

Tolerance of Highly Degraded Network Conditions for an H.323-based VoIP Service

Peter Holmes, Lars Aarhus, Eirik Maus

Norwegian Computing Center, P.O. Box 114 Blindern, Oslo, Norway
{peter.holmes, lars.aarhus, eirik.maus}@nr.no

Abstract. The empirical work presented here concerns the effects of very large packet loss and delay conditions upon the performance of an H.323-based VoIP service. Specifically, it examines call establishment performance and audio voice quality under these conditions. The work was performed as part of an investigation concerning the effects of GEO satellite environments upon H.323-based VoIP, but its findings are relevant to all kinds of networks which may suffer or become subject to highly aberrant transmission conditions. The call establishment tests showed that the H.323 protocols could establish calls successfully under severely impaired conditions. The listening tests showed that uncontrolled jitter was the most destructive parameter to listening quality. Packet loss was also influential, though somewhat less dramatically.

1 Introduction

The empirical work presented here concerns the effects of very large packet loss and delay conditions upon the performance of an H.323-based VoIP system. It was performed during Q2-Q4 2000 within the IMMSAT project (“*Internet Protocol based MultiMedia Services over a Satellite Network*”) [1] [2], a project carried out by Norsk Regnesentral [14] and Nera SatCom AS [16].

Two primary types of VoIP tests were carried out: *call establishment* and *audio voice quality*. The main objectives of prototype testing were:

- To verify that the H.323 protocols could be used to establish calls with an acceptable voice quality, in an environment that approximates the characteristics of a GEO (Geosynchronous Earth Orbit) satellite-based IP network.
- To find the *parametric limits* for delay, jitter, and packet loss where the H.323 protocols could no longer provide an acceptable VoIP service. That is, where call establishment becomes severely impaired / fails and where audio voice quality degrades such that usability of the service is severely or wholly impaired.

Regarding other related work, Bem, et.al. [3] provide a thorough introduction to the area and issues of broadband satellite multimedia systems, clarifying the basic physical and architectural distinctions amongst satellite systems, as well as some of the technical and legal issues to be addressed. Farserotu and Prasad [5] provide a more brief survey, and include concise descriptions of the basic issues and references to the latest work in areas such as enhancements to TCP/IP, enhanced QoS awareness,

IP security over SATCOM, onboard processing, switching and routing, and service enabling platforms. Metz [7] offers another concise introduction.

The IMMSAT results can perhaps be best contrasted to recent work by Nguyen, et.al. [8]. There, the work studied the effects of link errors and link loading upon link performance (e.g., packet and frame loss, delay, etc.) for an H.323-based VoIP service. The IMMSAT study focused upon identifying the levels of link delay, jitter and packet loss which could be *sustained* and *tolerated*.

2 H.323 Call Establishment

H.323 calls [9] can be established according to a variety of call models. A call model basically defines the system entities to and from which signals are exchanged. Readers can find a comprehensive H.323 tutorial at [10]. This section focuses upon signaling procedures and timers relevant to H.323 call establishment.

2.1 Basic signaling procedures for H.323 call establishment

The provision of H.323-based communication proceeds in 5 phases:

- phase A: call setup (RAS and H.225 messages [13]); connection of endpoints; H.245 call control channel establishment [11])
- phase B: initial communication and capability exchange (H.245 procedures); master/slave determination
- phase C: establishment of audio visual communication (H.245 procedures); media stream address distribution; correlation of media streams in multipoint conferences and communication mode command procedures
- phases D and E: call services and call termination, respectively [9].

In IMMSAT, call establishment testing concerned itself specifically with the success or failure of phases A-C. That is, call establishment was judged successful when voice could be transmitted between the endpoints.

2.2 Signaling and timing details within call setup

The discussion in this section specifically focuses upon call setup (phase A) within a Direct Endpoint Call Signaling via Gateway call model. This was the call model used during prototype testing. *It is important to note that in this call model, the Gateway is both a callee endpoint and a calling endpoint.* The RAS and H.225 signal exchanges in phase A for that call model are illustrated in Figure 1.

