February 2015
Geoff Huston

Notes from NANOG 63

The following is a selected summary of the recent NANOG 63 meeting, held in early February, with some personal views and opinions thrown in.

Operators and the IETF

One view of the IETF's positioning is that as a technology standardisation venue then the immediate circle of engagement in IETF activities is the producers of equipment and applications, and the common objective is interoperability.

What is the role of a technology standards body? Should it try and be all things to all people? Or should it try and maintain focus and resist pressures for change? On the one hand is the risk of over-popularity leading to a lack of focus and loosing traction with the traditional objectives of the standards body. On the other hand lies the risk of ossification and creeping irrelevance by closing the group to all forms of change and adaptation.

The Internet Society appears to have taken on an agenda of promoting the IETF to a wider audience within this industry, and Chris Grundemann of ISOC reported to the network operators and the recent NANOG meeting in San Antonio at the start of February on this effort. His presentation reported on a 2014 survey on IETF awareness and most survey respondents have heard of the IETF but they do not think that the work of the IETF is directly relevant to their work and would not place participation directly in the IETF as a high priority for their time and efforts. Of course some would argue that this is as it should be - the task of technology standardisation is a specialist task, and while many folk consume products based on its standards that does not mean that such consumers would helpfully participate in the process of forming such a standard.

The IETF has already dipped its toe into selective standardisation and the collective refusal to standardise NATs for many years led to a chaotic outcome where each vendor of a NAT product was forced to be creative as to how to implement the subtler parts of this technology. While the IETF relented and eventually moved into this area with standard specifications, it was very late and the damage was done. NATs were a random element that applications have to cope with as best they can. These days the IETF is far less selective as to what it standardises, and as long as the task is related to IP or a select subset of application behaves it appears that the only admission factor is one of sufficient drive to propel the proposal through what is these days a rather protracted review process. It does not seem to matter as much what the proposal is. (I'm still scratching my head over a current proposals in the IETF to standardize MPLS over UDP and the flooding of link state information using that well known distance vecor protocol BGP! Its challenging to place such concepts within a more conservative view of the Internet architecture.)

Would more input, and particularly more input from network admins and operators help the IETF's efforts? Would the process benefit from a broader set of inputs during the review stages of the development of a standard? Or do the community of network operators already play a role in the most critical part of any standards process - the acceptance of the standard by the market for products and

services that use these standards? I'm not sure I appreciate the nature of the problem that ISOC is trying to solve here by trying to pull these folk into the IETF's technology standardisation process.

The FCC and the Open Internet Ruling

The Proposed FCC order on the regulation framework for ISPs in the US will be voted on by the FCC's Commissioners on Feb 26. The proposal is rumoured to be one that pulls in ISPs into the common carrier provisions of the US Telecommunications Act, with just, reasonable and non-discriminatory access, regulate privacy and disability and avoid other measures

There is the view that a legislated solution to the regulation of ISPS would in theory avoid subsequent judicial overturning, which lends some weight to the common carrier measures. But Professor Christopher Yoo of the University of Pennsylvania argued at NANOG 63 that this may not be an optimal outcome. He argues that there has always been some suspicion of the common carriage provision. There is a ricj history of the application of this principle in the provision of warehouses, ferries, barges, inns, taverns, etc in common law, but he maintains that the common carriage provisions have been poorly understood. Back in the 1930's the US Supreme Court used a public interest provision in testing the applicability of common carrier provisions when rejecting them. and maintains that there is little further to learn from the historic references to common carriage than is relevant to today's situation. However, he notes that the US uses common carriage measures in utilities including gas, electricity and water, yet do not include such measures in other forms of carriage and transit.

The traditional focus of common carriage measures is natural monopolies. The objective is to regulate rates and require structural separation. Professor Yoo argues that the public interest fix is itself far less than ideal and the argument is unregulated vs regulated in terms of acceptable compromise The direct costs of such regulatory intervention within this industry would be high: regulatory filings, accounting rules, technological disintegration, etc and they would be challenging to apply when there is no consistency of product, no consistency of technology, and when the interface is complex. So when the service is not precisely uniform and substitutable, then it declaration as a common carriage service would be challenging. The risks that the industry would face were the access service regulated include the risk of imposing barriers to innovation, adding impediments to further private investment, the imposition of inertia and unresponsiveness, biasing investment to inefficient solutions in the market.

