The Vulnerability of Videoconferencing to Voyeurs

by Paul H. Fredette
Chief Operating Officer & Chief Technical Officer
Promptus InfoCrypt, Inc.
Technical Committee Chairman
BONDING Consortium

Abstract

High-speed ISDN videoconferencing, based on ITU H.320 standards, was previously believed to be very difficult or even impossible to intercept. This paper dispels this belief by describing how the audio portion of a conference can be intercepted using a single-channel of a standard, multi-channel videoconference.

Alerted to the possibility of audio-from-videoconferencing interception, the author conducted a detailed study by researching what gaps could exist in videoconferencing's inherent "security". He then successfully designed, developed and built a low-cost interception device using off-the-shelf components.

The study revealed the complications and solutions related to implementing a real-time interception device. It is the outcomes of this study that are described in this paper.

Introduction

High-speed ISDN videoconferencing, based on ITU H.320 standards, was previously believed to be very difficult or even impossible to intercept. This paper dispels this belief by describing how the audio portion of a conference can be intercepted using a single-channel of a standard, multi-channel videoconference.

Although revealing this vulnerability could alert voyeurs to the opportunities for interception, the most dangerous security risks are the ones that are not exposed.

Experts agree that the security methods considered to be the most secure are those which are exposed to analysis. This is why the most commonly used encryption algorithms are publicly disclosed.

After a detailed study of the interception threat to videoconferencing, it is now known that a real-time audio decoder can be constructed for less than US$100. This is a dramatic comparison with previous estimates of around US$50,000 using off-the-shelf test equipment.

PC software programs available for decoding multimedia, provide similar functions for free, although this avenue was not explored in our study.

A study of the vulnerability of the audio portion of the standard videoconference was initiated after it was observed that all of the audio is contained in only one of the network channels. At least one previous paper on the subject, indicated that the multi-channel nature of most videoconferences afforded users a high level of resistance to interception.

It is the author's contention, that reliance on the multi-channel nature of videoconference calls as a security tool is, at the minimum, surprisingly naïve.

Tapping a telephone call is relatively easy in either analog or digital form. This is because the speech signal is directly transmitted either in analog form or in a sampled format using ITU G.711 encoding. Telephone wire-tapping demonstrations are regularly seen at trade shows because G.711 test sets are a dime a dozen. Well, maybe a bit more, but decoder chips can be purchased for under US$5.

The Link to Listening In

Until now, there has been a missing link to solving the challenge of intercepting a videoconference: the lack of inexpensive test equipment to decode the speech portion of an H.320 signal. This issue is compounded by the multiple encoding options available on most videoconferencing equipment.

The primary audio compression method used on videoconferences today is the ITU-T G.722 standard. It's a popular choice because of its ability to provide better-than-phone-quality audio at less bandwidth than a telephone call. Whilst other algorithms such as G.728 (used for low bandwidth videoconferencing) do exist, the techniques described here are readily useable for G.722, G.728 and other published compression methods. In fact, the decoders for many lower-bandwidth speech processing algorithms are not proportionally more complicated just because the encoders are.

For example, MP3 audio decoding is far simpler than the encoder because no iterative compression methods are used. Once the encoder has decided what to send, the decoder need only follow its decision. If you doubt this, look at the typical cost of encoders, such as PDF file compressors, versus the zero cost of downloadable viewers.

Returning to G.722, the specification is quite complicated, but the complexity would not deter a serious voyeur. In fact, the author chose to begin this adventure by getting all the technical facts possible……from the Internet. All of the specifications can be downloaded for free at www.ITU.com as long as one is willing to provide an e-mail address.

In coding the decoder algorithm from scratch, one difficulty the author faced was the surprisingly large number of typographical errors found in two versions of the G.722 specification (the 1988 version and the current version). To establish a proper implementation, an Excel™ spreadsheet was used to compare the computed output with the test data provided by the ITU-T. This test data was also available as a free download from the ITU web-site.

Further coding on a Digital Signal Processor (DSP) was required to complete the computations in real-time. The author and a colleague performed the DSP coding and debugging in about two weeks.