Timers System entities start and stop various timers as they send and receive different signals. These include the T303, T310 and T301 timers, as specified for use within call setup by H.225 and Q.931¹.

Each calling endpoint starts a T303 timer at the moment it issues a *Setup* signal. By default [13], each calling endpoint should receive an *Alerting*, *Call Proceeding*, *Connect*, *Release Complete* (or other message) from its respective called endpoint within 4 seconds. Should a calling endpoint receive a *Call Proceeding* signal before T303 timeout, it should start a T310 timer. Its T310 timer runs until it receives an *Alerting*, *Connect* or *Release Complete* signal. Typical timeout values for the T310 timer are about 10 seconds [15].

If a calling endpoint receives an *Alerting* signal before a T310 timeout, it should start a T310 timer. This timer runs until a *Connect* or *Release Complete* signal is received. Its timeout value should be 180 seconds or greater [13].

Mandatory vs. optional signaling Not all signals appearing in Figure 1 are mandatory; certain are optional, and certain can be omitted in special cases. For instance, the *Alerting* ("terminal is ringing") signal can be omitted should the callee answer the phone (*Connect*) before it has time to ring. Otherwise, H.225 requires the callee endpoint to send an *Alerting* signal, regardless of call model. H.225 also states that a Gateway should forward an *Alerting* signal (as shown in Figure 1).

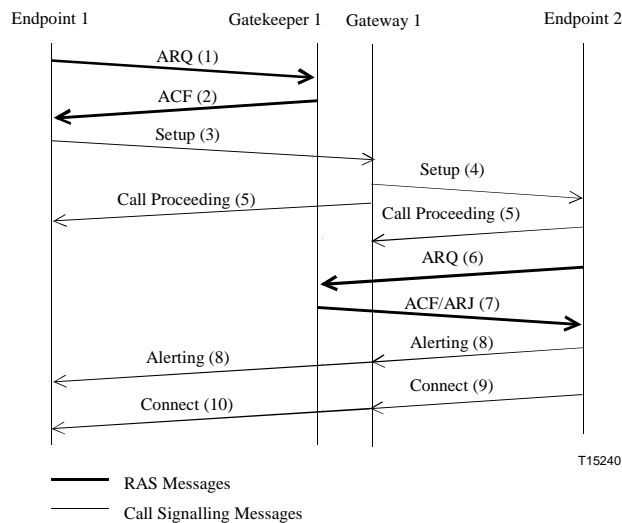


Fig. 1. Direct endpoint call signaling via Gateway: phase A (adapted from [9])

¹ It has been pointed out that certain equipment implementations use two timers in place of the three timers mentioned here (T303, T310 and T301) [15]; the discussion in this section accords with H.225 and Q.931.

When call setup signaling transpires via a Gateway, H.323 requires the Gateway to send a *Call Proceeding* signal to the calling endpoint, whenever the Gateway judges that it might require more than four seconds to respond to that endpoint. This is required in order to help prevent a T303 timeout (or its equivalent) within the calling endpoint. Perhaps surprisingly, H.225 does *not* require that the callee endpoint (e.g., endpoint 2) issue a *Call Proceeding* signal. In network environments with extraordinarily long delays, the absence of a *Call Proceeding* signal from a callee endpoint could yield a T303 timeout (or its equivalent) in the Gateway. In many implementations, a T303 timeout causes call setup to fail².

In short, use of the *Call Proceeding* signal by the callee endpoint can be highly valuable when facing *very* large delay conditions. Use of this signal helps serve to alleviate, though not decouple, end-to-end call setup timing constraints. In satellite environments, issuance of this message from callee endpoints helps mitigate the effects of multiple satellite hops during H.323 call setup.

3 Prototype Testing

The satellite-based environment parameters in focus within IMMSAT were delay, jitter, packet loss and bit error rate (BER). In a typical GEO satellite-based environment, values for these parameters are:

- *Delay*: 520-580 ms one-way delay (includes processing delay)
- *BER*: 10^{-6} average bit error rate
- *Packet Loss*: 1% average packet error (assuming 10^{-6} BER and 100 byte packets on average).

