May 2020
Geoff Huston

New IP and Emerging Communications Technologies

A "New IP" framework was proposed to the ITU last year[1]. This framework envisages a resurgence of a network-centric view of communications architectures where application behaviours are moderated by network-managed control mechanisms.

It's not the first time that we've seen proposals to rethink the basic architecture of the Internet's technology (for example, there were the "Clean Slate" efforts in the US research community a decade or so ago) and it certainly won't be the last. However, it this New IP framework is very prescriptive in terms of bounding application behaviours, and it seems to ignore the most basic lesson of the past three decades of evolution: communications services are no longer a command economy and these days the sector operates as a conventional market-based economy, and this market for diverse services is expressed in diversity of application behaviours.

What this market-based economy implies is that ultimately what shapes the future of the communications sector, what shapes the services that are provided and even the technologies used to generate such services are the result of consumer choices. Consumers are often fickle, entranced by passing fads, and can be both conservative and adventurous at the same time. But whatever you may think of the sanity of consumer markets, it's their money that drives this industry. Like any other consumer-focused services market, what consumers want, they get.

However, it's more than simple consumer preferences. This change in the economic nature of the sector also implies changes in investors and investment, changes in operators and changes in the collective expectations of the sector and the way in which these expectations are phrased. It's really not up to some crusty international committee to dictate future consumer preferences. Time and time again these committees with their lofty titles, such as "the Focus Group on Technologies for Network 2030" have been distinguished by their innate ability to see their considered prognostications comprehensively contradicted by reality! Their forebears in similar committees missed

[1] Telecommunication Sector Advisory Group (TSAG) contribution T17-TSAG-C83 [C83], presented at the September 2019 TSAG meeting.

computer mainframes, then they failed to see the personal computer revolution, and were then totally surprised by the smartphone. It's clear that no matter what the network will look like some 10 years from now, what it won't be is what this 2030 Focus Group pondering a new IP is envisaging!

I don't claim any particular ability to do any better in the area of divination of the future, and I'm not going to try. But in this process of evolution, the technical seeds of the near-term future are already visible today. What I would like to do here is describe that I think are the critically important technical seeds any why.

This is my somewhat arbitrary personal choice of technologies that I think will play a prominent role in the Internet over the next decade.

The foundation technology of the Internet, and indeed of the larger environment of digital communication, is the concept of packetization, replacing the previous model of circuit emulation.

IP advocated a radical change to the previous incumbency of telephony. Rather than an active time switched network with passive edge devices, the IP architecture advocated a largely passive network where the network's internal elements simply switched packets. The functionality of the service response was intended to be pushed out to the devices at the edge of the network. The respective roles of networks and devices were inverted in the transition to the internet.

But change is hard and for some decades many industry actors with interests in the provision of networks and network services strived to reverse this inversion of the network service model. Network operators tried hard to introduce network-based service responses while handling packet-based payloads. We saw the efforts to develop network-based Quality of Service approaches that attempted to support differential service responses for different classes of packet flows within a single network platform. I think some twenty years later we can call this effort a Grand Failure. Then there was virtual circuit emulation in MPLS and more recently variants of loose source routing (SR) approaches. It always strikes me as odd that these approaches require orchestration across all active elements in a network where the basic functionality of traffic segmentation can be offered at far lower cost through ingress traffic grooming. But, cynically, I guess that the way to sell more fancy routers is to distribute complexity across the entire network. I would hesitate to categorise any of these technologies as emerging, as they seem to be more like regressive measures in many ways, motivated more by a desire to "value-add" to an otherwise undistinguished commodity service of packet transmission. The longevity of some of these efforts to create network-based services is a testament to the level of resistance of network operators to accept their role as a commodity utility, rather than any inherent value in the architectural concept of circuit-based network segmentation.

At the same time, we've made some astonishing progress in other aspects of networking. We've been creating widely dispersed fault tolerant systems that don't rely on centralised command and control. Any student of the inter-domain routing protocol BGP, which is has been quietly supporting the Internet for some three decades now, could not fail to be impressed by the almost prescient design of a distributed system for managing a complex network that is now up to nine orders of magnitude larger than the network of the early 1990's for which is was originally devised. We've created a new kind of network that is open and accessible. It was nigh on impossible to create new applications for the telephone network, yet in the Internet that's what happens all the time. From the vibrant world of apps down to the very basics of digital transmission the world of networking is in a state of constant flux and new technologies are emerging at a dizzying rate.

