect on the average latency of the group; this is an effect of dilution, since such an enforcement has no effect on the leaf nodes (with no children), which made up the majority of nodes in the tree.

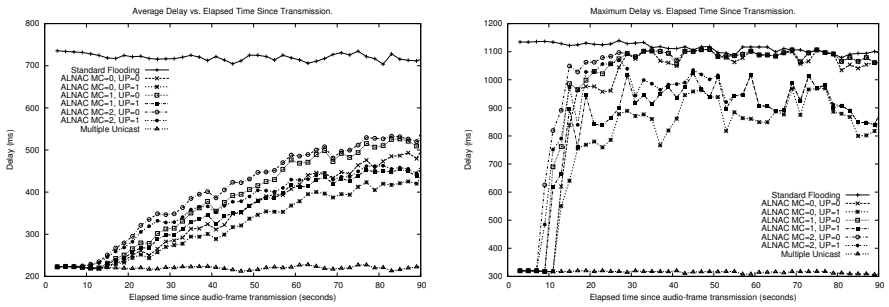
In figure 5(b), we see that the maximum delay experienced by nodes when using ALNAC was slightly worse than that experienced when using standard flooding. Maximum delay increased slightly with fewer enforced target children, but the combination of both broad packet dissemination, by transmitting nodes, and efficient delegations that made use of local inter-node distance information, lessened the impact of ALNAC on the group as a whole.

Table 5.4 shows the maximum stress placed on network links by the routing protocols in simulations using synthesised, high-interaction audio-conference traffic. In general, the network incurred higher stress from ALNAC than from standard flooding; this was the result of overlay-tree circumvention by ALNAC when routing to recently-speaking nodes, since the tree-construction protocol (TBCP), by design, tries to minimise network stress through clustering overlay nodes that are topologically close together.

In summary, ALNAC effectively reduced delay for recently-speaking nodes whilst minimising the impact on the whole group, but did so at the cost of slightly increased stress on network links.

6 Conclusions

In this paper, we have justified the case, and presented a design, for an ALM routing protocol that can dynamically adapt audio-packet routing to changing patterns in application usage.



(a) (b)

Fig. 4. Maximum and average delay experienced by recently-speaking participant nodes.

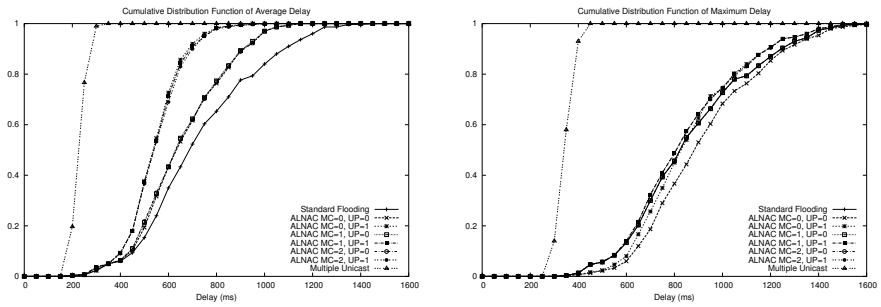
Table 2. Maximum network stress of routing protocols.

Routing Protocol	Maximum Stress
Standard Flooding	11
ALNAC (MC=0,UP=0)	15
ALNAC (MC=0,UP=1)	16
ALNAC (MC=1,UP=0)	15
ALNAC (MC=1,UP=1)	16
ALNAC (MC=2,UP=0)	15
ALNAC (MC=2,UP=1)	15

The approach we have taken essentially fuses the benefits — latency and scalability — of multiple-unicast and ALM group-communication techniques through consideration of patterns observed in conversation and through an understanding of a user's perception of audio-channel quality.

As further work, we plan to implement ALNAC within an ALM proxy client, creating a group-communication service for existing audio-conferencing applications, with which we will conduct subjective user trials. Such trials will help us to better understand the limitations of ALNAC: for example, as the size of a group increases, at what point does competition for conversation turns become so unfair for those participants who have not spoken in the recent past that they are unable to effectively communicate in the group.

Another interesting study would be to see how, say, a tree-first protocol using ALNAC would compare, subjectively, to a mesh-first protocol using standard flooding, where mesh-first protocols produce more-optimal trees though generally require more control overhead than the former.



(a) (b)

Fig. 5. Cumulative distribution of delay experienced by participant nodes.

References

1. Athelstan. Transcript of a Forty-Two–Member Meeting in the Corpus of Spoken Professional American English (CSPAE). `http://www.athel.com/sample.html`.
2. S. Banerjee, B. Bhattacharjee, and C. Kommareddy. Scalable Application Layer Multicast. In *ACM SIGCOMM*, Aug 2002.
3. Suman Banerjee, Seungjoon Lee, Bobby Bhattacharjee, and Aravind Srinivasan. Resilient Multicast using Overlays. In *Proceedings ACM Sigmetrics 2003, San Diego, CA.*, June 2003.
4. Y-H. Chu, S. Rao, and H. Zhang. A Case for End System Multicast. In *ACM SIGMETRICS*, pages 1–12, Santa Clare, CA, USA, June 2000.
5. Competition Commission. Lloyds TSB / Abbey National Merger Inquiry Open-Meeting Transcript. `http://www.competition-commission.org.uk/inquiries/completed/2001/lloyd%s/lloydstran.htm`.
6. S. Deering and D. Cheriton. Multicast Routing in Datagrams Internetworks and Extended LANs. *ACM Trans. Comp. Syst.*, 8:85–110, May 1990.
7. A. El-Sayed, V. Roca, and L. Mathy. A Survey of Proposals for an Alternative Group Communication Service. *IEEE Network*, 17(1):46–51, Jan/Feb 2003.
8. L. Mathy, R. Canonico, and D. Hutchison. An Overlay Tree Building Control Protocol. In *Proc. of Intl. workshop on Networked Group Communication (NGC)*, pages 76–87, Nov 2001.
9. Princy C. Mehta and Sanjay Udani. Overview of VoIP, Technical Report MS-CIS-01-31. Technical report, University of Pennsylvania, February 2001.
10. H. Sacks. *Lectures on Conversation.* Blackwell, Oxford, UK, 1992.
11. Ulrich Schmitz. Eloquent silence. *Linguistik-Server Essen (LINSE)*, 1994.
12. Su-Wei Tan and Gill Waters. Building Low Delay Application Layer Multicast Trees. In *Proceeding of 4th Annual PostGraduate Symposium (PGNet 2003)*, pages 27–32. John Moore University, Liverpool, UK, June 2003.
13. E. Zegura, K. Calvert, and S. Bhattacharjee. How to Model an Internetwork. In *IEEE Infocom*, pages 40–52, Mar 1996.