3.1 Prototype Configuration

The network configuration and prototype elements within the IMMSAT prototype are given in Figure 2. Details about the specific kind of equipment used is provided in the figure, and explicated in [6]. References to the equipment include [17][18][21] [22].

In the figure, the H.323 client types are indicated with **NM** for the Microsoft NetMeeting 3.01 clients, **IPT** for Siemens LP5100 IP telephones [20], and **ISDN** for Siemens Profiset 30 ISDN telephones³ [19]. The configuration enabled testing the effects of both *one* and *two hop* calls⁴.

² H.225 and Q.931 allow the *Setup* signal to be retried. Since TCP underlies the H.225 call signaling messages, however, many implementations choose to abort and clear the call, rather than retry *Setup* [15].

³ These acronyms are followed by a number (1 or 2) which distinguishes different instances of the same kind of system entity.

⁴ Two hop calls can occur, for example, when two mobile satellite terminals require (e.g., audio) media transcoding in order to communicate. In such cases, the audio media is transmitted from the first terminal to the satellite gateway via the satellite, transcoded in the

All testing was performed using various combinations of H.323 clients, IP telephones, ISDN telephones, H.323 Gateway, and Gatekeeper (for most tests). A Direct Endpoint Call Signaling via Gateway call model was employed. Voice encoding and decoding was fixed to G.711 in the ISDN/H.320 network, and to G.723.1 (6.3Kb/s), in the IP/H.323 network. During testing, transcoding was always performed in the H.323 Gateway.

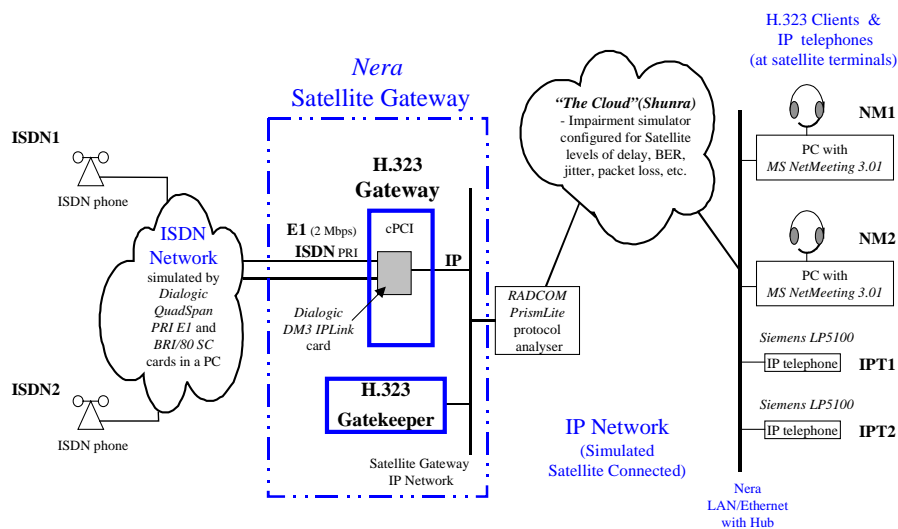


Fig. 2. IMMSAT Prototype, including network and telephony configuration details

3.2 Test Strategy and Approach

A test strategy was chosen in which an effort was made to independently identify critical values for each individual parameter⁵. For call establishment tests, a critical value was a value for a single parameter (e.g., delay, packet loss, etc.) at which call establishment began to fail. Once a critical value was identified for a specific parameter during a test series, two or three calls were made in the parametric space neighboring upon that critical value. Unfortunately, time available for testing yielded only very small sample sizes for each test series.

Early, yet significant experimentation with the impact of varying BER upon call establishment indicated that it was necessary to raise this rate to values greater than

gateway, then transmitted on to the second terminal, again via the satellite. In this example, this process transpires for audio media transmission in each direction.

