A Touchstone Technologies White Paper

1500 Industry Road, Suite H
Hatfield, PA 19440
USA
+1.267.222.8687
www.touchstone-inc.com

Things that go bump on your network: Five reasons why you need to monitor your communication networks

By: Mark A. Stacy, CTO

Copyright 2012 Touchstone Technologies, Inc.

Contents

Why monitoring matters	**2**
The top five reasons	**2**
Call quality problems	**3**
Network delay	**5**
Service levels	**6**
Insight	**7**
User satisfaction	**8**
Summary	**9**

Why monitoring matters

Today's converged networks provide unique challenges that didn't exist in the previous generation of communication systems. We will examine five major areas and the solutions that enable network engineers and administrators to overcome them.

The top five reasons

Below, we will summarize the top five reasons to monitor your communications network. Read further on in this paper for an in-depth analysis of these topics.

- **Call quality problems**
  There are many factors that can contribute to call quality issues. Traditional methods of determining how "well" a network is performing are only marginally applicable to real-time communications.
- **Network delay**
  One-way or "path" delay can be a real problem. It can greatly affect the ability to communicate effectively.
- **Service levels**
  Whether you are a service provider or consumer, the level of service is a yardstick by which quality is measured.
- **Insight**
  Once there is a problem, it's too late to fix it. An early warning system is invaluable in knowing what is going bump on your network.
- **User satisfaction**
  Ensuring the highest quality service allows your sales teams to have the competitive advantage. Whether as a consumer or a provider.

Copyright 2012 Touchstone Technologies, Inc.

Call quality problems

Let me begin by saying that I believe in, and embrace, this "new" technology. I remember my first encounter with it very well. Back around 1995, before I was in this industry, my friend Billy called me and said "Come over and check out this cool new software phone I found." As it turns out, Billy had "discovered" VocalTec's Internet Phone application and, a like-minded counterpart in Italy.

At that time, I was pursuing another business idea and likened Billy's discovery to Ham radio; just a worse quality version. Fast forward about five years and I found myself gainfully employed and working for a company that specialized in videoconferencing test solutions. They were working on active test solutions for H.323 had built a videoconferencing "bridging" application that analyzed H.320-based ISDN conferences and had decided to work on similar tools for SIP. That was my first task: build a software-based SIP testing solution.

At first, I didn't really understand why people needed to test or monitor their phone networks but they were buying these tools so there must be some reason, right? Then I found out first hand. I happened to be at a SIPit Interop event when I called home from the phone in my hotel room. The hotel was so new that I was the first person to ever stay in that room. My wife answered the phone and we were talking when she asked where I was calling her from. I told her and she asked me to call her from my cell phone. When I called back, she told me that the other "connection" was so bad that she couldn't listen to it anymore.

I did mention that she asked me to call her on a ***cell phone*** for a ***better experience***, correct? Well that was my first experience with a VoIP phone and it wasn't very good.

There are many reasons that call quality issues arise. Some can be attributed to anomalous conditions on the network and some can be an indication of pending network meltdowns. Some factors far outweigh others when it comes to voice quality problems but network conditions can vary wildly depending on circumstances.

