September 2025
Geoff Huston

AUSNOG 2025

The Australian Network Operators' Group, AUSNOG, held its 19th meeting at the start of September. Rather than simply relate the content of the presentations I'd like to take a few presentations and place them into a broader context to show how such topics fit today's networked environment.

Network Operations and Data Centres

Perhaps unsurprisingly for a network operators' meeting, the topic of network management and operations has been a constant topic of presentations and discussions. From a suitably abstract perspective, packet switched networks are simply a collection of switch engines and transmission elements, and the ability to pull data from the processing elements that control the switching function has created endless possibilities.

In looking over the history of this area, we started with a simple model where the network's switching systems had consoles and network operators accessed the units performed both configuration and data extraction using console commands.

> The Internet tends to call these "routers", although the distinction between "routing" and switching" is pretty blurred these days. It may be the just a somewhat artificial distinction between the standards body that looks after the switching function (IEEE) and the body that looks after the routing function (IETF).
>
> The historical definition of the distinction was based on the ISO 7-Layer conceptual model of network functionality, where "switches" operated on the media layer (MAC addresses) and "routers" operated at the network layer (IP addresses.

It's highly likely that there are more than a few network management systems in use today that operate in the same manner, still using *expect* scripts to perform these functions of network monitoring and control using scripted interaction with the network element's consoles. The Simple Network Management protocol (SNMP) model was an early entrant into the network management suite of tools that broke away from the idiosyncrasies of command line languages. SNMP could be considered a remote memory access protocol, where each processing element could write data to labelled memory locations which could be read by a remote party via SNMP *fetches*. The SNMP protocol also allowed for remote writing to enable configuration operations. It's a very simple model, was widely adopted, and continues to enjoy widespread use, despite a very primitive security framework and limited repertoire of functionality. There have been numerous efforts to improve on SNMP. There is Netconf with YANG, which appears largely to be a shuffling of the

data modelling language from ASN.1 to XML, and not an awful lot else. There has also been a change in the tooling models, moving from individual tasks to orchestration of large pools of devices, with tools such as ANSIBLE, NAPALM and SALT. However, there are still a few invariants in the network monitoring toolbox, and these include the venerable first response go-to tools of *ping*, *traceroute* and a collection of convenient BGP Looking Glasses.

We've also moved to a different control paradigm, moving away from the dedicated network management console that queried a set of managed devices into a tighter level of integration by using container-based toolkits that place the monitoring and diagnostic tools into the routers directly. Rather than rely on vendor-provided diagnostic capabilities this programmable approach of loading custom images allows the network operator to code their own diagnostics and load these tools into the routers within the network. It's a step forward from the previous practice of deploying a small computer beside every router to collect data as to the view of the network from the perspective of the adjacent router and gain this perspective directly from the router itself.

The story of monitoring systems is like safety regulations, in that their incremental revisions are written using the blood of prior victims! In network management we tend to be reactive in what we develop and how we approach the task. But such an incremental approach to the task runs the risk of rampant complexity bloat as we simply add more and more incremental fixes to a common substrate. Collecting megabytes of network device status information each and every second is quite a realistic scenario today, However, the question is who has the time and skills to shovel through all this data to rephrase it as a relevant and timely indicator of current network anomalies and use the data to identify options to how to rectify them? The current fashionable answer is "This is surely a task for AI!"

One of the dominant factors in the network operations and management domain is the pressure of scale, and the view from Amazon AWS is a perspective from one of the larger of the hyperscalers. Scaling for Amazon is unrelenting, with an 80% annual increase in network capacity is a unique challenge. Such growth rates are slightly higher than the growth in silicon capability, so even if an operator would be willing to regularly swap out their silicon capability (at some considerable cost), they would still find it challenging to stay ahead of this magnitude of scaling pressure.

How should a service provider respond to such unrelating scaling pressures? For Amazon the answer is a combination of "stripped simplicity" and "automation!" They are claiming that some 96% of network "events" in their network are remediated or mitigated without any human operator intervention at all. That is a pretty intense level of automated response.

This is accompanied by something that is shared across all the hyperscalers, which is an intense focus on the absence of external dependencies. They take the mantra of "Own our own destiny" very seriously. This and the desire to automate everything are the key factors behind their capability to manage networks with over 1M devices.