⁵ Clearly, certain parameters (e.g., delay and jitter) are interdependent. In this document, the term 'individual parameter' is used to mean a parameter which could be independently adjusted within the satellite simulator.

1/50000, in order that call establishment should *occasionally* fail. The same was true in regard to the perception of BER's effect upon audio voice quality⁶. Since such a rate is far outside the expected performance range for most networks, including GEO satellite, further independent testing of this variable was terminated.

Complete details about the approach and test process for call establishment and audio voice quality testing can be found in [6]. Some details about audio voice quality testing are included below, however, due to their relevance in this context.

The voice quality tests included two aspects, one-way *listening quality* and two-way *conversational quality*. To rate the subjective listening quality of audio voice, Recommendation P.800's *Quantal-Response Detectability Testing* scale was employed (see [12], Annex C). An abbreviated version of this scale is included in the legend for Figure 4.

In order to employ a consistent audio reference during listening quality tests, playback of a pre-recorded audio source via the Gateway was used. *It must be noted that the time and resources available for listening quality testing were so limited that only a single subject was employed to perform the listening tests and to rate performance. Testing procedures were made as "blind" as possible [6].*

As part of the listening quality tests:

- silence suppression was turned off in all prototype elements (whenever and wherever possible)
- audio frame size was 30 msec
- number of audio frames per packet was clearly noted as a significant variable (either 1 or 2 audio frames per packet)
- listening quality for both one and two hop calls was checked.

4 Analysis of Results

Since the total number of trials in each test series was so very limited, a judgement was made to analyze the results on the basis of apparent, rather than statistical trends.

4.1 Call Establishment

Figure 3 summarizes two major sets of call establishment tests. Each major set included eight *test series*. These two major sets include:

- investigations of the effects of varying delay (test series A-H) and
- investigations of the effects of varying packet loss (test series J-R).

The relevant variables for each test series are: the type of client terminal initiating the call (the *caller*), the type of client terminal answering the call (the *callee*), whether the call involved one or two simulated satellite hops and the settings for the satellite simulator. When reviewing Figure 3, it is important to note that the horizontal axes employed therein are not linear.

⁶ This is consistent with the findings of Nguyen, et.al. [8].

Effects of delay In regard to the effects of delay upon call establishment, there were two major observations. The first was that in the face of delay, call establishment performance did not seem to be influenced by the addition of a second (simulated) satellite hop. This can be seen in Figure 3 by comparing: test series C to D, test series C to E, and test series F to H. It is expected that this behavior is related to the characteristics of a Direct Endpoint Call Signaling via Gateway model which fully employs the *Call Proceeding* signal, as explained in section 2.2.

With respect to this first observation, the comparison of F to G is *not* conformant. It is presumed that this result is a consequence of the second observation described below.

The second major observation is that increases in delay apparently seemed to impact calls employing IPT (as either caller or callee) *more adversely* than calls which did not involve IPT. Rather than including here a detailed clarification involving signal exchanges and timing, a higher-level explanation is offered instead.