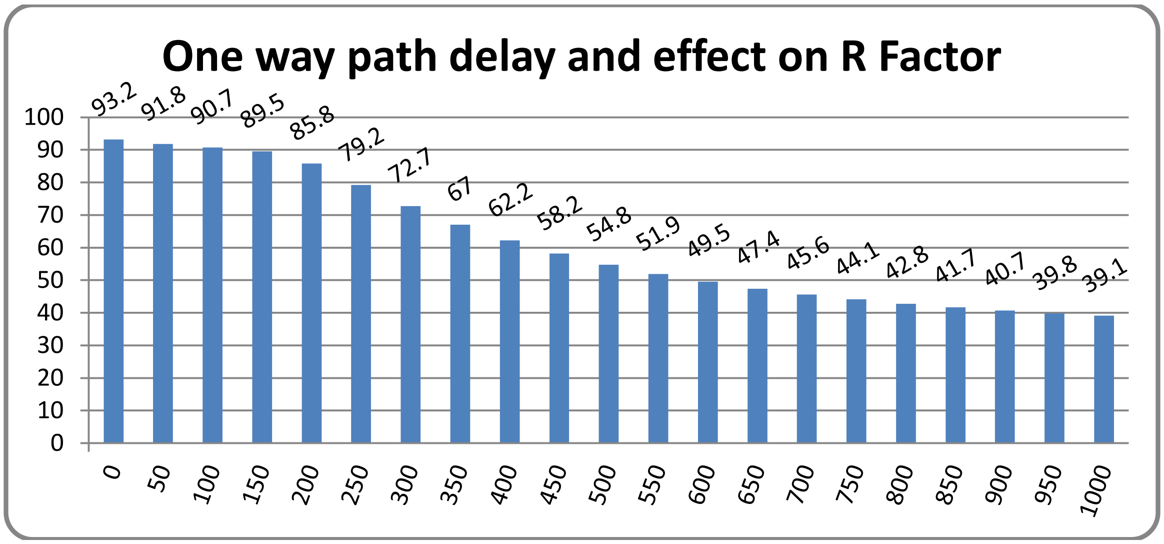
Let's look at the most commonly identified causes of VoIP quality issues:

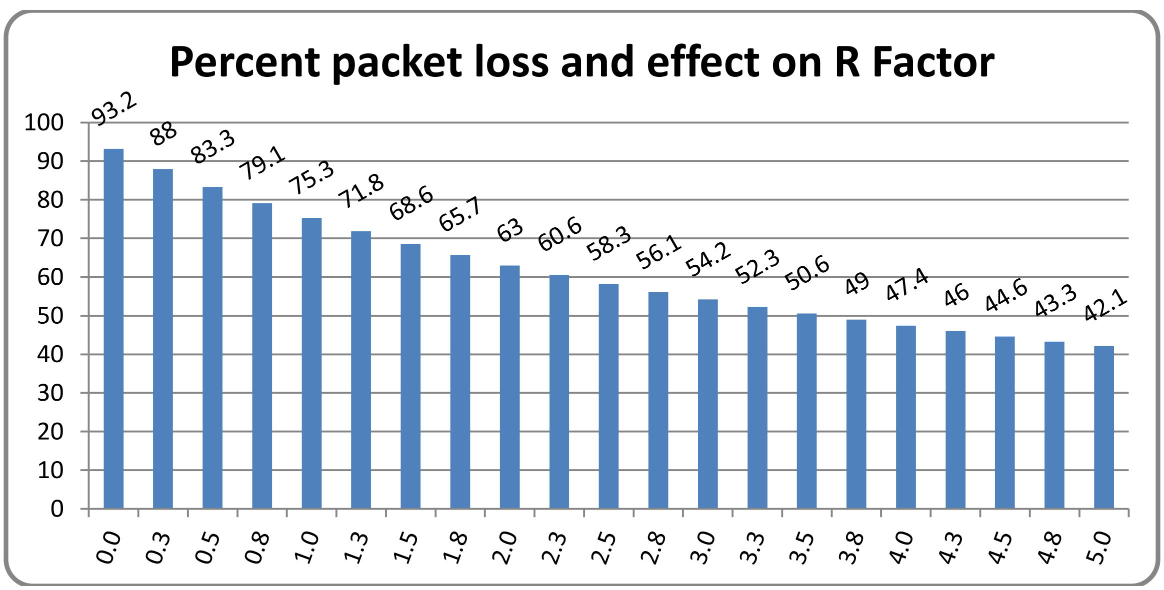
- **Jitter**
  Jitter is the deviation in the inter-arrival times between packets. Jitter is not the same as delay. You can have jitter when packets arrive ***before*** they are supposed to. It's rare, but it can happen. Session Border Controllers (SBCs) come to mind when I think of early (or negative) jitter.
- **Delay**
  We'll talk about this in some detail in the next section, but one-way or path delay can have a dramatic effect on the ability to converse naturally.
- **Packet loss**
  This is the VoIP killer. Many companies have tried to solve this problem and some have done very well. Ditech Networks is one company that has dedicated a lot of time, money, and resources to develop Packet Loss Concealment (PLC) devices that can interpolate and "solve" packet loss issues but at some point, those little lost packets will prevail.