The trend over many years in the routing and switching world in the equipment vendor world was to use ever larger units of equipment with ever larger port populations. To permit this equipment to be used in a broader range of operational contexts these large units were loaded with a continually expanding software base. The venerable time-division phone switches used a few million lines of code. Today's routers have a few hundred million lines of code! All this adds to the complexity of the operation of these devices, increases the risks of failure and increases the difficulty of fault isolation and rectification.

We are seeing a reaction to the trend to larger and more complex devices, and there is now a level of interest in simple single chip switches. Such single chip devices have fewer ports, less memory and perform fewer functions in total, but have the advantage that the overall switch architecture is simpler. With up to 64M of high-speed memory on the same ASIC chip as the switch fabric, such single chip switches can also provide improved performance at a lower price point. They have a lower port population per switch, so there are potentially many more such switches, but with their more modest size and power requirements they can be built into very dense switching frames with the additional factor of redundant switch capacity. A typical building block is a 12.8T 1xRU 32x400G ports switch, all with a single chip single control plane with 64M of high-speed local memory. 32 of these devices are assembled into a 100T rack, with a 3 tier Clos switching fabric, and 42 of these racks get us to the petabyte capacity!

The automation approach is to react to a known signal from a switch device by taking the device into an out-of-service condition and let the redundant framework take care of the associated traffic movement. What is perhaps unique here is the deliberate approach to design and construct an automated network from the start, distinct form the more conventional approach of retrofitting automation of network operations to an existing network. The standard automated response is to pull the unit out of service, mediate the device and roll it back into service, all in a fully automated manner.

It should also be noted that this approach of using a custom designed switch chip and its own firmware bypasses the typical router function bloat (and related complexity issues found in today's routers. The functions of these units can be tightly constrained and the conditions that generate a management alarm are also similarly constrained. This approach facilitates the automation of the environment at these truly massive scales.

Finally, in the topic of network management and operations for extremely large high-speed data centres, Nokia presented on AI data centre design and their take on Ultra Ethernet (UE). These special purpose facilities can be seen as an amalgam of a number of discrete network applications namely linking GPUs to memory (RDMA), storage access, in-band communication and out-of band control. The most demanding of these is the GPU memory access application. This is based on RoCEv2 (Remote Direct Memory Access Over Converged Ethernet v2), which is a networking protocol which enables high-throughput, low-latency data transfers over Ethernet by encapsulating Infiniband transport packets within UDP/IPv4 or UDP/IPv6 packets. RoCEv2's Layer 3 operation allows it to be routed across subnets, making it suitable for large-scale data centers. UE also requires a lossless Ethernet network, which achieved through Priority Flow Control (PFC) and Explicit Congestion Notification (ECN), to maintain its performance.

The GeoPolitics of Submarine Cables

Funding a submarine cable in the telco era was a process of managed controlled release. Submarine cables were proposed by consortia telco operators, who would effectively become shareholders of a dedicated corporate vehicle that would oversee and finance the construction and operation of the submarine cable. The route of the cable reflected the majority of interests that were represented in that cable company. The cable could be a single point-to-point cable system, a segmented system that breaks out capacity (branches) at various waypoints along the cable path, or a multi-drop cable system that supported multiple point-to-point services.

The telco model of cable operation was driven largely by the mismatch between supply and demand in the telephone world. The growth in demand for capacity was largely based on factors of growth in population, growth in various forms of bilateral trade and changes in relative affordability of such services. These growth models are not intrinsically highly elastic. The supply

model was based on the observation that the cost of the cable system had a substantial fixed cost component and a far lower variable cost based on cable capacity. The most cost-effective approach was to build the largest capacity cable system that current transmission technologies could support.

The risk here is that of entering cycles of boom and bust. When a new cable system is constructed, placing the entire capacity inventory of the cable into the market in a single step would swamp the market and cause a price slump. Equally, small increments in demand within a highly committed capacity market would be insufficient to provide a sustainable financial case for a new cable system, so new demands would be forced to compete against existing use models, creating scarcity rentals and price escalation.

