White Paper on VOIP, Ethernet, Internet, ATM and other Topologies with Regards to Broadcast and Professional Intercom Systems.

Telex Communications, Inc.

Revised May 2005

The intent of this paper is to provide an overview of a number of voice and data transport mechanisms and their strengths, weaknesses, and application to Broadcast and Professional Intercom Systems.

A number of factors will be taken as "given" in this discussion, a few of the more critical ones are presented here.

1) Broadcast and Professional Intercom Systems are used for live "real-time" activities involving the coordination in real time of the activities of a number of individuals. Delay times in communications in excess of 100 msec become problematic in many cases.

2) Broadcast and Professional Intercom Systems are used for "mission critical" activities such as live production of Nightly News, the coordination of activities in refueling a nuclear reactor, or in military communications. System "up times" in excess of 99.999% (<5 minutes downtime per year) are typically required.

3) Noisy, distorted, intermittent, delayed communications are ALWAYS preferable to no communications at all.

A very limited "layman's" knowledge of computers and audio is needed to best understand this guide. A number of networking manufacturers and organizations maintain extensive up to date glossaries on their web sites. One particularly thorough glossary can be found at:

http://www.cisco.com/univercd/cc/td/doc/cisintwk/ita/ita_book.pdf

Baseline for Discussions

The discussions will specifically relate to the RTS ADAM family of Intercoms, and the principles will apply to most Broadcast and Professional Intercom Systems. In the ADAM series of intercoms, there are three "types" of connectivity used between equipment.

Keypanel to Matrix – this requires a full duplex, bi-directional audio path; one balanced +8 dBu audio in one direction and another in the opposite direction. Also required is a bi-directional pair for 9600 bit per second RS-485 data.

Matrix to Matrix via Telex Intelligent Trunking – this requires a single audio path per trunk (typically trunks are "2 way", so an audio path is required in BOTH directions, but this is not ALWAYS the case) and it requires a single bi-directional 9600 bit per second (or faster) RS-232 between each matrix and the trunk master.

Matrix-to-Matrix Bus Expansion – this is a serial connection running at approximately 200 Mbits/second, carrying audio and data. There are no recorded instances of this signal being "embedded" in any form of the connectivities to be discussed, so it will be ignored in this paper.

NETWORKS

No, not ABC, NBC, CBS, BBC, NHK, Fox, or Turner. As the "N" in "LAN" "WAN" and the "net" in "Internet" a network is most simply an interconnection of equipment via a common path. Think of a road; with wires instead of pavement, with data instead of cars, with nodes instead of on ramps. (Yes, this is where the whole "information superhighway" reference came from).

In this discussion, a network can be thought of as a set of wires and equipment, which transport data (whether that data represents audio, control, web pages, corporate data, print jobs, or software updates is immaterial) from one place to another (or to multiple places). The key things to keep in mind are that YOUR data is likely sharing the path with other data, and the path may be empty or overloaded, or anything in between.

So… "Local Area Network", i.e. LAN, is a network in a given LOCALity. This could be between your two PCs and your cable modem at home, or it could be the Network interconnecting 350 users in a 5-story 200,000 square foot office building.

"Wide Area Network" (WAN) is typically a group of LANs in geographically separate areas (campus, cities, states, countries) connected and organized to function as a single entity.

"Virtual Private Network" (VPN) is a method of making a PUBLIC network (such as the Internet, where the data being transported is generated by separate entities and is viewable by multiple entities) seem as if it were private (a Network in which one individual or company has full control and exclusive access to the data) – typically this means embedding an encryption scheme on all the data, so that even if (when) someone intercepts the data on the Internet, it can neither be "read" or altered. This does NOTHING to improve the speed or reliability limitations of a public network.

Bandwidth Requirements and Data Reduction

High quality audio is a very high bandwidth thing. A single channel (mono) of CD quality Audio (such as the quality of signals in RTS ADAM and ZEUS intercoms) is “natively” in excess of a 700,000 bits per second signal. (44.1 kHz times 16 bits). That, in most cases, is WAY TOO HIGH of a data rate to be cost effectively transported across a data network – it is because of that problem that MP3 compression for music files was developed. With MP3 and other schemes, the 1,400,000 bits per second required for stereo has been reduced to 256,000 or 128,000 bits per second with little audible degradation to the signal.

Even 128,000 bits per second is a VERY HIGH load on a data network – as an example, a computer on a network in an office building may require peak data rates of 50,000 bits per second when sending a LARGE complex file to a network printer – and the duty cycle of such jobs for a single user may be five 1 Megabyte jobs in a day, for a duty cycle of 100 seconds in a 28,800 second (8 hour) day, or roughly 0.3% duty cycle, for an average data rate of 150 bits per second.

