NETEVENTS

GLOBAL PRESS & ANALYST SUMMIT

DRAFT

Opening Keynote Presentation: Hallmarks of the Impending Internet Revolution

Glenn Ricart Ph.D

Founder & CTO, US Ignite

Good morning, everybody.

Technology is a wonderful and interesting thing. About every 30 to 40 years, something major changes. It's an S-curve. If you look up, Everett Rogers has shown that the adoption of technology is at first slow, then very rapid. And then, at the conclusion, it tails off and another technology comes and takes its place. You can see this happening in television. First, black and white TV; about 40 years later, colour TV; about 40 years later, high-definition TV. Hopefully, it won't take 40 years to get to ultra-high-definition TV, but these things change every period of time.

And the Internet and Ethernet is about how old? You can tell from this celebration that it's about 40 years old. So it's about time. It's about time for some major inflections and some major change, and I think you're going to hear all morning about some of those major changes, some of the folks who are leading those changes at some of the leading companies and other places where that change is happening. I think that the changes are going to really be profound in the sense that they're going to impact people's lives. And I have begun working with a non-profit organisation, US Ignite, on how these changes might impact people's lives.

This is a picture of the South Auditorium of the White House when we launched US Ignite on June 12th, 2012. When US Ignite was launched, and US Ignite is about new applications of these new technologies, we had John Holdren, the Head of the Office of Science and Technology Policy for the United States. We had Subra Suresh, the Director of the National Science Foundation. We had Julius Genachowski, the Chairman of the Federal Communications Commission. We had Larry Strickland, the Head of the National Telecommunications and Information Administration. All of these folks were there to celebrate the launch of a new initiative and non-profit

organisation. US Ignite is not a government organisation. It's a non-profit. It works with the government but it works independently and together with other parties.

So what is US Ignite? It’s a public/private partnership. It's government working together with the private sector in a non-profit way.

We have a major, simple goal, to take these new technologies and turn them into new applications to benefit the citizens of the world. We want to see next generation applications. And you see there, in orange, the areas that are being impacted. At the top, clean energy. Going around clockwise, advanced manufacturing, public safety, transportation, healthcare, education, important things that impact people everywhere. What can these technologies do for them?

And we're working together with a number of constituencies. We work with industry, with government, research and education, foundations, cities, and other organisations who are allied in this goal to help make the next generation of technology turn into beneficial applications for citizens of the world.

We have three goals at US Ignite. The first is to create 60 compelling, transformative applications based on things you couldn't do today, based on new technologies, things that you've been hearing about and will hear about today: software defined networking, local cloud computing, taking gigabit to the end-user, reducing latency, things that are going to change the way the internet works today. Second, that we get 200 community test beds, 200 cities, who are eager and willing to adopt these new technologies and their applications. And third, to coordinate best practices among these communities, among the industry partners, and to make sure that government, industry and communities are working together to make this goal happen.

This next slide shows the partners for US Ignite. You'll see here some industry leaders. You'll see some communities. You'll see a number of folks who are working together to make US Ignite happen.

The next generation infrastructure I think will have three big, important spokes. Number one is the use of virtualisation and software defined networking. Those are not quite the same, but they're close enough that I'm going to put them into one category for this morning. And for all those who will come later and say they're not the same, I agree, they're not quite the same, but I still put them in about the same category. These are things that are going to change the business model for how networks work and are charged for.

The second is local cloud computing, and I will speak about this in my talk. And the third is that we don't need necessarily more speed. We need more responsive services to the end-user, and I'll talk about that also. So I'm going to mainly speak about numbers two and three, because I believe other folks today will talk about the first one.

Okay. So I thought I would tell you my punch line first. The whole talk revolves around three changes: first that today we think that more bandwidth is bigger, and I think that's going to change. The new version is going to be more responsive is better. And I use the word locavore. That word is used in the food industry to mean eating

foods grown near you, and I'm using it to mean consuming computer cycles and networking located near you.