Copyright 2012 Touchstone Technologies, Inc.

The ITU-T specification G.107 defines the E-model as “a computational model for use in transmission planning”. This document provides the mathematical formulae to determine an “R Factor” for the call from which we can derive MOS scores. While there is no input in the E-model for jitter, we can use this scale to show the impact of delay and packet loss on call quality. The following two charts demonstrate this. Chart A shows the effects of one-way path delay on the conversational R Factor. We’ve input a delay of zero milliseconds to one second (one thousand milliseconds). Chart B shows the effects of packet loss on the R Factor. We’ve input zero to ten percent.

An R Factor of less than 70 is generally considered unacceptable.





Copyright 2012 Touchstone Technologies, Inc.

Network delay
In a perfect scenario, there would be no network delay. However, if we roughly estimate the delay across North America at 3,000 / 186,000 = 0.016 seconds (the rough distance across the farthest two points of the Continental United States divided by the speed of light) you get 16 milliseconds (ms) as the theoretical minimum transmission time. In reality, it is closer to two to four times that. A sample of carrier data publically available shows average delays of 41-42ms in the U.S. , a New York to London average of around 79ms and a trans-Pacific average of around 104ms.

So how does this affect voice quality? Well according to these statistics and the G.107 e-model, it doesn't really in these cases. A delay of around 125m is the point at which the e-model begins to dip below the "gold standard" of an R Factor of 90. Delays greater than 125ms begin to impair not only the flow of packets, but the ability to converse naturally. Consider that you have an urgent message that you have to impart in a limited timeframe. When you call your party you hastily explain the situation and propose a quick solution. You wait a nominal amount of time and resume the conversation. The problem is that if there is a 250ms delay in each direction, there is always a half second gap between when you finish speaking and the time you would hear an instant reply. By not hearing a reply right away, you believe it is okay to resume the conversation when in fact; the other party is still trying to respond to your last comments. Basically, it's akin to the problem inherent in real-time communications on satellite-based systems.

So, without placing probes that measure network delay by synthetically injecting traffic onto the network, how can path delay be measured? The answer lies in the IETF (Internet Engineer Task Force) RFCs (Request For Comments) 1889 and (subsequently) 3550. In these documents, a mechanism for path delay measurements are built into sender reports that have a reception report block. Because these reports are time stamped precisely, we can deduce the path delay between devices to the point of measurement ***without adding any active devices to simply measure delay.***

Now, I have no issue with measuring path delay with active devices, but when it's inherently available, why not use that? It's a true measure of the media path, whereas active testing may take a completely different route. Which brings me to my next topic (and my issue with nearly all VoIP SLAs offered by most carriers)...

Path delay is not the same as a "ping" across the network. Roundtrip delay divided by two is not the same thing as path delay. And, just to be clear, a ping almost certainly will not take the same path or even be in the same service class as a media stream. A delay of 50ms in one direction and 400ms in the other is far more difficult to deal with than a 225ms delay in both directions.

So how is this applied in the real world? Network delay can be caused by router or switch congestion, bandwidth shortages, SIP trunk provider networks in general and is an indication of degradation of service either caused by internal network problems or service delivery issues. Having something that can detect and alert you to these problems can save a significant amount of time and effort; as well as the name calling and finger pointing that can occur with something as difficult to pinpoint as path delay can be.

Copyright 2012 Touchstone Technologies, Inc.

Service levels

Today's converged networks are a mixed blessing. They enable us to reduce costs by moving our separate voice and video communications to our data networks. The problem is just that: many of these networks were built as endpoints on the "information superhighway" and were meant to transport data and not necessarily real-time communications.

Data is ***not necessarily*** time sensitive, voice and video ***certainly*** are.

I have reviewed the Service Level Agreements (SLAs) of several North American Tier 1 carriers for their Business VoIP packages and initiatives. Remarkably, they are virtually identical to their ***Business Internet SLA***. Let me reiterate: guaranteeing and delivering time-critical applications such as voice and Unified Communications and Collaboration (UC&C) is ***not*** the same as delivering data.