The cable consortium model was intended to smooth out this mismatch between the supply and demand models. The cable company would construct a high capacity system, but hold onto this asset and release increments of capacity in response to demand, while attempting to preserve the unit price of purchased capacity. Individual telephone companies would purchase 15 to 30 year leases of cable capacity (termed an "Indefeasible Right of Use", or IRU), which would provide a defined capacity from the common cable, and also commit the IRU holder to commit to some proportion of the cable's operating cost. The cable company would initially be operated with a high level of debt, but as demand picked up over time the expectation was that this debt would be paid back through IRU purchases, and the company would then shift into generating profit for its owners once it had achieved the break-even level of IRU purchases.

In the telephone world these capacity purchases would normally be made on a "half circuit" model. Each IRU half circuit purchase would need to be matched with another IRU half circuit, such that any full circuit IRU was in effect jointly and equally owned by the two IRU holders. This framework of cable ownership and operation did not survive the onslaught of Internet-based capacity demand. Since the 1990's the Internet has been growing along the lines of a demand-pull model of unprecedented proportions. Over the past three decades the aggregate levels of demand for underlying services has scaled up by up to eleven orders of magnitude.

Today's Internet is larger by a factor of a hundred billion or more than the networks of the early 1990's. This rapid increase in demand exposed shortfalls in the existing supply arrangements, and the Internet's major infrastructure actors have addressed these issues by working around them. Around 2010 the brokered half circuit model of submarine cable provisioning was rapidly replaced with models of "fully owned" capacity, where a single entity purchased both half circuits in a single IRU transaction, and even to models of "fully owned cables" where the cable company itself was fully owned by a single entity. It was not coincidental that this coincided with the rise of the hyperscalers, and these days Alphabet, Meta, AWS and Microsoft account for 70% of global cable usage, and this growth has happened in the last decade. Their model is one that integrates transmission, data centers and service and content delivery. These days only one half of announced projects are completed, yet 100% of the hyperscaler projects have been commissioned to operational status. The submarine cable map looks denser every year.

In terms of geopolitics, national jurisdictions do not stop at the low tide mark of a territory, but extend out into the sea and the sea floor. The 1982 United Nations Convention on the Law of the Sea defined sovereign territorial waters as a 12-mile zone extending out from a country's coast. Every coastal country has exclusive rights to the oil, gas, and other resources in the seabed up to 200 nautical miles from shore or to the outer edge of the continental margin, whichever is the further, subject to an overall limit of 350 nautical miles (650 km) from the coast or 100 nautical miles (185 km) beyond the 2,500-metre isobath (a line connecting equal points of water depth). What does this mean for submarine cables? The point where the cable enters territorial waters

surfaces requires the permission of the country. It also implies that any cable construction and repair operations carried out in these sovereign areas also requires the permission of the country who has exclusive rights to the sea and seafloor.

Where the seabed lies on within contested areas, and the South Chinas Sea is current example of this situation, its more complex. Cables are expensive assets and are easily disrupted, and the preferred option is to avoid laying cable within contested areas. However, there is a further factor here, which concerns the simmering tensions between the United States and China. A Reuters news item in late 2020 reported that the United State had warned Pacific Island nations about the security threats posed by Huawei Marines "cut price bid" to construct the East Micronesia Cable System, echoing earlier concerns voiced by Nauru on the potential roles of Huawei Marine. The report mentioned the requirement placed on all Chinese firms to cooperate with China's intelligence and security services, and the potential implications this had regional security in this part of the Pacific. This was picked up by the United States, who voiced a similar warning on the potential security threats posed by the Huawei Marine bid to construct this cable. By mid 20201 the entire project had reached a stalemate and the project was ditched. The Hong Kong-Guam (HK-G) cable system was originally proposed in 2012 as a 3,700 kilometer undersea cable connecting Hong Kong and to Guam with 4 fiber pairs, with design capacity of 48 Tbps. The application to the FCC land the cable in Guam was withdrawn in November 2020. The Pacific Light Cable Network was a partnership between Google, Facebook and the Chinese Dr Peng Group, proposed to use 12,800 km of fiber and an estimated cable capacity of 120 Tbps, making it the highest-capacity trans-Pacific route at the time (2018), and the first direct Los Angeles to Hong Kong connection. The cable was completed in 2018, but a US Justice Department committee blocked approval of the Hong Kong connection on the US side, citing "Hong Kong's sweeping national security law and Beijing's destruction of the city's high degree of autonomy as a cause for concern, noting that the cable's Hong Kong landing station could expose US communications traffic to collection by the PRC." In early 2022 the US FCC approved a license for a modified cable system that activated the fibre pairs between Los Angeles and Balen in the Philippines and Toucheng in Taiwai. The other four fibre pairs are unlit under the terms of this FCC approval.