Second, today we think about putting things into a massive data centre, into a massive cloud, concentrating many, many, many computers all in the same place. I think that is going to change, and we're going to see the cloud coming closer to you to be more responsive and better support real-time applications.

The third is that today we think that if we have a big enough internet pipe, that will be fine and traffic won't conflict, and that's true. But I think over time, and especially for wireless, it's going to be much more important to use our resources more efficiently, and tomorrow multiple, dynamic virtual pipes will be purpose built, will be bespoke per application.

So those are the points I'm going to make. And now that you know all of them, I'd like to go and start with more bandwidth is better. I have an internet connection in my home which is about 20 to 25 megabits per second, from Comcast, and of course I would like it to be faster. But would it really do much good? I did an experiment.

Here's the experiment that I did. I measured how many web page loads per second I could get. On the right-hand side is my connection unaltered, and you can see I did this experiment a number of times. It's not always the same. You can see the number of web page loads there. Okay, pretty decent. It's about one per second, a little bit less than one per second. That's very typical. Then I went and I found an old Ethernet hub in my closet which was only 10 megabits. And so I put that into the circuit. I did a speedtest.net. It said it was about 6.5 megabits.

I then took the same readings running it through that hub to slow things down, and you can see that I got fewer page loads per second, as you would expect. But I didn't see a huge dropoff. From speedtest.net, the speed tripled, between 6 and 22, but I didn't get three times the number of page loads, did I? In fact, it only went up 23%. So clearly something else is going on.

Why is it that when I tripled the speed I didn't get triple the page loads? Well, there's actually a very good reason. This is what happens between my computer and the place giving me all those page loads. There are all these routers and links in between, and in fact the only one whose speed I changed was the last one.

Now, of course the weakest link will determine the speed, but as these things go, it had to go through all these other links in order to get to the server which was in a massive date centre. Where was that server? I ran a technical program which allows me to see where all of those things were. So I tracked down each one of those things, and in each one of those things you can see the number of milliseconds, over three different tries, and where it went.

I live in Salt Lake City, so the first things you see up there are Salt Lake City. If you know your American geography, it then sent it east to Denver, Denver sent it east to Chicago, Chicago sent it east to Washington DC. It rattled around Washington DC for a while, was handed off to Yahoo. The website I was going after was finance.yahoo.com. Yahoo went and took it through a series of cryptic links, but I did

do some work to find out where these were located, and it sent it all the way back across the United States to right here, Sunnyvale, and that's where the server was that provided the service.

So my traffic went to Washington DC, rattled around for a while, came back to Sunnyvale, and you can see that it took about 65 to 80 milliseconds. The amount of time that people can distinguish that something takes is about 50 milliseconds, so this is a noticeable amount of time just to go and get my request to Yahoo and get it back. That's not providing for any time for the server to do its work of actually fetching the information I'm interested in.

So that's why the number of page loads does not change dramatically, even though my end system was much slower when I changed the speed down to an effective 6 megabits per second. So going to a gigabit on that last link is probably not going to make that much difference.

This next slide shows another experiment. This one was done by Microsoft Research. It shows you how long it takes to load a web page with a bandwidth of 1 megabit. You can see all these are all about the same. When they did it again at 10 megabits, you can see that when they invoked something called SPDY, speedy, which goes back and forth fewer times, that the bottom two lines there are lower and it takes less page load time.

So we think that just from these two graphs the point to take is that as the speed goes up, the importance of the length of that chain goes up. As the speed goes up, the fact that there are more links in the chain makes more difference. So as the internet grows and as it gets faster, the length of that chain makes more difference.

So what's the obvious thing that's going to happen in the future to have a more responsive Internet? Make this chain shorter. That's about the only thing we can do. The number of times a packet gets handled makes a difference. So we want to go to a local cloud or locavore computing.