Now compare that AVERAGE data rate for the user with occasional COMPLEX print jobs of 150 bits per second to the 700,000 bits per second for a CONTINUOUS audio path, and you find that a single audio signal may “load a network” as much as 4,000 or more users.

There are four basic ways to “bit rate reduce” or compress audio signals, we will discuss these very briefly as a way to familiarize the reader with the compromises that will ALMOST CERTAINLY have to be made to use a data network for Broadcast and Professional Intercom signals.

1) Lossless data compression. This is the same technology as is used with “zip” files on a computer, and used with modems to increase effective transmission speed. With a complex, non-coherent signal such as audio, this is usually very inefficient, and results in very little data reduction.

2) Lossy data compression, also somewhat esoterically referred to as “psycho-acoustic compression”. This is the method that MP3 and mini-discs employ to reduce the data bandwidth by a factor of 5 or 10 times. This can be very effective, but require non-trivial processing power (with attendant cost and speed issues). When done well, there is little or no audible effect on the sound quality.

3) Bandwidth limiting. The 44,100 samples per second referred to previously allow for a full 20 kHz audio bandwidth. In reality, the human voice rarely exceeds 5 kHz, and perfectly acceptable “telephone quality” signals can use just 3 kHz. In the earlier example, the “accuracy” was specified as 16 bits; again, in reality, human speech can be clearly understood with as little as 8 bits of data. Combining 3 kHz (requiring 6,000 samples per second) bandwidth with 8 bits of accuracy reduces the 700,000 bits per second to 48,000 bps (a 93% reduction!)

4) Duty cycle limiting. A variant of item 2 above, if a threshold for audio level (VOX) is set, then the gaps between words, sentences, and general silences can be set to generate NO data, further reducing data rate. If a person speaks for a total of 45 minutes in an 8 hour period, this alone can reduce total AVERAGE bandwidth requirements by 90%. The RTS Terminology for this gating or duty cycle limiting is “VAD” Voice Activity Detection.

In summary, the four techniques outlined above are often all used in commercial MUX/DEMUX/VOIP equipment manufactured to put telephone audio signals on networks. The result is that perfectly acceptable telephone grade audio can be compressed to only require PEAK data rates of 16,000 to 40,000 bits per second, which when combined with duty cycle limiting may result in an AVERAGE data rates of 2,000 or less bits per second.

Several Internationally agreed upon standards exist for the specific purpose of reducing data bandwidth to different degrees depending on audio quality and latency requirements. Some of the relevant standards are G.711 G.723 and G.729 A&B. These standards can effectively transmit a voice grade audio signal at data rates between 29,000 to 224,000 bits per second BEFORE duty cycle limiting as described above.

Codec	Bit Rate	Coding Delay	Playout Delay	Bandwidth	MOS*
G.711	64	125 μs	20-60 ms	160–224 kbps	4.3
G.729AB	8	10 ms	20-120 ms	32-112 kbps	3.95
G.723	5.3, 6.3	30 ms	60-120 ms	29-45 kbps	3.5, 3.9

*MOS (Mean Opinion Score) or ACR (Absolute Category Rating) is a widely known quality measuring method. The scale ranges from 5 (excellent) to 0 (unacceptable). The typical desirable range for VOIP transmission is from 3.5 to 4.2.

NOTE: The Playout Delay and Bandwidth depend on the configured amount of audio per packet. The bandwidth values are for bi-directional audio without VAD (Voice Activity Detection) enabled

Now instead of having audio which requires the bandwidth of 4,000 “office users” as described earlier, we are down to that bandwidth required for 10 or so users……completely acceptable in most networks.

THE INTERNET

We all know what the Internet is ..right? It’s the…uhh….well…..you know…..web pages, America On-Line, Kazaa, Amazon place….that stuff.

Without getting into a bunch of heavy technical details and history, it is basically an unregulated conflagration of interconnecting networks, data sources (web sites, web-cams, file servers, etc.), and users. It has been designed to be as fault tolerant and “self healing” as possible, using sophisticated means to get a signal (data) from point “A” to point “B”, even if it means sending said data around the world 5 times, through points C, D, E, F, G3, H12 and Q13 multiple times and taking forever (okay, from milliseconds to multiple seconds) to get there.

Add in the “World” as its domain (the first “w” in “www”) with 7 Billion people and 24/7 use, mix in various groups who like to try to cause chaos, and the REAL transit time for ANY data on the Internet is indeterminate…data may get from New York to Los Angeles in 50 milliseconds, while data from Dallas to Fort Worth takes 2 seconds.

The point is that while it is ubiquitous, cheap and reliable, it’s definitely not the way to go if you want to get data from point “A” to point “B” in as little time as possible CONSISTENTLY.