The basic point that I took from this presentation is that any approach to regulation of the internet access service has its problems. It's not a black and white case in any respect. The issue is one of a case of trade offs. Allowing local access monopolies to continue without any form of regulatory constraint, and leave the threat of competition as being the only form of constraint for local access providers seems to construct a weak case for avoiding common carrier regulatory measures. On the other hand the lack of regulatory imposition on the products and services provided by the Internet to end users should encourage innovation, and through that to encourage competition in these markets.

The Status of the IANA Transition

ARIN's John Curran referenced the 14 March 2014 announcement that the USG planed to transition oversight of the IANA functions contract to the global multistakeholder community. He references the NTIA's conditions, namely supporting the multistakeholder model, security and stability and robustness. It explicitly states that it would not accept a government or multi-government proposal.

The IANA Stewardship Coordination Group (ICG) is chartered to coordinate the development of a proposal from the various communities. The initial community drafts were due Jan 15, and the next step is for the ICG to synthesise these proposals into a single proposal in response to the NTIA's requirements.

The end point is a proposal by July 2015 with demonstrated community support based on the supposition that this proposal provides for appropriate oversight and accountability for the IANA function.

Of course post this process, and assuming the US Congress would not disrupt the execution of this plan, then the issues about how to integrate governmental and intergovernmental mechanisms into the broader framework of the global communications endeavour will remain as issues that will not disappear any time soon.

There are now some rumours over a further extension to the IANA Functions contract, and comments about the level of "undue haste" on the part of the NTIA to call for a community process across a diverse global multi-stakeholder community certainly appear to have some substance.

There is little choice but to work within the parameters of the timetable as proposed by the NTIA, but on the other hand we've now experienced more than a decade of term-by-term extensions for the IANA Functions contract, and if the past is any indicator of the future then that would tend to water down the force of any assertion that "this is the last such contract, truly!"

The RING

Job Sniders has been presenting on this topic for some years now. The basic concept is a distributed set of slave systems that simultaneously execute a network diagnostic (e.g. ping or traceroute) allowing a user to "see" how they appear to the network from a diverse set of remote vantage points.

He has now added the function of a full mesh ping on a 30 second timer in both 4 and 6. This will generate an alarm to a node's operator if the node has been unreachable for 3 minutes - which is a relatively fast indication of local outage.

BGP Route Hijacks

This presentation looked at a number of specific examples of route hijacking. The examples he selected included:

Network hijacking to support the creation of bitcoin farms and bitcoin mining via hijacked pool of servers, which, in turn, may use a hijacked pool of routes. The scope of a Canadian hijack was limited to a single IX and its peers at Torix. 51 prefixes, 19 ASNs affected by the hijack.

Network hijacking in Turkey in March 2013. The Turkish authorities first tried to impose a set of DNS blocking filters on ISPs. This encouraged users to redirect their DNS queries to the various open resolvers. The authorities then tried to null route IP addresses of the more popular open resolver services, but in so doing they caused national breakage for a large number of users. Then they tried local spoofing on these addresses. The false routes intended to block the access to open resolvers did not mimics the originating AS, nor the original prefix sizes, making the effort highly visible.

Spammers. The problem noted here is that the RADB has no admission policy, so spammers were not only hijacking the prefix, but using RADB to make a bogus route entry! They hijacked an idle AS and then moved on to the DB.

Syrian outage - advertised routes blocked. mis-origination of 1500 prefixes, including the Youtube prefixes via Telecom Italia (hijacked 208.117.232.0/24 and announced this to TI)

Route Leaks (customer re-advertisement from transit to transit) (https://blog.cloudflare.com/route-leak-incident-on-october-2-2014/)

The presentation was a casebook of example situations of route hijacks, but Andree Toonk did not indulge in any speculation about possible cures!

Automating Network Configuration

Some aspects of technology appear to have moved rapidly, while others seem stuck somewhere in the paleolithic era. Network Management still appears to be in a very primitive state, where every active component that is deployed by a service provider is a distinct managed entity, and the configuration management function is performed by a command line interface and, if you ware feeing particularly brave, or foolhardy, via SNMP write.

As we shift into every larger operations with the number of managed units moving into a population of thousands or even tens of thousands of units then the management tools need to improve dramatically.

DYN is a case in point, operating across 20 data centres across the Internet and automation of their equipment configuration management is a major objective. NETCONF (RFC6241) is their chosen direction, using XML encoding and SSH-based access mechanisms (which is a good match to a Juniper environment). On this they are looking at Ansible (open source IT automation tool, which uses an SSH push model without a local agent. Junos has a defined NETCONF module for Ansible) and Jenkins (open source CI/CD tool that can automate cron management and event sequencing). The presentation then looked at DYN's approach to integrating these tools into their operational environment, within a framework they have built ("Kipper").