What can we observe about emerging technologies that will play a critical role in the coming years? Here's is my personal selection of recent technical innovations that I would classify into the set of emerging technologies that will exercise a massive influence over the coming ten years.

Optical Coherence

For many decades the optical world used the equivalent of a torch. There was either light passing down the cable or there wasn't. This "on-off keying" (OOK) simple approach to optical encoding was continuously refined to support optical speeds of up to 10Gbps, which is no mean feat of technology, but at that point it was running into some apparently hard limitations of the digital signal processes that OOK is using.

But there is still headroom in the fibre for more signal. We are now turning to Optical Coherence and have unleashed a second wave of innovation in this space. Exploiting Optical Coherence is a repeat of a technique that was been thoroughly exercised in other domains. We used phase-amplitude keying to tune analogue baseband voice circuit modems to produce 56Kbps of signal while operating across a 3Khz bandwidth carrier. Similar approaches were used in the radio world where we now see 4G systems supporting data speeds of up to 200Mbps.

The approach relies on the use of phase-amplitude and polarisation keying to wring out a data capacity close to the theoretical Shannon limit. Optical systems of 100Gpbs per wavelength are now a commodity in the optical marketplace and 400G systems are coming on stream. It’s likely that we will see Terabit optical systems in the coming years using high density phase amplitude modulation coupled with custom-trained digital signal processing. As with other optical systems it’s also likely that we'll see the price per unit of bandwidth on these systems plummet as the production volumes increase. In today’s world communications capacity is an abundant resource, and that abundance gives us a fresh perspective on network architectures.

5G

What about radio systems? Is 5G an "emerging technology"?

It's my opinion that that 5G is not all that different from 4G. The real change was shifting from circuit tunnelling using PPP sessions to a native IP packet forwarding system, and that was the major change from 3G to 4G. 5G looks much the same as 4G, and the basic difference is the upward shift in radio frequencies for 5G. Initial 5G deployments use 3.8Ghz carriers, but the intention is to head into the millimetre wave band of 24Ghz to 84Ghz. This is a mixed blessing in that higher carrier frequencies can assign larger frequency blocks and therefore increase carrying capacity of the radio network, but at the same time the higher frequencies use shorter wavelengths and these millimetre-sized shorter wavelengths behave more like light than radio. At higher frequencies the radio signal is readily obstructed by buildings, walls, trees and other larger objects, and to compensate for this any service deployment requires a significantly higher population of base stations to achieve the same coverage. Beyond the hype it’s not clear if there is a sound sustainable economic model of millimetre wave band 5G services.

For those reasons I'm going to put 5G at the bottom of the list of important emerging technologies. Radio and mobile services will remain incredibly important services in the Internet, but 5G represents no radical change in the manner of use of these systems beyond the well-established 4G technology.

IPv6

It seems odd to consider IPv6 as an "emerging technology" in 2020. The first specification of IPv6, RFC1883, was published in 1995, which makes it a 25-year-old technology. But it does seem that after many years of indecision and even outright denial, the IPv4 exhaustion issues are finally driving deployment decisions and these days one quarter of the Internet’s user devices use IPv6. This number will inexorably rise.

It’s hard to say how long it will take for the other three quarters, but the conclusion looks pretty inevitable. If the definition of "emerging" is one of large-scale increases in adoption in the coming years, then IPv6 certainly appears to fit that characterisation, despite its already quite venerable age!

I just hope that we will work out a better answer to the ongoing issues with IPv6 Extension Headers, particularly in relation to packet fragmentation before we get to the point of having to rely on IPv6-only service environments.

BBR

Google's Bottleneck Bandwidth and Round-trip time TCP control algorithm (BBR) is a revolutionary control algorithm that is in my mind equal in importance to TCP itself. This transport algorithm redefines the relationship between end hosts, network buffers and speed and allows end systems to efficiently consume available network capacity at multi-gigabit speeds without being hampered by poorly designed active packet switching elements.

Loss-based congestion control algorithms have served us well in the past but these days, as we contemplate end-to-end speeds of hundreds of gigabits per second, such conservative loss-based system control algorithms are impractical. BBR implements an entirely new perspective on both flow control and speed management that attempts to stabilise the flow rate at the same rate as a fair share of available network capacity. This is a technology to watch.