VOIP

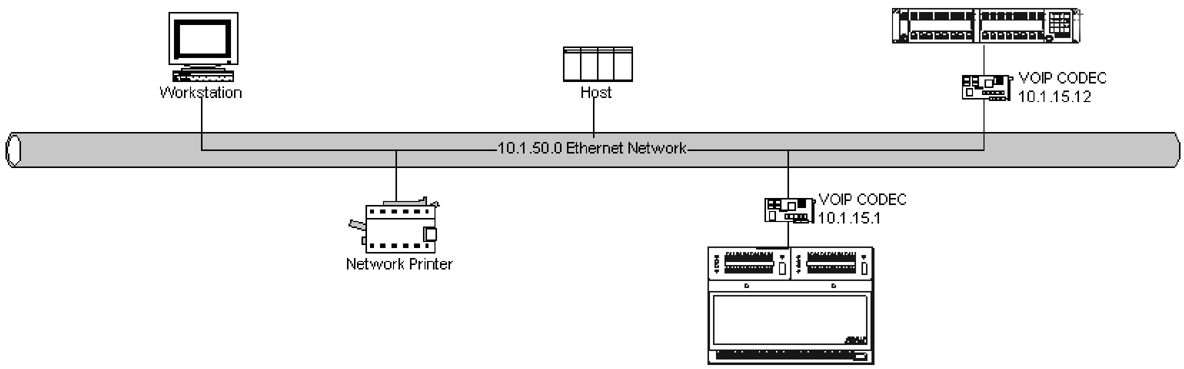
VOIP is the acronym for Voice Over Internet Protocol.

This is IMPORTANT, Get it straight, this is NOT VOTI (Voice Over THE Internet).

With current technology and topology, the Internet is improving rapidly for such uses, but is still nowhere near reliable enough in terms of data throughput, latency or uptime to be used on a regular basis for critical communications. The vast unregulated PUBLIC Internet has delays and transit times so extreme that even with a cable modem or DSL connection capable of 2 MILLION bits per second, audio sent across town via internet can, in extreme cases, take seconds to arrive (That’s why when you use Media Player or RealAudio to listen to Internet audio, the player buffers (or stores up) 20 seconds or so of audio before you hear the FIRST sound). This is certainly not real time audio. The specific amount of delay which will be encountered depends on many factors including network congestion, transit distance, compression algorithms, and other factors.

VOTI should not even be considered unless a MINIMUM 768K DSL connection is available.   VOIP on the other hand can work VERY reliably when as little as a 64K channel of a fractional T-1 or E-1 is available on a dedicated (non-Internet) data network.

The “IP” in VOIP is Internet Protocol. IP is a specification for the bundling, addressing, routing, and error correction of data sent over a network connection. That physical connection can be a Local Area Network (LAN), a Wide Area Network (WAN), or a number of other Network topologies.   The key to making VOIP suitable for (as example) Keypanel to Matrix connection is to insure that the Network BETWEEN the 2 points has the available data capacity and that the latency of the network is low. Once this has been done, it is a relatively simple matter of converting the audio and data signals into IP compliant bit streams using any number of off the shelf converters, setting the addresses (IP address) of the converters so that the source talks to the destination and vice versa, and VOILA! Your intercom system has VOIP capability.



The above discussion also applies to using VOIP to provide the trunk audio and data connections between matrices.

Ethernet

Ethernet is ONE (the dominant ONE) standard for the Networks referred to above. Perhaps the most important characteristic of Ethernet is that it is a “collision based” transmission method. What is “collision based” you ask?.

Imagine four cars coming into an intersection, each from a different direction. All have completely blacked out windows, and a compass to tell them which direction (where) they are to go. Car #1 gets to the intersection (without realizing there is even such a thing as an intersection) 20 seconds before any of the other cars, and soars right through, no problem.

Cars #2 and #3 are not so lucky. #2 traveling from the west and #3 from the south enter the intersection at the same instance –COLLISION!. No real damage done, they each back up and try again – COLLISION!!. No real damage done, they each back up, but this time car #2 is a bit slower, so car #3 gets safely through the intersection BEFORE car #2 can collide with it.

These collisions take time…the transit time for the intersection was both longer than it would have been had there been NO collisions, and more importantly, it was indeterminate; there was no accurate way of predicting the transit time before the cars set out on their respective journeys.

Car #2 now enters the intersection just as Car #4 arrives from the North – COLLISION!. Car #2 this time (again) is slower to back up and retry, so on the next try, Car #4 gets safely through the intersection, and then, when all other traffic (the network term IS “traffic”) is gone, Car #2 finally makes it through the intersection and on to its destination.