How soon will this happen? Well, interestingly enough, it's already happened, and it’s starting to happen more broadly. AOL has begun putting in what they call micro data centres - this is a micro data centre they put in in Dulles, Virginia - to be closer to the end-user so they can be more responsive. And studies have shown that people buy more and that people will be able to stay with your service longer if you're more responsive. So one of the things that Mike Manos, the Chief Technology Officer of AOL Services, has said is that we expect that micro data centres will enable us to roll out five times the amount of total compute capacity in less than 10% of the cost and the physical footprint of a traditional data centre based deployment. And if you want to go look this up on the web, just Google or Bing, your choice, micro datacenter and AOL, and you'll find this quote and the other pictures.

So this is what's going to happen. It's also happening in research and education in the United States. There's a research project by the name of GENI. And the GENI Project not only facilitates this local cloud computing, but it also facilitates federating these local clouds, bringing the local clouds together so that when you need more compute

than one of these local data centres can provide, you can do that by aggregating them together. So you can have the best of both worlds. For simple things, very responsive; it's close to you. For more complex things, you can federate services from multiple local clouds. It will take a bit longer, but still you'll be able to have that ability to aggregate lots of compute power on what you need.

Okay. So what I've been talking about is: instead of more bandwidth is better, more responsiveness is better. And to get to more responsiveness, we need to decrease latency, decrease jitter, and a major step to be able to do those two things is to be responsive by having local facilities. There are other ways of reducing and jitter and those are important also, but I don't have time to talk about all of those this morning.

Let me move on to the second big point, which is that today we tend to think that, with a big enough pipe, it doesn't matter, I can just toss all my devices and applications together on the same Internet link and it’s all going to work fine.

Here’s a slide showing all the Internet-connected devices at my house. The smartphone and laptop travel with me, but the rest of the stuff stays home. And you can see that I have a couple of desktops, one of which I bought for $200 but still does a lot of interesting things like record my television. I have my own home-built television recorder. There’s a network printer. There’s Vonage VoIP. So, all of these things are contending for the bandwidth on my internet connection.

And what happens is the tragedy of the commons, the traditional, typical tragedy of the commons. Today's internet only works well because the utilisation is so low. If we had high utilisation, it wouldn't work well at all. In fact, I have Comcast, as you know, and Comcast says that if you use more than 4% of your capacity on a long-term basis, your usage is excessive and they have the right to go and charge additional money because you are using excessive bandwidth.

On my phone, which happens to be a Verizon Wireless phone, they begin slowing down my data if I use more than 0.03% of the bandwidth available to my phone. Obviously spectrum is limited and the more data I send, the more of that spectrum I’m using. To be fair the 0.03% is on a 3G data plan, if you have unlimited data usage, and it is a long-term average. But the utilisation being low is very important.

Applications cheat. Things like Google Search will violate the rules of sharing the road. They will go and grab more than the bandwidth that they're entitled to. Voice over IP services will try to grab more than they're entitled to. Things that think they are better and want to be more responsive than other applications will grab more than they're entitled to. And there are entire companies, like Vidyo and OpenClove, who are based on the proposition of making video work well even when they're having to fight and claw for the bandwidth they need.

So this will eventually come to a head. And the question I would ask is what would data centres do? What would data centres do in this case? And the answer is, of course, what the data centres would do is they would go virtual. They would have a virtual server per application. Instead of sharing all these apps and putting them all on one server - we used to call that time sharing - they now go and they have a server per

application, a virtual server per application. The server configuration is matched to the application it's running. They allocate these virtual servers as needed, dynamically, based on the load that's being presented. It provides for more fault resilience.

And remember I suggested another difference is the business model; it's easy to bill the server to the application. You're dedicating a virtual server to an application, so you say "hmm…, we're going to bill this server to the application." And the incremental cost model means that there are more total dollars, because you get to charge every time more work gets done.

