Routing (Packets) vs. Gatewaying (Messages)

I recently attended a panel at the http://www.icce.org a panel titled "Next Generation Mobility". When Q&A time arrived the moderator said that there were no wireless microphones available. I thought that strange because we were all carrying very powerful devices festooned with radios. Each mobile phone can, in fact, be a wireless microphone.

As I listened I to the panelists I was increasingly aware of the implicit acceptance of assumptions that no longer made sense. One topic was how to add more radios to a device that already had seven or more radios! And then I thought about the advertising slogans that said that it didn't matter which ISP you used – they all brought you to the same Internet. Each antenna seems necessary when you focus on the narrow purpose of each chip but when you look at the device the distinctions between the paths the bits take are arbitrary.

It's useful to look at the difference between connecting devices to a computer using USB versus Ethernet. I cannot use a USB device unless I have a driver for that device for the specific platform (operating system and machine) I am connecting it to. To connect a device over Ethernet I don't need any of that – any program can connect to the device directly. In fact the device doesn't even to be nearby – it can be connected via a Wi-Fi router even if the machine only has a wired Ethernet connection. With USB the device cannot be very far away. USB relays have been troublesome and have failed in the marketplace.

Why does USB require installing a different driver for a Samsung phone versus a Motorola phone just to access information on the SD card? Part of the reason is historic going back to the days when computing was expensive and you needed to use the CPU in the computer to make your printer or scanner work. Today the need for such software shims (drivers) limits our ability to connect devices.

With Ethernet (or Wi-Fi) none of the elements along the path need do anything more than pass along packets. The protocols require that the devices at the end points of the relationship handle packet losses and delays gracefully.

This ability to focus on the relationships between the end points and not have to deal with intermediaries is the big idea of the Internet but isn't really about telecommunications networks. One way to think about it is that all networking is social and exchanging bits is just a means we use.

It is important to distinguish between the powerful idea behind the Internet – the ability to ignore the details of how the bits are exchanged – with what we do with that ability. The Web is an important example. We can access web sites by merely typing in an address because we don't need to worry about who is providing the path. It is a very simply idea and it doesn't matter if the web server is hosted by a very large provider like Amazon or a small device in your home.

What makes the web such a powerful example is that I can sit in my house, type a URL and find myself connected to a site anywhere in the world without having to be aware of third parties in the middle. There is no difference between a server sitting in my home and one on another continent. In fact most PC's come with a web server built in as standard software! If you use Skype or other Voice over IP application on your device then it contains a server for voice protocols available to anyone anywhere whom you allow to connect.

Yet connecting the simplest Bluetooth device is impossible without the right devices in the middle explicitly paired and connected using a predefined profile. And if that intermediate device is a cell phone it must, in turn, have the right carrier available and the carrier must approve the particular applications.

Whew, I'm tired just describing all those steps along the path. No wonder applications like connected healthcare seem so difficult.

The idea that you only need to be concerned with what happens at the end points of a relationship is basic to the understanding of the Internet and is known as the *End-to-End Argument*. To fully appreciate the paper we need to understand it as an insight inspired by the Internet rather than an architectural document. We need an articulate understanding of this principle in order to explain the counterintuitive aspect of the Internet.

We're accustomed to designing systems for specific applications such as telephone calls. The Internet was created by software people who could create their own solutions. All they need was a common medium for exchanging packets. The first applications were simple enough – just exchange files. But the real value came from applications we could not anticipate such as the web.

The Internet is transformative because it gives us a blank slate as a resource. Rather than depending on lower layers we have to discover what is possible with the resource.

One reason this works so well is that as we discover new applications we also create a market for more capacity which, in turn creates more opportunity.

Meaning comes from context and needn't depend on the intermediaries understanding the message. You don't need an intermediary to understand what "Wikipedia" means. It is only necessary that the person you are talking to is aware of it. Another nuance of the Internet is that we don't actually pass a copy of the encyclopedia along the path – we just pass the reference and then "look it up in Wikipedia".

When we build smarts into the infrastructure we cede control to intermediaries and this makes it difficult to reinvent our environment.

Phone calls are a good example. As Skype shows we don't need to have a phone company even know that bits are part of a phone call. And yet many still commonly accept the idea of needing a phone company and paying them for phone calls. Why should we care or even know whether a phone call is carried on a wire or a fiber or over a wireless link? Yet we have entirely different regulatory and business policies based on the accidental properties of the way a particular phone call is handled – very strange when you think about it.

We're accustomed to building solutions yet VoIP works so well because we didn't built for VoIP into the Internet. Instead we've taken advantage of the capacity provided for the web. Not only is this counter-intuitive but it is a vital part of the dynamic. By taking advantage of an opportunity we are not beholden to intermediaries. What makes this difficult to accept is that there are no guarantees that VoIP will work.

This creates a conundrum for policy makers who are used to making promises.

Even more so when I see an advertisement in which a carrier claims telemedicine is a product of their network. The reality is that the value is created outside of their network and, to a degree, despite their need to make the network a profit center. No wonder they emphasize remote video rather than the mundane day-to-day medical information which, in fact, doesn't need much capacity at all.

I use the term packet routing for exchanging bits without their interpretation. In effect the bits find a route through the network just as a car finds a route through the road system. There's an additional twist in that the Internet doesn't require every packet make it through –we only ask for "best efforts". The intermediaries can't do more than best effort because they don't have a view beyond the packets in isolation. Thus the packets taking other routes may indeed go astray – something the owner of a single route can't know.

We've noted that applications can be resilient in the face of lost packets. This is in sharp contrast with approaches that depend on the gateways understanding message formats, and in some cases, imposing policies. Since the smarts are inside the network it isn't easy to route around such failures.

We can understand why this hands-off approach doesn't appeal to service providers and chip manufacturers because their specialized skills are no longer the primary source of value. Yet, for society as a whole it is important that we create the opportunity for all of us to create value rather than limiting ourselves to the "smarts" that others provide.

After writing this essay I was at a store and looked at the label on an HDMI cable which said it had a reverse channel for audio and supported Ethernet. This is very strange: if bits are bits why does a cable tell me what applications it supports? All that matters is that it carry bits and the more capacity the more I can do with it. This represents a very simple shift in thinking with profound implications that I plan to explore in future essays.

The Internet is "merely" this principle writ large. TCP is a sophisticated protocol for sharing the common facilities without depending on intermediaries doing anything more than routing packets. If we are to realize the benefits of that accomplishment we need to understand how routing IP packets differs from traditional telecommunications which was limited by intermediaries whose business depending on carrying valuable messages.

We miss the point of the Internet when we focus on the artifacts of the network – the wires and the gear. The Internet is about transcending such details and shifting giving us the ability to create own solutions – even before we understand the possibilities. This is process of discovery which drives our society forward.