OK, now instead of cars, imagine various bits of data inserted into the “intersection” from various sources at various times going various directions. Unpredictable, uncontrollable collisions will occur.

To bring us back from the world of automobiles to data, it should be noted that unless the traffic on an Ethernet system is so near to capacity as to have virtually non-stop collisions, the delays introduced from the collision based system are reasonably trivial (well within the 100 msec) and as such do not impede our ability to use Ethernet based systems (as most LANs and WANs are) for keypanel or matrix interconnections. The key here is “***near to capacity***”, try to add in the bandwidth demands of a few dozens of audio trunks (bandwidth “hogs”) for intercom to an already over-burdened office LAN, and your intercom AND MIS staff are both going to be unhappy – massive collisions, massive delays.

Collisions on LANs can be somewhat mitigated by using a number of techniques, including use of switches and routers instead of hubs.

Useful bit of trivia….many folks (even those who know better) will at times use the terms “LAN”, “Ethernet”, “the network” and even “CAT-5” interchangeably, even though they are not. (“CAT-5” is the most common type of cable used for LANs, Ethernet based and otherwise).

ATM

ATM stands for **A**synchronous **T**ransfer **M**ode. Very, very briefly it is a protocol which can be superimposed “on top” of a network topology (such as Ethernet) to allow certain messages or data to be given priority so that they DO arrive within a certain time, as opposed to being FULLY at the mercy of a collision based system.

The “cut to the chase” bit here is that if a network system supports ATM, and we use ATM interface equipment to tie to that network, it is much more likely that the issues of delay and indeterminacy (also referred to as ‘Latency Thresholds’) are minimized greatly.

Reliability, etc…

Quick TRUE story to sum this all up…….

A bunch of Intercom Engineers went to a VOIP Telephony Conference in New York in 2001…audience was full of two types….”telephone types” and “data systems types”.

Telephone types talked about phone systems, reliability, up time, etc.. and said that they had a CUMULATIVE ACHIEVEMENT of something like 99.9999% up time.

Data types talked about system up time, network access time, etc. and talked about having a GOAL of something like 99.5% up time.

Doesn’t sound like a big difference until you subtract each of those numbers from 1.000000 and realize that the phone types were talking about down time of .000001 versus .005 for the Data types (5,000 times better for the phone folks).

You’ll remember that in Item #2 at top of article we talked about Broadcast and Professional Intercom Systems needing something like 99.999% up time – a bunch closer to “telephone types” than “data types”.

Item #2 (reliability) is critical. It is WAY MORE important than almost anything EXCEPT not having ANY communications at all. So…any current methodology for sending Broadcast and Professional Intercom signals over a data network of any type is going to be LESS RELIABLE than a direct hard wire (or dedicated fiber optic) connection.

<u>So what does all this mean….</u>

What this means is that VOIP, LAN, WAN, ATM are TODAY (and for the foreseeable future) appropriate ADJUNCTS to a reliable intercom system. Where the <u>only</u> (or best, or most affordable, or easiest to use) way to interconnect parts of a communications system is by one of these methods, then it is entirely appropriate to do so. Because in those cases, lower reliability, or lower bandwidth communications is preferable to no communications at all.

<u>And how do RTS Matrix Intercom Systems fit into all of this?</u>

Each of the techniques and scenarios presented above have been accomplished with RTS Matrices using a combination of RTS and industry standard data hardware.

RTS also offers a full line of VOIP interfaces to be used as ADJUNCTS to RTS Matrix Intercom Systems. From cards that plug into ADAM, to VOIP enabled Keypanels to stand-alone interface boxes, to virtual PC based Virtual KeyPanels, RTS offers VOIP solutions which can be very useful in certain circumstances, when the capabilities and limitations noted earlier are taken into account.

The RTS line of VOIP products are denoted as RVON (**<u>R</u>**TS **<u>V</u>**oice **<u>O</u>**ver **<u>N</u>**etworking) And embody many of the “G dot” CODECS described earlier and have been optimized for minimized delay (Latency), low data bandwidth, audio quality, and error recovery, so that they can make the most of the benefits of data networks while minimizing the bad effects of the limitations described earlier.

Summary

Use of VOIP for Professional Intercom systems is on the increase, and the reliability and performance are continually improving. There are many considerations to be made in choosing IF and HOW VOIP is implemented in specific installations. The basic consideration is if the convenience/economics/operating needs that can be met by using VOIP outweigh the requirement for failsafe 99.999% connections…..if “Yes” then VOIP is a good solution for that PORTION of your communications system.

RTS has a worldwide staff of knowledgeable sales professionals and systems engineer fully versed in VOIP and are available to assist with your design and requirements at any time.

Ralph K. Strader

Ralph K Strader

President, Critical Communications Systems

Telex Communications
May 2005