Here is an excerpt from one of North America's largest carrier's Business VoIP Service Level Agreement:

> "The VoIP Jitter SLA provides that [the company's] contiguous U.S. Internet Network (as defined in the Guide) monthly jitter performance will not exceed 1.0 millisecond.
>
> Performance is measured by periodically collecting data across the contiguous U.S. Internet Network, from which a monthly average is derived."

Okay, we could agree that jitter of less than one millisecond on average would be acceptable, ***but it is an average of the entire network… over an entire month!*** And, as we mentioned above, jitter is not an input into the e-model so therefore it has no effect on the R Factor or the derived MOS score so, while low jitter is desirable it is not a guarantee of high quality. What about path delay? What about packet loss?

Well, that's interesting because the contract goes on to say:

> "The VoIP MOS SLA provides that [the company's] U.S. Internet Network MOS performance will not drop below 4.0 where MOS is calculated using the standards based E-model (ITU-T G.107).
>
> Performance is measured by periodically collecting data across the contiguous U.S. Internet Network from which a monthly average is derived."

Collecting data? What data? What about equipment impairment factors (***G.729 can't ever achieve a 4.0 MOS score***!) and packet-loss robustness factors (which are codec specific)? What assumptions do you input into the e-model about them? Hmm, that's about as non-descript as it gets!

Business communications are critical business

Nowhere in that particular SLA does it ever address call quality on a ***per call basis***. What is more important than that? Isn't that what we're trying to define in an SLA? By sampling the data, you are basically saying: "We make no differentiation between the secretary calling her child to see how their day went and a conference call between members of the board of directors. In fact, we may place more emphasis on her call than on theirs".

At the end of the day, especially as video-based services such as telepresence join the arsenal of the "conference room warrior", the next-generation SLA will be a per-call model and will take into account all of the factors which comprise a true, business-class worthy SLA.

So whether you are a carrier providing services or a consumer of services, you need to monitor your communications network for quality on a per-call basis to determine what the real state of your communications affairs are.

Copyright 2012 Touchstone Technologies, Inc.

Insight
In the world of test and measurement, nothing is more important than accuracy. False positives lead to unnoticed, undiagnosed and other bad un-things. False negatives turn relative calm into the Chicken Little syndrome: first one acorn, then two, then the sky is falling.

An accurate monitoring solution is not only critical for gathering analytical data, it is imperative for predictive alerts and alarms. It is the key component to managing your communications network.

A real-time communications monitoring solution should help to accurately and proactively identify anomalous conditions. In a monitored environment, network engineers can identify segments that aren't keeping up to specification either on a constant or on a periodic basis. A good monitoring solution would allow you to trend this data to identify when these conditions occur and potentially, where.

A typical service truck roll can cost a company upwards of $2,500 (USD). By monitoring and identifying problems automatically most problems can be diagnosed, and potentially solved, across the network without the necessity of a physical presence.

Furthermore to customer service, what could potentially be more beneficial to a company's reputation than answering a customer complaint with something like: "Yes Mister Jones, we're aware of the problem you had on your call at 2:32 pm. It appears as though the call as it was delivered to our network was below our threshold levels. Our system notified us of this within the first 30 seconds of the call and although our equipment did the best it could to improve the quality of the inbound audio stream, we have credited your account for the full 8 minutes and 48 seconds of the call".

While that may be an extreme example, having that kind of information can be critical to a company's customer-facing side. Another example of where a monitoring solution can provide forward-looking insight is in the area of trending. The ability to identify trends, either by time of day, day of week, by SIP trunk or any other imaginable combination of data mining can be highly useful to engineers doing network planning, or simply trying to get a handle on how things bump up against each other on the network.

One example that I can relate is that of a customer who had an existing problem prior to monitoring their network. Occasionally, during the middle of the night in a contact center, all of the phones would go silent. Calls dropped, phones lost dial tone and everything went silent for about a minute. Then everything would return to normal for several weeks and then ***something would go bump on the network again***. After installing the monitoring solution it became immediately apparent the next time it happened: a full-on automated FTP-based network backup kicked on at 3:00 am and occasionally, during peak usage times, it would bump up against the voice traffic and exceed the networks capacity.