So, remember, what would data centres do in this situation we're about to face on the network? They would go virtual. What will happen on the network? I think it will also go virtual. Instead of a virtual server per application, you have a virtual network per application. You create a virtual network for the application. The network configuration is matched to your application. You allocate as many of these as you need to run your application dynamically, based on the load. It is more fault resilient. It's easy to bill the network to the application and there's incremental cost for service providers. And in fact it might not be the end customer that pays. It might be somebody who is bundling the service.

So I think this is what's going to happen. Here's an example of a typical home. Ordinary internet comes in. What else might you want to do? You might want to add another virtual network, perhaps to provide telehealth services. Somebody is elderly. They may have a diabetic ulcer which needs to be checked up on, and you'd like to have privacy and you'd like to provision a high-bandwidth, high-quality video link from the home uplink on demand. Maybe you would like have a public safety application, where every home has WiFi that can be accessed by public responders or by citizens who are reporting an emergency and want to beam, for example, video of that emergency.

How about interactive online education? The MOOCs that were talked about yesterday. How can we get high quality, low cost, specifically for that piece of education? Or what if there is something that is very life dependent? What if there is dialysis control, that you need a high-reliability connection? You need to make sure this connection always works, because dialysis machine at home needs to be supervised. So these could all be virtual networks using network function virtualisation - I'm sure you will hear that term several times today - to be able to provide multiple services into the same home over the same pipe.

So, again, here's my summary. I think that tomorrow responsiveness and low latency is going to be the order of the day. It's going to be provided in part by using lower latency equipment and products. They exist already for the finance industry. I think they're going to come into more widespread use. I think that keeping things local and being a locavore about your computing is going to be important. The cloud comes to you to be more responsive. And we'll end up with multiple, dynamic, virtual pipes which are purpose built and designed to run with the properties of each application.

So, if you would like to hear any more about this next generation infrastructure, I invite you to come join us at an application summit, June 24th to 26th in Chicago.

We're going to give demonstrations of applications in education, advanced manufacturing, healthcare, and public safety, based on the new technologies I've just talked about, and applications which, as I said earlier, are only possible because of these new technologies being used.

So that is what I wanted to talk about this morning. Thank you very much.

Manek Dubash - Editorial Director, NetEvents

Thanks very much for that, Glenn. An interesting range of new ideas I've not heard before. I'd like to say first of all, if I can, that this sounds like a very US centric -- perhaps a US centric problem. I don't know. Perhaps a US centric solution. We have an international audience here. How does it play elsewhere?

Glenn Ricart

Well, first of all, I believe that the issues and the problems are universal. There are a few communities who have a longer runway before they have to bump into them. For example, in South Korea there is so much bandwidth that abundance will continue to work for more years than in places where there's not much abundance. But it is true that lots of folks are worrying about this.

Last month, I was meeting with the European Commission in the Future Internet Assembly in Dublin, Ireland. Ireland currently holds the presidency of the EU. And there's a public/private partnership now being formed in the EU. It's called the FI-PPP. And if you go to FI-PPP.eu, you can see the public/private partnership that the EU is creating to stimulate advanced applications. We've also seen that in Japan there is a similar next generation project to GENI. They're also creating a public/private partnership to do the same kinds of things as US Ignite is intending to do in the United States.

There's been great interest from Australia. I have no doubt something will happen there as well. And I believe other countries will also come along. In the Southern Hemisphere, Brazil has been very active in this. And there's also lots of interest from RedCLARA, and Argentina and Chile have expressed a lot of significant interest in making these things happen in their areas as well. One of the demonstrations that you're going to see at the US Ignite summit comes from Canada. And a 4K video feed from Poland, with a virtual 3D augmentation coming into the Electronic Visualization Laboratory in Chicago using these new advanced technologies.

Manek Dubash

Okay. Thank you. So I'm also interested in where's the money coming from. At the moment, things like micro data centres or data centres close to where the data is required tend to be paid for by people like the finance industry, who can afford to do that and they absolutely want the minimum latency. Now, that's all very well and good, but they've got deep pockets; most of us don't.