A DSP demonstration board and software for the TI TMS320C31 DSP was used to code and to debug the program: the cost of the demonstration equipment totalled about US$75. The only simplification the author used was in the selection of one of several possible signal intercept points. The ISDN “S” bus was selected for extracting the electrical signals directed to the DSP for processing, because it is the ISDN equivalent of in-building telephone wiring. An off-the-shelf BONDING IMUX had the appropriate hardware, so it was used it to complete an initial proof-of-concept design.

The actual hardware needed is now in the prototype stage, comprises less than US$100 in material and will not require software beyond the decoder algorithm. Copies of the decoder software in both Microsoft Excel™ format and DSP assembly language format (for the DSP development equipment) are available as free downloads on our web-site, www.infocrypt.com

The implementation process was made somewhat complicated by the number of errors found in the specifications. However, an engineer with a signal processing background, such as the

author, could resolve these inconsistencies by comparing two versions of the specifications and examining the computed results compared to the provided test results.

BONDING Dispersion Does Not Enhance Security

Most videoconferences today use a technique for aggregating multiple telephone channels (BONDING) in order to achieve the desired conference quality. The resulting dispersion of information in the network was considered by some to make interception almost impossible, except perhaps by the most determined and well-funded voyeur.

In reality, BONDING only corrects information dispersion *if* it occurs. Most channels between two points follow a similar if not equal path through the network.

Access circuits typically have channels combined onto 1 or 2 wire pairs or optic fiber along with other telephone circuits. However, since only a single channel is necessary for decoding the audio portion of a videoconference, BONDING provides little or no protection from dispersed channel effects.

Worse, experiments have shown that the more recent videoconference terminals, with integrated BONDING devices, transmit data aligned so as to guarantee the interception of all the audio data on a single channel. This was dismissed earlier when the author's initial study indicated that external IMUX equipment would reduce the probability of single-channel intercept to 50%. The remaining 50% however, would only require the interception of two channels, independent of the number of bonded channels.

There is a simple solution to this interception threat……encrypt all of your transmissions.

G.722 Decoding Simplified

A G.722 listening decoder is much simpler to implement than the encoder. Commercial software products typically provide a complete algorithm set of encoder and decoder. Prices for such software has been quoted at about US$18,000; this may have discouraged previous attempts at implementing for interception.

The designers of G.722 made the algorithm very flexible, but all modes keep the telephone quality audio in 4 bits. The high-frequency portion of the decoder is not required to determine the content of the conversations and this portion can be ignored in searching and decoding channels.

Another simplification lies in the ITU-T H.221 specification for multiplexing the audio and video portions of the conference. H.221 uses frame alignment signals called FAS and BAS however these are NOT required to snoop the audio. This apparent complexity could lull users into a false sense of security.

Tiptoeing Through T1s Trivialises Tapping

Searching for a single channel within the telephone network is actually quite simple because of the level of diagnostic capability provided by network switches.

Many readers also may be familiar with multi-channel T1 (1.544 Kbit/s) or E1 (2.048 Kbit/s) transmission lines used to carry phone calls. The hardware described above would work

quite readily on a T1 line instead of the ISDN “S” interface used for our example implementation.

The change in hardware is only in the selection of the interface chip and a channel selector. This would not affect the cost of the intercept unit by more than a few dollars. All that is needed is a little extra patience to scan through the 23 (30 for E1) channels until a conversation is heard. Again, the threat still remains when high-speed digital access facilities are used.

Conclusions

Think of the effect of an audio bug in a videoconference room. The audio component of a videoconference is likely to contain the most sensitive portion of the information content of a conference.

Now consider that this audio is almost as easily tapped as Plain Old Telephone Service (POTS) and that wire-taps do occur all the time.

Since the most dangerous security risks are the ones that are not exposed, knowing this vulnerability allows for a plan to protect against it.

The simple solution to this vulnerability is to encrypt all of your transmissions. Also use at least 128 bit encryption keys to avoid practical DES decryption devices from deciphering your data.

Audio tapping is not the only security concern for videoconferencing, and other aspects of physical security for conference rooms should also be considered. These other security aspects have not been considered within the scope of this paper.

Promptus InfoCrypt, Inc.
207 Highpoint Avenue
Portsmouth, RI 02871
USA

119 Willoughby Road
Crows Nest NSW 2065
Australia

www.infocrypt.com