QUIC

There has been a longstanding tension between applications and networks. In the end-to-end world of TCP the network's resources are shared across the set of active clients in a manner determined by the clients themselves. This has always been an anathema to network operators, who would prefer to actively manage their network's resources and provide deterministic service outcomes to customers. To achieve this its common to see various forms of policy-based rate policers in networks, where the 'signature' of the packet headers can indicate the application that is generating the traffic which, in turn, generate a policy response. Such measures require visibility on the inner contents of each IP packet, which is conventionally the case with TCP.

QUIC is a form of encapsulation that uses a visible outer wrapping of UDP packets and encrypts the inner TCP and content payload. Not only does this approach hide the TCP flow control parameters from the network and the network's policy engines, it lifts the control of the data flow algorithm away from the common host operating system platform and places it into the hands of each application. This gives greater control to the application, so that the application can adjust its behaviour independent of the platform upon which it is running.

In addition, it removes the requirement of a "one size that is equally uncomfortable for all" model of data flow control used in operating system platform-based TCP applications. With QUIC the application itself can tailor its flow control behaviours to optimise the behaviour of the application within the parameters of the current state of the network path.

It’s likely that this shift of control from the platform to the application will continue. Applications want greater agility, and greater levels of control over their own behaviours and services. By using a basic UDP substrate the host platform’s TCP implementation is bypassed and the application can then operate in a way that is under the complete control of the application.

Resolverless DNS

I was going to say "DNS over HTTPS" (DoH) but I'm not sure that DoH itself is a particularly novel technology, so I'm not sure it fits into this category of "emerging technologies". We've used HTTPS as a firewall-tunnelling and communication privacy-enhancing technology for almost as long as firewalls and privacy concerns have existed, and software tools that tunnel IP packets in HTTPS sessions are readily available and have been for at least a couple of decades. There is nothing novel there. Putting the DNS into HTTPs is just a minor change to the model of using HTTPS as a universal tunnelling substrate.

However, HTTPS itself offers some additional capabilities that plain old DNS over TLS, the secure channel part of HTTPS, cannot intrinsically offer. I'm referring to "server push" technologies in the web. For example, a web page might refer to a custom style page to determine the intended visual setting of the page. Rather than having the client perform another round of DNS resolution and connection establishment to get this style page, the server can simply push this resource to the client along with the page that uses it. From the perspective of HTTP, DNS requests and responses looks like any other data object transactions and pushing a DNS response without a triggering DNS query is, in HTTP terms, little different from, say, pushing a stylesheet.

However, in terms of the naming architecture of the Internet this a profound step of major proportions. What if the names were only accessible within the context of a particular web environment, and inaccessible using any other tool, including conventional DNS queries? The Internet can be defined as a coherent single namespace. We can communicate with each other by sending references to resources, i.e. names, and this makes sense only when the resources I refer to by using a particular name is the same resources that you will refer to when you use the same name. It does not matter what application is used and what might be the context of the query for that name, the DNS resolution result is the same. However, when content pushes resolved names to clients it is simple for content to create its own context and environment that is uniquely different to any other name context. There is no longer one coherent name space but many fragmented potentially overlapping name spaces and no clear way to disambiguate potentially conflicting uses of names.

The driver behind many emerging technologies is speed, convenience and tailing the environment to match each user. From this perspective resolverless DNS is pretty much inevitable. However, the downside is that the internet loses its common coherence and it's unclear whether this particular technology will have a positive impact on the Internet or a highly destructive one. I guess that we will see in the coming few years!

Quantum Networking

In 1936, long before we built the first of the modern day programable computers British mathematician devised a thought experiment of a universal computing machine, and more importantly he classified problems into "computable" problems where a solution was achievable in finite time, and "uncomputable" problems, where a machine will never halt. In some ways we knew even before the first physical computer that there existed a class of problems that were never going to be solved with a computer. Peter Shor performed a similar feat in 1994, devising an algorithm that performs prime factorization in finite time in a yet-to-be built quantum computer. The capabilities (and limitations) of this novel form of mechanical processing were being mapped out long before any such machine had been built. Quantum Computers are an emerging potentially disruptive technology in the computing world.

There is also a related emerging technology, Quantum Networking, where quantum bits (qubits) are passed between quantum networks. Like many others I have no particular insight as to whether quantum networking will be an esoteric diversion in the evolution of digital networks or whether it will become the conventional mainstream foundation for tomorrow's digital services. It's just too early to tell.