The problem with a CLI tool is that vendors do not implement uniform CLIs, so not only is the automation tasks a set of expect scripts that perform screen scraping to link to the unit's command logic, a multi-vendor environment requires a suite of vendor-specific CLI libraries to translate generic concepts into specific command read and write sequences. SNMP has been around for as long as CLIs, but few operators are prepared to use the SNMP WRITE functions, so its more of a monitoring and reporting tool. Shifting the data model from ASN.1 to XML or JSON, using a REST'ful API and connecting to the web, and using NETCONF with the YANG data model have all been tried. Today NETCONF and YANG enjoy a certain level of popularity, but it can be complex to set up, and there are a variety of open source management agents (such as Ansible) that are trying to synthesis a simple, scalable and effective operational management environment.

It was noted that the days of the lone netops Perl hacker may be numbered, but at the same time much of the Internet's infrastructure is still managed using their code, and the progress towards various managed approaches such as Schprokits, Cumulus and their ilk is slow.

One interesting area of speculation is the trend towards ever simpler network elements whose response can be configured, and placing more service delivery concepts into the overlay, and pushing functionality into the active elements via configuration comments. This allows a certain level of centrally managed services. SDN and Openflow for network element management if you want!

BCOP Publication Options

The RFC document series has a certain level of "cachet" in the Internet, but the IETF process of generating an RFC is at best protracted and messy. While the IETF in the past has tried to broaden its criteria in RFCs to include operational practices (Initially the "Informational" document sub-series, then subsequently the Best Current Practice" sub-series) there has been a continuing commentary that the process is still very IETF-centric, and the BCOP folk (which was initially a NANOG initiative, but other NOG groups appear to have taken this up, thanks in no small part to ISOC folk spreading the word) want a path to publish their BCOP documents as RFCs without necessarily driving the draft through an IETF review process.

This presentation summarized the current status of the discussion with the RFC Editor on this topic. The current options are to use the Independent Stream submission process or to charter a BCOP WG

in the Ops Area of the IETF that would not meet at IETF meetings, but would submit drafts into the RFC publication process in the same manner as all other WGs. Both options have their issues related to the current provisions in RFCs on derivative works and the form of document review.

It seems to me that the subsequent discussion at the NANOG meeting was not all that successful in gathering input from network operators, as it was more of a proxy discussion by members of the IESG who were in the room that left little room for any others to comment.

DNS Track at NANOG

New gTLDS

This is the 3rd round of top level domain expansion in the DNS, following the initial expansion of 7 new TLDs in 2000, and then a further 8 over the period 2003 - 2009. This round, encompasses more than a thousand new gTLD over some 15 months. Of these 500 new gTLDs have been delegated to far. All are signed with DNSSEC (minimally for the TLD zone itself), with a common Whois interface and with services offered over IPv6 as well as IPv4. Most of the names in the queue are awaiting contract completion, and these names will continue to be released through the next couple of years.

Operational issues: some of these new domain names may be rejected by software (e.g. with software using a backdated version of the Public Suffix List and its variants used in certain apps). Certain email junk filters on domain part may perform false positives because of false assumption of invalid domain name.

DNSVIZ CLI

DNSVIZ is a popular tool to analyse DNSSEC-signed domain names. Casey Deccio of Versisign presented on a CLI version of the tool, that generates a JSON version of the responses gathered in resolving the name and two other tools that interpret the JSON data. They are: dnsget - query / response to auth server(s), dnsviz – vixualiser and dnsgrok - interprets the JSON. The software is on https://github.com/dnsviz/dnsviz

A Standard for DNS over TCP

While there is a common belief that the RFCs are unequivocal about DNS servers supporting queries and responses over TCP (as in a normative "MUST"), the language used in the RFCs falls short of this. John Kristoff is advocating submitted an internet draft that proposes making support for TCP by DNS resolvers a MUST.

BIND update

There are new knobs in BIND that allow for thresholds in fetches per server and fetches per zone to make the local server authoritative about the domains that have unresponsive authoritative servers once the query rate exceeds the configured threshold, in order the scale back the query rate being passed to the authoritative server.

DNS Attack Traffic Profile

Attack profiles using the DNS increased mid 2014 and there are some billions of queries using random strings being queried as part of DNS DDOS attacks on authoritative name servers.

The attack profile uses a high volume of queries each with a unique name, and the task of the recursive resolver is to detect that they are forwarding an attack and throttle back their query rate to the auth server. It appears that the BIND query throttle controls may be useful in this context.