More recently the incidence of cable outages appears to have increased, with cable outages in the Red Sea, the Baltic Sea and the East China Sea that are potentially indicative of deliberate actions by some parties to disrupt communications. There is little that a cable operator can do to protect a submarine cable from such deliberate acts, particularly in shallower waters. In cases where it is feasible, multiple diverse cable routes are used to try and minimize the disruption caused by a single event.

Routing

Routing Registries have played a role in the inter-domain routing system since the early days of the Internet. These registries are used as a source of routing intention for a network, listing the prefixes that a network intends to originate and the intended policies that are associated with such originating announcements. A routing registry can also be used to list the ASes of adjacent networks and the policy of that adjacency (such as "customer", or "peer", for example). The registry allows other network operators to generate filters that can be applied to BGP peering sessions. These filters allow a network operator to only admit those routes that correspond to the network operator's route admission policy. They are also extremely useful in limiting the propagation of route leaks. Routing registries are a useful asset in the management of the inter-domain routing space.

However, these registries are far from perfect. Very very far! There are many route registries, and they contain differing information. Understanding which entry is the current "truth" where similar entries differ in detail across a number of routing registries is not readily resolved. Often the problem is simple neglect, where current information is initially added to the registry, but subsequent changes are just not recorded into the registry. But the problem is a little deeper than just inconsistent information. The issue is that this is an unauthenticated system where it's often possible in a registry for anyone to enter any information. How can a client be assured that the information that they are downloading from a registry and using to create routing filter lists in a production router is the "correct" information? The simple answer is that they simply can't provide any such assurance.

Such assurance is challenging. The entire RPKI system and the associated framework of route origination authorities could be regarded as nothing more than a replacement for the route registry's "route" object, that associates an address prefix with an originating AS number, and then generates a filter list to match the collection of such authenticated objects. Even then the level of assurance in this RPKI framework is by no means complete, in that the prefix holder has provided its authority via the use of the associated digital signature over the ROA object, by there is no similar assurance that the network references by the AS number is aware of this authority, let alone has provided any indication of its intent to generate such a route advertisement. The other aspects of route registries, such as AS adjacency and inter-AS routing policies have no counterpart in the RPKI, although work has been underway in the IETF on the ASPA object, an odd hybrid of a signed object that contains partial AS topology information and just one aspect of policy constraint. After 8 years of gestation in the IETF there are no expectations of when or even if the IETF will be done with this particular draft!

As well as the inability to authenticate these route registry objects there is also the issue of abuse of these fields. It's hard to determine if this abuse is deliberate or intentional, but the results are the same. In this case the route registry construct is the "AS Set", which is a macro operator that associates a list of ASes with a name. The members of an AS Set may be ASNs or other AS Sets, and this recursive definition allows the size of the set of prefixes that may be originated to quickly balloon if not carefully defined. As Doug Madory noted in his presentation to AUSNOG on the topic of "The Scourge of Excessive AS Sets", there is no inherent quality, integrity, and authenticity controls over content of AS Sets, and there are no limits to the number of AS members or AS-SETs, or the depth of recursion that can be used. There is also no agreed upon understanding of different use cases of AS-SETs. There are some pretty impressive examples such as the AS Set "AS MTU" which expands to 43,824 ASNs (of a total of some 80,000 ASes announced in the BGP default-free routing tables. This macro expands to some 1.3M lines of configuration to set up a filter list! Its by no means unique, and evidently some 2.192 AS Sets have expansions of 1,000 or more ASNs.