Architectural Evolution

Why do we still see constant technical evolution? Why aren't prepared to say "Well that's job done. Let's all head to the pub!" I suspect that the pressures to continue to alter the technical platforms of the Internet comes from the evolution of the architecture of the Internet itself.

One view of the purpose of the original model of the internet was to connect clients to a service. Now we could have each service run a dedicated access network and a client would need to use a specific network to access a specific service but after trying this in a small way the 1980's the general reaction was to recoil in horror! So we used the Internet as the universal connection network. As long as all services and servers were connected to this common network, then when a client connected, then they could access any service.

In the 1990's this was a revolutionary step, but as the number of users grew, they outpaced the growth capability of the server model, and the situation became unsustainable. Popular services were a bit like the digital equivalent of a black hole in the network. We needed a different solution and we came up with content distribution networks (CDNs). CDNs use a dedicated network service to maintain a set of equivalent points of service delivery all over the internet. Rather than using a single global network to access any connected service all the client needs is an access network that connects them to the local aggregate CDN access point. The more we use locally accessible services, the less we use the broader network.

What does this mean for technologies?

One implication is the weakening of the incentives to maintain a single consistent connected Internet. If the majority of digitally delivered services desired by users can be obtained through a purely local access framework then who is left to pay for the considerably higher costs of common global transit to access the small residual set of remote-access only services? Do local-only services need access to globally unique infrastructure elements.

NATs are an extreme example of a case in point that local-only services are quite functional with local-only addresses and the proliferation of local use names leads to a similar conclusion. It is difficult to conclude that the pressures for Internet fragmentation increase with the rise of content distribution networks. However, if one looks at fragmentation in the same way as entropy in the physical world, then it requires constant effort to resist fragmentation. Without the constant application of effort to maintain a global system of unique identifiers we appear to move towards networks that only exhibit local scope.

Another implication is the rise of specific service scoping in applications. An example of this can be seen in the first deployments of QUIC. QUIC was exclusively used by Google's Chrome browser when accessing Google web servers. The transport protocol, which was conventionally was placed into the operating system as a common service for applications was lifted up into the application. The old design considerations that supported the use of common set of operating system functions over the use of tailored application functionality no longer apply. With the deployment of more capable end systems and faster networks we are able to construct highly customised applications. Browsers already support many of the functions that we used to associate only with operating systems, and many applications appear to be following this lead. It's not just a case of wanting finer levels of control over the end user experience, although that is an important consideration, but also a case of each application shielding its behaviour and interactions with the user from other applications, from the host operating system platform and from the network.

If the money that drives the Internet is the money derived from knowledge of the end user's habits and desires, which certainly appears to be the case for Google, Amazon, Facebook and Netflix, and many others, then it would be folly for these applications to expose their knowledge to any third party. Instead of applications that rely on a rich set of services provided by the operating system and the network we are seeing the rise of the paranoid application as the new technology model. These paranoid applications not only minimize their points of external reliance, they attempt to minimise the visibility of their behaviours as well.

Change as a Way of Life

The pressure of these emerging technologies competing with the incumbent services and infrastructure in the Internet are perhaps the most encouraging sign that the Internet is still alive and is still quite some time away from a slide into obsolescence and irrelevance. We are still changing the basic transmission elements, changing the underlying transport protocols, changing the name and addressing infrastructure and change the models of service delivery.

And that's about the best signal we could have that the Internet is by no means a solved problem and it still poses many important technology challenges.

Where does this leave the New IP proposal?

In my view it's going nowhere useful. I think it heads to the same fate as a long list of predecessors as yet another rather useless effort to adorn the network with more useless knobs and levers in an increasing desperate attempt to add value to the network that no users are prepared to pay for.

The optical world and the efforts of the mobile sector are transforming communications into an abundant undistinguished commodity and such efforts to ration it out in various ways, or adding unnecessary adornments are totally misguided efforts. Applications are no longer being managed by the network. There is little left of any form of cooperation between the network and the application, as the failure of ECN attests. Applications are now hiding their control mechanisms from the network and making fewer and fewer assumptions about the characteristics of the network, as we see with QUIC and BBR.

So if all this is a Darwinian process of evolutionary change than it seems to me that the evolutionary attention currently lives in user space as applications on our devices. Networks are just there to carry packets.

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.

Author

Geoff Huston AM, B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net