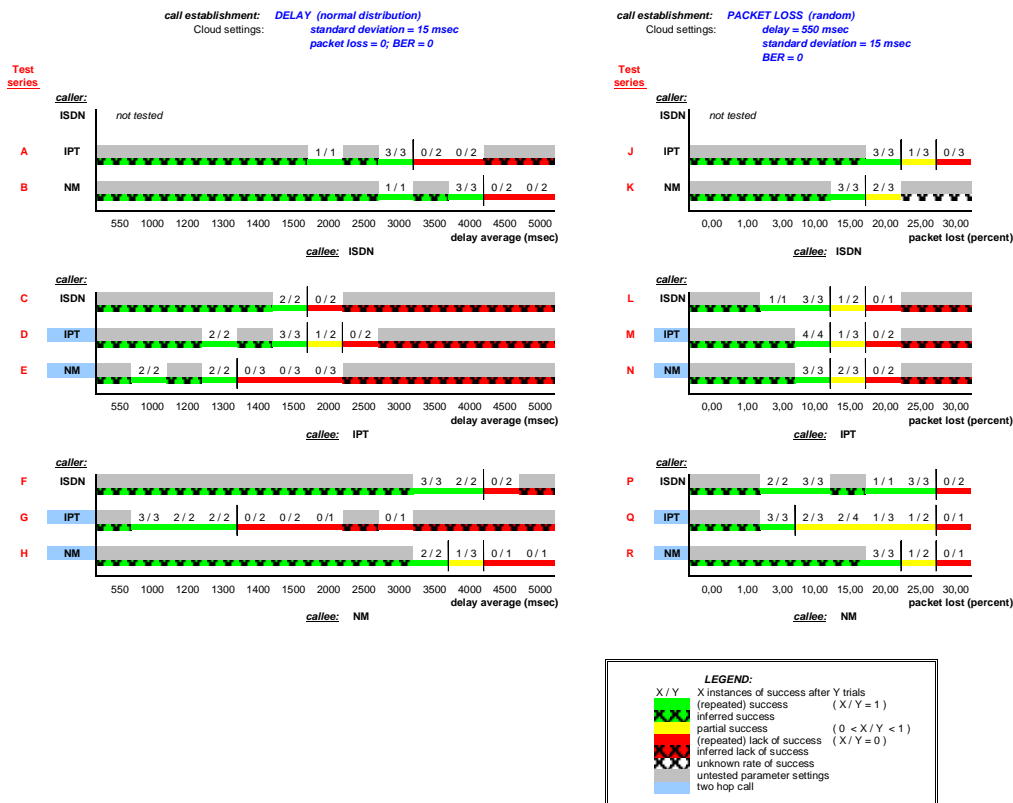


Fig. 3. Results of call establishment testing for delay and packet loss

IPT as a caller endpoint: In test series F, G and H, the *second* leg of the call involves signaling between the Gateway and NetMeeting. Series F and H both demonstrate that the second call leg *can* tolerate up to 4000 ms delay. Series G begins to fail at 1400 ms, however, which implies that the cause of failure lies on the first leg of the call. A reasonable hypothesis in this case is that timeouts occurred within the IPT terminal, the originating caller.

IPT as a callee endpoint: In test series B, E and H, the *first* leg of the call involves signaling between NetMeeting and the Gateway. Series B and H both demonstrate that the first call leg *can* tolerate up to 4000 ms delay. Series E begins to fail at 1400 ms, which implies that the cause of failure lies on the second leg of the call. It is also interesting to notice that the other two series having IPT as callee endpoint (series C and D) also demonstrated failure at 2000 ms. A reasonable hypothesis in this case is that IPT did not issue a *Call Proceeding* signal to the Gateway, a situation which caused the Gateway to timeout and clear the call. As mentioned earlier, H.225 does not require that callee endpoints issue the *Call Proceeding* signal. Still, these test series seem to demonstrate the value of that signal being issued by the callee endpoint.

It must be mentioned here that though reasonable, the two hypotheses above could not be *completely* verified. The circumstances regarding this condition are explained further in [6].

Effects of packet loss Analysis of the call establishment testing for packet loss did not reveal any apparent trends with respect to the variables under study (e.g., caller vs. callee type, hop count, etc.). The only apparent trend is that the success rate of call establishment seemed to significantly diminish when packet loss reached 15-20%. It should be mentioned here that some of the *successful* call establishment trials employing packet loss rates of 20-25% took one minute or more to establish. Despite the lack of usable audio quality at this level of loss (see below), successful call establishment could still be of value in other kinds of contexts (e.g., a “sign of life” in an emergency).

4.2 Listening Quality

Figure 4 summarizes six major *cases* of listening tests: one matrix for each major case (case matrices A-F). The approach used for testing is completely described in [6]. In the figure, each case matrix is also colored according to a “usability scale” defined for IMMSAT. The colors essentially map P.800’s seven point quantal-response detectability scale ([12], Annex C) onto a three point scale; this was done in order to ease perception of the significance of the listening test results. The legend in Figure 4 explains this mapping. Note that in certain case matrices, parts of the matrix are colored despite the fact that no trial was performed. These instances of coloring were *inferred*, based upon the performance of other trials within the same matrix.