Any unsuspecting operator that uses route registry tools to generate configurations based on these AS Sets may be in for an unwelcome surprise. These excessively large sets may break route configuration tools or generate broken configurations.

What should we do about it? I think its late in the day to propose radical changes to the way routing registries and managed and the model of data acquisition, so proposals that assume that such change is feasible strike me as unrealistic. We could suggest as an alternative to performing radical changes to the mode of operation of route registries that that we work alone the lines of increasing the collection of RPKI-signed objects, using for example, signed prefix lists as an analogy of the route object and some form of AS propagation policy as an RPKI analog of the abstraction of the import and export clauses of AS objects. Given the pace of work in the IETF in the RPKI space,

which is somewhere between glacial and geological at present, then even if this were the preferred path there is no expectation that anything is going to happen here anytime soon!

Quantum Cryptography

I am happy to say that I am one of the overwhelming majority when I claim that I have absolutely no clear understanding of quantum physics and the magical properties of quantum entanglement and superimposition. I've yet to hear a clear explanation of qubits, state superposition and entanglement, and state collapse under measurement or observation. But the industry's hype machine is putting in the extra hours, and already we have quantum 2.0!

There is quantum networking and quantum computing, but of which are high-cost high profile flagship technology projects of the past decade or so. Large sums of money have been spent ofn various research projects and the early implementations of quantum computers are already out there. Among these projects, there's Google's Willow, Microsoft's Majorana and D-Wave's Advantage-2 which are moving the progress indicator forward. But while the underlying physics of quantum mechanics is really impressive, the performance of these early quantum computers is definitely not so. Finding the prime factors of 21 is now a quantum computable problem, but performing the same function on a 40-digit integer under a second is still some time away.

So why are we talking about quantum computers today? Some, if not many, of the answers lie in the secrets we want to keep and the cryptography that protects such secrets. But if the secret needs to be a secret for, say, 20 years, is that we are not assessing the difficulty of 'cracking' the cryptography using today's computers, but the feasibility of performing the same function 20 years from now is an interesting question. In past years this was a simple question, based on the scaling properties of improving the etching quality in large scale integrated circuits on silicon wafers. The shorter the wavelength of the light source and the better the optics, the finer the detail that could be etched on the silicon chip. The improvements in feature size on a chip followed a largely predictable model, called "Moores Law" after Intel's co-founder Gordon Moore, who first observed this regular doubling of the number of transistors on each silicon chip every year (subsequently modified to every two years). The greater the gate count, the more capable the processing capability of the chip, and a regular increase in the gate count implied a regular increase in the capability of the processor. After 20 years the continued application of this doubling predicts a compute capability some 512 times greater than the current capability. The second salient factor concerns the nature of digital cryptography. The challenge is not to try and construct insoluble scenarios, but to construct scenarios are impractical to compute using current computing technology. One such example is the task of computing the prime factors of a composite integer. So far we know of no mathematical shortcuts to this problem, and we are left with just the Sieve of Eratosthenes as the generic approach to this task. If you take two very large prime number s and multiply them together, then the task of calculating the original numbers is computationally infeasible.

Quantum computing holds the promise of improving the speed of some of these tasks. Shor's algorithm, developed in 1994, is an algorithm for a quantum computer to find the prima factors of large composite integers. To "break" the RSA-2048 cryptography in under 8 hours it is estimated that we would require a quantum computer of some 20M qbits. Today we are looking at around 1,000 qbits as the current state of the art in quantum computing.

So, if you want to keep a secret encoded using RSA-2048 for, say 20 years, then the biggest threat to this secret over this period is not the continued refinement of conventional processing capability with integrated circuits. The biggest threat to the continued integrity of RSA-2048 in the year 2045 is that by then we will have constructed large scale quantum computers than can perform this

prime number factorization task in hours rather than centuries or millennia. The big question is: When is this likely to occur?

What we are after is the time when cyphertext that is encrypted using RSA and ECC algorithms are susceptible to being broken by quantum computers that meet the necessary scale and reliability parameters. This computer of such scale is termed a Cryptographically Relevant Quantum Computer (CRQC). When that may happen is literally anyone's guess, but the more money that gets pumped into finding solutions to the engineering issues of quantum computers, the earlier that date will be. The year 2030 has been talked about, and it is not considered to be completely crazy date, even though it's well on the optimistic side (Figure 2).