Copyright 2012 Touchstone Technologies, Inc.

User satisfaction
At the end of the day, it's all about user satisfaction. Whether your perspective is that of a service provider, integrator, or consumer, this is the key. The more aware you are of what is really happening on your network, the better equipped you are to make informed and rational decisions about issues or planning.

As I mentioned previously, the SLAs that are being offered need to be carefully scrutinized prior to acceptance. They are Internet SLAs with an e-model calculation that is based on undisclosed network data that "is gathered periodically". How does that affect my phone service? For my money, I believe that if you cannot provide me with an SLA based upon a per-call model where I can see the following metrics, I'm not buying.

- Time/Date
- Duration
- Calling/Called party numbers
- CODEC Type
- Average Jitter
- Average Path Delay
- Total Packets Expected
- Lost Packets
- R Factor/MOS (conversational)
- Early/Late Discarded Packets
- Ontime Playout
- Gap Fill

So let's run through these to make sure we understand each one. I'll assume that the first three are self evident.

CODEC
The compressor/de-compressor that translates analog signals to digital packets and vice-versa. The CODEC is important to know because each one has different characteristics that dramatically affect the e-model.

Average Jitter
As mentioned previously, while not an input to the e-model, excessive jitter can have an impact on a device's jitter (or de-jitter) buffer's ability to maintain an acceptable rate of play out.

Average Path Delay
As detailed previously, path delay has an impact on the ability to communicate naturally. Delay values of more than 125ms can cause disturbances.

Total Packets Expected
Derived from the sequence numbers of the RTP media streams, this lets us know how many packets we should have received.

Lost Packets
Derived from the sequence numbers of the RTP media streams, this lets us know how many packets we lost. This is derived by subtracting the Total Packets Received from the Total Packets Expected (conversely, you can derive the Total Received by adding expected plus lost).

R Factor/MOS (conversational)
The conversational versions take into account path delay where the listening versions do not. Therefore, for transmission monitoring and planning, the conversational values have significantly more value than the listening versions.

Early/Late Discarded Packets
This lets us know how many packets a jitter (de-jitter) buffer would have discarded based on early or late arrival. These would not be able to be played out.

Ontime Playout/Gap Fill
The amount of time that the CODEC could play voice and the amount of time it had to fill with silence (or comfort noise).

Copyright 2012 Touchstone Technologies, Inc.

Summary

Test and Measurement solutions are not sexy. They tend to be the proverbial "red-headed stepchild", the one often considered last or only when needed; often summoned only after an "Oh no" moment. However, the increasingly complex environments which we are creating are going to demand these types of solutions as an integral part of their ecosystem.

The solutions need to be focused enough to understand the complexities, smart enough to detect the anomalies, and intelligent enough to determine hiccups from Pneumonia. A good IP communications monitoring solution should possess the following criteria at a minimum:

1. The ability to determine quality on a per-call basis with highly detailed metrics supporting all Key-Performance-Indicators (KPIs).
2. The ability to provide enough detail in its Call Detail Records (CDRs) to drive a billing system.
3. The ability to not only alert or alarm on metrics that fall below or exceed thresholds but the intelligence to know the difference between a short-lived blip and a full-on issue.
4. The ability to provide an aggregated database of call records that can be analyzed in general (how is my network performing?), by organization, or by individual location.
5. The ability to support carrier-class call volume requirements.
6. The ability to live in a vendor-agnostic environment by passively watching live traffic on a real-world network at any point on the network.
7. The ability to provide remote diagnostic information including but not limited to capturing and transmitting back to a technician a capture of the live network traffic.
8. The ability to be configured and upgraded remotely.
9. The ability to handle unusual network problems gracefully.
10. Last but most importantly, the ability to be trusted as an accurate measuring device.

For more information on Touchstone's products, please visit:

http://www.touchstone-inc.com

Copyright 2012 Touchstone Technologies, Inc.