The parameters which *distinguish* each major case are: the type of client terminal initiating the call (the caller), the type of client terminal answering the call (the callee), whether the call involved one or two (simulated) satellite hops, whether the listening test was performed using a headset or a handset, and the number of audio

frames per packet. For the cases in which IPT was the callee, an additional parameter was the setting for the HiNet jitter buffer; this buffer could be set to either ‘Long’ or ‘Short’. The HiNet jitter buffer is a buffer internal to the Siemens IP telephones. The parameters which distinguish the trials *within* each major case are the settings for the satellite simulator. It is important to note that the axes employed for each major case matrix are not linear.

The analysis here is discussed using *comparisons* amongst major cases A-F (i.e., “Case Comparisons I-V” in Figure 4’s legend); these are depicted as (red) arcs in the figure. The legend succinctly describes the variation in parameter values for each case comparison. Note further that the case comparisons can be viewed as tree rooted at case B.

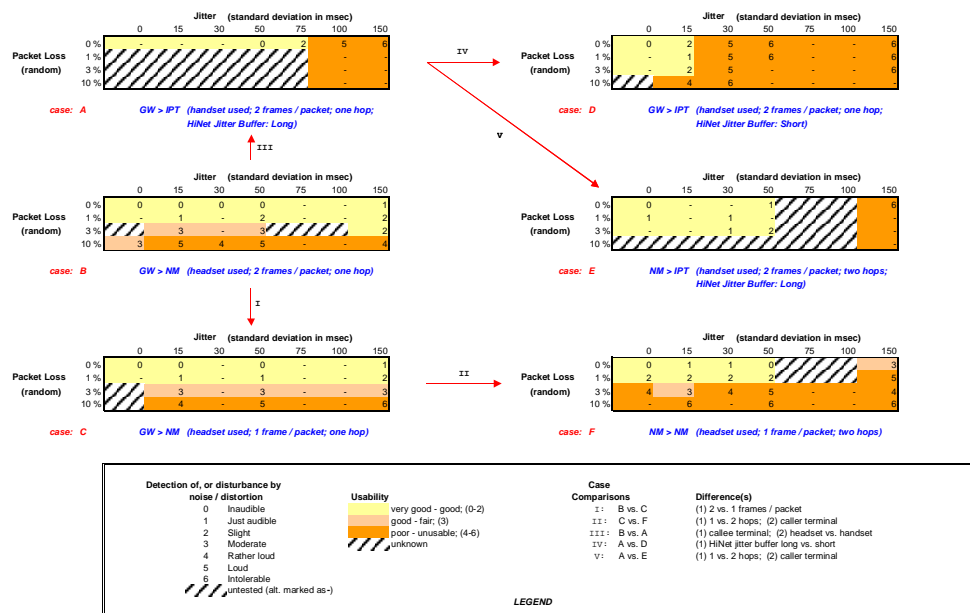


Fig. 4. Results of listening tests for jitter and packet loss

Case B depicts a listening test where the configuration supported “good – very good” audio quality for jitter values as high as 150 msec and packet loss up to 1%. Audio quality began to noticeably degrade to “fair” when packet loss reached 3%. Comparison with case C (comparison I) indicates that use of two vs. one audio frames per packet does not appear to be of any significance.

Differences are noticeable in comparison II, however; case F involves two hops, while case C involves only one. This comparison illustrates that audio quality degrades quite noticeably when a second hop is introduced, an expected result due to the cumulative effects of packet loss, delay and jitter across hops. Worth noting in case F is the near-usability of the trial having 150 msec jitter and 0% packet loss per hop. This result seems to indicate that NetMeeting is internally operating with a

relatively “large” jitter buffer. No user-level access to this parameter was available via the NetMeeting application, however, in order to confirm this hypothesis.