The DDOS Sessions

In some ways the details are less material than the larger picture. The Internet always had the potential issue that the aggregate sum of I/O capacity of the edge was massively larger than the interior, and the sum of multiple edge outputs was always greater than the input rate of any single edge.

Sometimes this is useful. With TCP the limit of a flow throughput is, in theory, either the limit of the edge or the limit of the network in the middle. By ensuring that the edge used highly capable I/O capacity then the performance of the edge device was such that it was in a position to make use of whatever network capacity is available.

The DDOS picture exploits this imbalance and drives multiple edges in a way that either saturates a network component or saturates a victim edge device.

The attacker has a variety of exploits here, and addressing one weakness simply allows others to be exploited. The current vogue is the massive bandwidth attack using simple query / response UDP protocols where the repose is larger than the query, with a spoofd source address. Send enough queries to enough servers at a fast enough rate and the victim is overwhelmed with useless traffic. The DNS UDP query protocol has been used in this manner, as has the NTP time protocol. You would've thought that folk would not expose their SNMP ports and at the very least use decent protection, but evidently SNMP is exploitable, as is chargen and similar. The mantra we repeat to ourselves is that we could stop all this form of attack by isolating the query packets, and the signature of these packets is a false source address. So if we all performed egress filtering using BCP 38 and prevented such packets from leaving every network then the network would again be pristine. Yes? Not really.

TCP has its own issues. SYN flooding is now a quite venerable attack vector, but this form of attack can still be effective in blocking out legitimate connection attempts if the incoming packet rate is sufficiently high. SYN flooding is possible using source address spoofing as well. A variant of this is for the attack system to complete the connection handshake and hold the connection open, consuming the resources of the server. For this form of attack its necessary to take over a large set of compromised end hosts and orchestrate them all to perform the connection. Unfortunately these zombie armies have been a persistent "feature" of the Internet for many years.

So the attacks get larger and larger. What happens then?

If you can't stop these attacks then you have to absorb them.

Which means deploying massive capacity in your network and across your servers in a way that can absorb the attack traffic and still remain responsive to "genuine" traffic. With significant levels of investment in infrastructure capacity this is a viable approach. But it has the effect of increasing the ante. If you operate a server whose availability is critical in some sense, then you can no longer host it on your own local edge infrastructure. And picking the cheapest hosting solution in some cloud or other is also a high risk path. If you want resiliency then you have little choice but to use a hosting provider who has made major investments in capacity and skills, and has the capability to provide the service in the face of such attacks. These attacks are creating differentials in the online neighbourhoods. Those who can afford premium services can purchase effective protection from such virulent attacks, while those who cannot afford to use such highly resilient service platforms operate in a far more exposed mode without any real form of effective protection against such attacks.

But of course this differentiation in hosting is also apparent in the ISP industry itself. Smaller local providers are being squeezed out through the same means. In order to survive such attacks they are being forced to purchase "scrubbing" services from far larger providers, who are in a position to absorb the incoming traffic and pass on the so-called "genuine" traffic.

The result is of course further pressure to concentrate resources within the industry - the larger providers with capacity and skills can respond to this onslaught of attack traffic, while the the smaller providers simply cannot. They are marginalised and ultimately will get squeezed out if there is no effective way to ameliorate such attacks without resorting to a "big iron" solution.

The picture is not entirely bleak, or at least not yet. There have been some attempts to provide an attack feedback loop, where a victim can pass outward a profile of attack traffic and neighboring networks can filter transit traffic based on this profile and pass the filter onward. One of the more promising approaches is to coopt BGP, and instead of flooding reachability information, propagate a filter profile of the attack traffic. (BGP Flow Spec RFC 5575)

SDN

No operators meeting would be complete without a presentation or two on currently fashionable technology, and of course what's fashionable today is applying the concepts of centralised orchestration of network elements, using the concepts of SDN. One such application is in supporting lambda path networks, where the elements of the network are not packet switches but wavelength switches.

At this stage there is a lot of effort in exploring the possibilities in such forms of centralised control over network elements, but its unclear to what extent such an approach will offer superior properties in network design and management in production networks. There has been much interest in this approach in the context of data centres and support of virtualised networks. The application of the same approach to wide area networking is still yet to prove its viability.

Disclaimer

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

Author

Geoff Huston B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region. He has been closely involved with the development of the Internet for many years, particularly within Australia, where he was responsible for the initial build of the Internet within the Australian academic and research sector. He is author of a number of Internet-related books, and was a member of the Internet Architecture Board from 1999 until 2005, and served on the Board of Trustees of the Internet Society from 1992 until 2001.

www.potaroo.net