Glenn Ricart

Indeed, and that is why, instead of giving you a finance example, I picked on a different example. I picked on AOL. AOL does not have that kind of deep finance interest and yet they are finding that they can increase their revenues by going to this locavore model, and I believe that you're going to see more and more adopters. So this is a very exciting event today. I was talking about the fact that things are going to be changing and be very different going forward. I think this is one of the things that you'll be able to track for the next two or three years and you'll see a big change.

Manek Dubash

Okay. Thank you. Questions from the floor. Yes, Alan.

Alan Weissberger - Chairman and President, IEEE Communication Society

Alan Weissberger, IEEE Communication Society Center, Clara, California. I'm very intrigued with this concept, Glenn, but I have a couple of issues with a slide you showed on the different types of applications. I'll ask them both and you can respond. How are you going to get that ultra-reliable connection for dialysis control or any other type of automated healthcare?

The second one is observation about the community WiFi application. In the Greater Bay area, we were promised three separate community WiFi networks. One was for San Francisco, one was for the city of Santa Clara and one was for the whole Bay area. They've all not materialised. The only one that has persisted is the Mountain View network by Google. So how is that going to happen if in Silicon Valley in the Greater San Francisco Bay area, the hub of technology, we couldn't make it so? Thank you.

Glenn Ricart

Certainly. Let me address those in the order you presented them. The first one was the question on how do you get ultra-reliable connection to the home for something like dialysis. And I have to be a little bit more technical, but you're IEEE and I'm an IEEE senior member so I think we can go right ahead and talk about the fact that within the main part of the network we'll send a packet down multiple paths simultaneously, to provide for the fact that if there's a failure in any one path the packet will be delivered on the other paths.

This is being worked right now in a demonstration project by Doug Comer, a name which should be familiar to many of you, including the international press, together with George Adams and John Geske of Kettering University in an advanced manufacturing scenario. So that's how it’s done within the main part of the network, where you can have multiple paths at the same time. Going into the home, there’s only one pipe, but you can still make sure that other traffic doesn’t interfere. I choose to think of it as a software defined network that has priority separation and guaranteed bandwidth, so we know that the person playing Xbox games in the basement will not be interfering with the communications with the dialysis machine on the second floor.

The second question was about community WiFi. So why the heck should community WiFi be any more successful than it has been in the past? And I argue because it's going to be built out of what's already there. If you go driving down your street, is there any chance that you will not be within WiFi range of at least one access point? No chance, right? Everybody's home has one. You may not be able to access it. What if there were a requirement for a second SSID which is used only for public safety purposes, which is bound securely, perhaps cryptographically, to an application that is provided as part of the community emergency response system that could attach to any WiFi that it can see from any of the major providers in that neighbourhood, and therefore be able to take advantage of that access?

So I think there's two factors that would make it successful in the future. One is it's not general open access; it's just for public safety. The second one is that it's depending only on what's installed already.

Manek Dubash

Although I'd argue that actually the geeks would actually replace their service provider's router with something rather more powerful and therefore cut off that route of access, and they're more likely to have the fat pipes.

Glenn Ricart

And I think that they are about 2% of the population even in Silicon Valley.

Manek Dubash

You could be right there. Okay. More questions? Yes, sir.

Phil Keys - Principal, Maido Media

Phil Keys, Maido Media. I think a lot of the presentation was more around how to handle high bandwidth applications. Could you make some comments on if there's anything in here that would help Internet of Things type of applications where you have lots of very low -- a small bandwidth but lots of them trying to head upstream?