Consider now comparison III (case B vs. case A). In case A, the most significant difference from case B is that instead of NetMeeting, the Siemens LP5100 telephone is used as the callee terminal. For the LP5100 telephone, jitter has a severely destructive impact upon audio quality as soon as it reaches 100 msec. This was the case even though the HiNet jitter buffer was set to ‘Long’. When that buffer was set to ‘Short’ (see case D, comparison IV), the effect of jitter was devastating.

In contrast to the LP5100 telephone’s poor performance in the face of jitter, the results seem to indicate that its audio quality was somewhat better than NetMeeting, when faced with the same levels of packet loss. Though this is not explicitly shown in comparison III, one may choose to infer this by considering how well the LP5100 performed with respect to packet loss in case E — a two hop call (comparison V).

In summary, the one-way listening tests showed that uncontrolled jitter was the most *destructive* parameter to audio quality. Packet loss was also influential, though somewhat less dramatically.

4.3 Conversational quality

Even when the intelligibility of audio is unaffected by its transmission medium, the exclusive effect of delay can have severe consequences upon the character of a conversation (see e.g., [4]). Since this area has already been so well investigated, the tests for conversational quality were only quick, general checks performed using the typical values expected for delay, BER, packet loss, etc. In general, the conversational quality between clients at these levels was satisfactory. Of note, the two hop calls tended to include some (very) small amount of added noise / distortion, when compared to the one hop calls.

For the sake of experience, conversational quality was occasionally tested when delays were varied between 1500-4000 msec and other satellite parameters were held at their typical values. As expected, users could ultimately manage to successfully communicate. With such large delays, however, it was necessary for them to adopt an unnatural “single duplex” mode of conversation.

5 Conclusions

Generally, all tests performed using the IMMSAT prototype indicated that H.323-based VoIP services in a GEO satellite-based IP network are practical and usable. The effects of *typical*, GEO satellite environment conditions upon H.323-based VoIP services are expected to be negligible. These results are consistent with those of Nguyen, et.al. [8]. Further remarks about call establishment and audio voice quality follow below.

5.1 Call Establishment

The call establishment tests showed that the H.323 protocols could establish calls successfully under severely impaired conditions — conditions which were at least an order of magnitude worse than in typical GEO satellite-based environments.

The call establishment tests also indicated that the client type on each end of the call can have great significance with respect to the amount of delay which can be tolerated during call establishment. For example, it appeared that one of the clients failed to issue the *Call Proceeding* signal. This signal is not required by the standard but, when facing extreme conditions, its use can yield greater tolerance of delay.

Lastly, it was interesting to see that both one and two hop call configurations performed equally well in the face of delay. It is expected that this behavior is due to the characteristics of a Direct Endpoint Call Signaling via Gateway model which fully employs the *Call Proceeding* signal.

5.2 Audio Voice Quality

The experiments showed that audio quality appreciably degrades when moving from a one hop call to a two hop call. This kind of result is expected due to the cumulative effects across multiple hops. The experiments showed that client type also has an impact upon voice quality; use of headset instead of a handset also makes a difference here.

Still, jitter is the network characteristic that seemed to have the greatest impact on listening quality. Even though the highly adverse conditions tested in this investigation are not likely to be observed in most networks, the ability to handle jitter is crucial for offering an acceptable VoIP service. Using a larger jitter buffer in certain system entities may help alleviate the problem, though at the cost of an increase in overall delay.

The experience of large delays when trying to converse can be frustrating and confusing, even to users who have been informed of and prepared in advance for such conditions. Users practiced with large delays *can* adapt and reconcile themselves with the condition and ultimately use the service, but conversations must transpire in a “single duplex” mode in order that communication be achieved.

Acknowledgements

The authors wish to thank the project team at Nera SatCom: Ronald Beachell, Harald Hoff and Paul Mackell. A special thanks as well to Ronald Beachell, for his diagram of the IMMSAT prototype configuration (Figure 2).