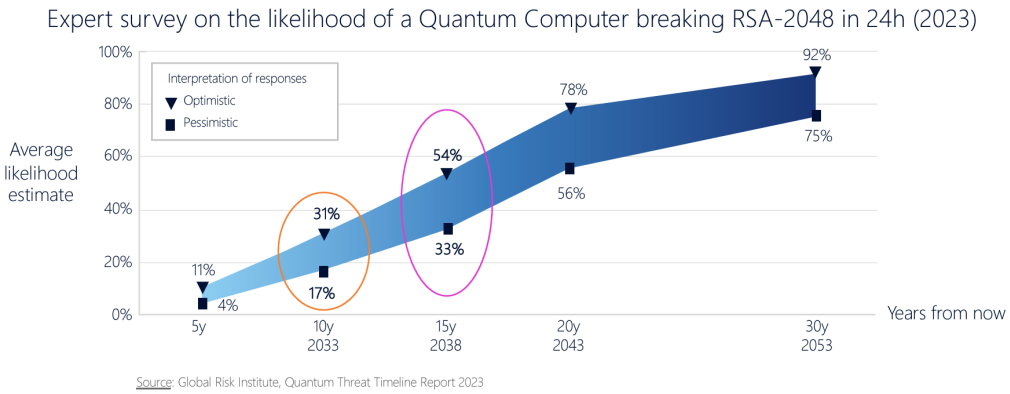

Figure 1 – Expectations of a CRQC timeline ("In-Flight Data Protection in the Quantum Age", Chris Jansen - Nokia,2Presentation to NANOG 92)

This date of 2030 is well within a two-decade horizon of keeping a secret, so it you want to ensure that what you are encrypting today remains a secret for the next twenty years, then it is prudent to assume that quantum computers will be used to try and break this secret sometime within that twenty-year period. Even though capable quantum computers are yet to be built, we still need consider the use of quantum-resistant cryptographic algorithms today in certain areas where long-held integrity of the encryption process is important. The present danger lies in an attacker performing data capture now, in anticipation of being able to post-process it at a later date with a CRQC. There is even an acronym for this, Harvest Now, Decrypt Later (HNDL). The response is to use a new generation of cryptographic algorithms that are not susceptible to being broken by a CRC engine. Easy to say but far harder to actually achieve. In the US the National Institute of Standards and technology has been running a program on Post Quantum Cryptography for some years. The challenge is that crypto algorithms are not provably secure, but can be considered to be secure until they are shown to be broken! The approach has been to publish candidate PQ algorithms and invite others to test them for potential vulnerability. The current state of this program is summarised on Wikipedia at https://en.wikipedia.org/wiki/NIST_Post-Quantum_Cryptography_Standardization

What applications require the integrity of long-held secrets? I have seen a number of presentations in recent months on the topic of using Post Quantum Cryptographic (PQC) algorithms for DNSSEC, the security framework for the DNS. It's challenging to conceive of a scenario where breaking the integrity of a DNSSEC private key would expose a long-held secret after 20 years. So no, there is no immediate need for PQC in the DNS. What about TLS? TLS is used for both authentication and session encryption. Its challenging to conceive an authentication task that requires 20-year secrecy. However, the case for PQC in session encryption is far easier to make. In this case there may well be communications that require secrecy for an extended period, and in

this case, there is a need to provision TLS with PQC options. On the whole, this is an easier outcome than contemplating PQC in the DNS. The larger key sizes in PQC present serious issues to UDP-based transport as fragmentation is an ever-present reliability factor, while larger key sizes can be more readily accommodated in a streaming transport such as TCP or QUIC.

And should a network operator use PQC for network layer encryption? I can see the case in certain specialised scenarios, particularly in cases where radio links are used, where the need for enduring secrecy of all communications carried over a network is vital, but for the case of the public Internet I do not see a convincing rationale.

Resources

The video records and slide packs from AUSNOG 2025 will be available online soon. Look for them at https://www.ausnog.net/events/ausnog-2025/program

Disclaimer

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

Author

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

www.potaroo.net