Glenn Ricart

Right. I think that smart things are definitely going to take off. I'm on the Kick-starter list for some new Spark Devices. My son's got an Arduino running and also a WiFi Wireless Scale. So I think this is definitely happening. We will definitely see the Internet of Things. The reason I didn’t address it is because, for the most part, you can run the Internet of Things perfectly well on today's internet. It is a new, it is happening, but I'm trying to address things that, remember, wouldn't be possible if we didn't have an advanced technology to support it. And the Internet of Things is today occupying small enough bandwidth that, in general, it can be done with today's internet.

The one exception I will make is that some medical devices need, in the United States, to be compliant with a law called HIPAA, the Health Insurance Portability and

Accountability Act, and that requires for extra security for those connections. And although there are a number of different ways of providing security that would be HIPAA compliant, one way to increase that security is to use software defined networking or virtualisation to put it onto a separate virtual network, in the way I showed it on my slide.

Manek Dubash

Okay. Thanks, Glenn. I think we've got time for one more question. Bob. Actually, I'll come to you in a second.

From the floor

If I understand correctly, you've got one relatively fat pipe which you partition in order to handle the different services. Not being rude, but that isn't exactly a new development, and quad play, triple play to the home use that technology. What is to stop people grabbing some of that extra data, that extra capacity?

And at the beginning you were talking about apps, and several of them, like the medical, for instance, are being done at the moment by the MTOM community. Do you have MTOM players on board in your organisation?

Glenn Ricart

I couldn't quite hear all of the question. I believe it has to do with partitioning of a pipe is not new, which is correct, but the partitioning in this way is new and the ability to more easily bill that to a particular customer is comparatively new, although of course, as you pointed out, things like MPLS can already partition things.

From the floor

Yes, but the other quick thing, Manek, is what's to stop those people grabbing some of that capacity? You know, the bad guys, the Skypes and everything of the world. And the third thing, quickly, is those apps you had at the beginning. Many of those have been realised at the moment by the MTOM community.

Manek Dubash

The end-to-end community.

Glenn Ricart

What about the end-to-end community?

From the floor

Machine to machine.

Glenn Ricart

Machine to machine? Those are important.

From the floor

So, two questions. One is what's to stop apps like Skype grabbing all the bandwidth. And the other one was --

Glenn Ricart

Right. So what's preventing Skype from grabbing all the bandwidth? I believe that, as we go forward and have these multiple virtual networks or partitionings of any kind, that you need some way of managing those partitions, and there are today some ways. For example, in Coax there's policy based management to divide up the bandwidth on DOCSIS 2.0 and higher networks. But I think what we're going to find is that the things we used to do for operating systems and time sharing will reappear for networks, so that we can go and better manage according to priority, price, utilisation, to make better utilisation of our pipes.

Manek Dubash

Although I would say, on that point, that my router already allows me to do that, to prioritise different applications.

Glenn Ricart

You can prioritise, but I dare you to go and say give this between 10% and 15% but no more than 15%, and leave at least 10%. You can probably say this is first priority, second priority, third priority.

Manek Dubash

Well, actually, my router is quite sophisticated and I can do stuff like that, but it's not that common.

Glenn Ricart

Excellent. You can tell me afterwards.

Manek Dubash

Okay. It's an AVM FRITZ Box. Wayne.

From the floor

Yes. Just a quick question. Don't CDNs already take care of the local services that you're referring to as something we should be doing?

Glenn Ricart

Are your contents static?

From the floor

No, but content management networks aren't always static either.

Glenn Ricart

That's true, but when you need some computation, for example you're doing a 3D model and you'd like to be able to rotate that model in real time, you need something that's more than a CDN. So, in a way, the locavore network I'm talking about is the next evolution of the CDN, if you'd like to think about it that way, where you now have programmable modules you can download. Not only is the content cached, not only is it replaceable static based on being able to fetch pieces of the page, but it's also dynamic in the sense that you can go and insert arbitrary code to do arbitrary things. An example I'm using is 3D rotation. Make sense?

Manek Dubash

Okay. I think we're going to have to cut it short there. Glenn, thank you very much.

Glenn Ricart

Thanks for the great questions.

[End]