References

1. Beachell, R.L., Bøhler, T., Holmes, P., Ludvigsen, E., Maus, E., Aarhus, L., "1551-IMMSAT; IP Multimedia over Satellite Study Report", Copyright 2000, Nera SatCom AS / Norsk Romsenter. Also available as: "IP Multimedia over Satellite: Study Report", NR Report 967, December 2000, ISBN 82-539-0473-8. See: <http://www.nr.no/publications/nrrapport967.pdf>
2. Beachell, R.L., Holmes, P., Thomassen, J., Aarhus, L., "1523-IMMSAT-TR; IP Multimedia over Satellite: Test Report for the IMMSAT Prototype", Copyright 2000, Nera SatCom AS / Norsk Romsenter. Also available as: "IP Multimedia over Satellite: Test Report", NR Technical Note IMEDIA/10/00 (Classified), December 2000.
3. Bem, D., Wieckowski, W., Zielinski, R., *Broadband Satellite Systems*, IEEE Communications Surveys & Tutorials, vol.3, no.1, 2000, pp. 2-15; see: <http://www.comsoc.org/pubs/surveys/1q00issue/pdf/zielinski.pdf>
4. Douskalis, B., *IP Telephony - The Integration of Robust VoIP Services*, Prentice-Hall, 2000.
5. Farserotu, J., Prasad, R., *A Survey of Future Broadband Multimedia Satellite Systems, Issues and Trends*, IEEE Communications, vol. 38, no. 6, June 2000, pp. 128-133.
6. Holmes, P., Aarhus, L., Maus, E., "An Investigation into the Effects of GEO Satellite Environments upon H.323-based Voice over IP", NR Report 973, May 2001, ISBN 82-539-0479-7. Available as: <http://www.nr.no/publications/report973.pdf>
7. Metz, C., *IP-over-Satellite: Internet Connectivity Blasts Off*, IEEE Internet Computing, July-August 2000, pp.84-89.
8. Nguyen, T., Yegenoglu, F., Sciuto, A., Subbarayan, R., *Voice over IP Service and Performance in Satellite Networks*, IEEE Communications, Vol. 39, No. 3, March 2001, pp. 164-171.
9. *Packet-based multimedia communications systems*, ITU-T Recommendation H.323 v3 (with change marks), ITU-T October 1999.
10. From DataBeam: "A Primer on the H.323 Series Standard", see: http://www.h323analyser.co.uk/h323_primer-v2.pdf
11. *Control protocol for multimedia communication*, ITU-T Recommendation H.245 v5 (with change marks), ITU-T 1999.
12. *Methods for Subjective Determination of Transmission*, ITU-T Recommendation P.800, ITU-T August 1986: <http://www.itu.int/itudoc/itu-t/rec/p/p800.html>
13. Call signaling protocols and media stream packetization for packet based multimedia communication systems, ITU-T Recommendation H.225.0 draft v4 (with change marks), ITU-T 1999.
14. Norsk Regnesentral (Norwegian Computing Center) WWW site; see: <http://www.nr.no/ekstern/engelsk/>
15. Personal communication from Espen Kjerrand, Ericsson AS Norway, May 2001.
16. Nera SatCom AS WWW site; see: <http://www.nera.no/>
17. Dialogic IPLink cPCI Solutions; see: http://www.dialogic.com/products/d_sheets/4798web.htm
18. Dialogic QuadSpan™ Voice Series; see: http://www.dialogic.com/products/d_sheets/3967web.htm
19. Siemens Profiset 30isdn: see <http://www.siemens.no/telefoni/isdn/profi30.htm>
20. Siemens optiPoint 300 advance IP telephone (formerly LP5100 IP); see: <http://www.ic.siemens.com/CDA/Site/pss/1,1294,208375-1-999,00.html>
21. The Cloud, from Shunra Software Ltd.; see: <http://www.shunra.com/products/thecloud.htm>
22. RADCOM Prism and PrismLite protocol analyzers; see: <http://www.radcom-inc.com/radcom/test/prism.htm>