The ISP Column
A monthly column on things Internet

November 2012

Geoff Huston

NANOG 56

NANOG held its 56th meeting in Dallas on October 21 through 24. I found the two and a half day program to be once more quite diverse and interesting. The following are my impressions of the presentations of this meeting.

NANOG is of course at http://www.nanog.org, and the NANOG 56 program material, including presentations and video streams, can be found at http://www.nanog.org/meetings/nanog56/.

At the start of this meeting it was reported that the meeting had attracted some 580 attendees. For NANOG this is a large number of attendees, and it clearly underlines the prominent position of NANOG in the network operations community in North America, and by virtue of the international attendance at the meetings from 20 countries other than the United States and Canada, across the wider Internet. The profile of attendees includes not just network operators but also includes attendees from research, education, government, content, operator, vendor and carriage sectors. The organization appears to be growing but at the same time the culture and relevance of NANOG to the Internet's technical domain is being maintained by this community.

DHS and Cybersecurity Programs

The opening presentation was from Doug Maughan of the US Homeland Security Advanced Research Projects Agency, talking on the major current activities in cybersecurity. One possible perspective is one of "more threats, less resources", but at the same time Doug noted that there is a strong organisational focus within the broader DHS environment on cybersecurity, noting in particular that "safegarding and securing cyberspace" as one of the five major DHS missions.

In particular, with regard to cybersecurity, the DHS has focussed on the deployment in secure DNS, secure inter-domain routing and IPv6.

Their operational support model for technology development is more inclusive than a conventional R&D funding model. It includes support of pre-R&D workshops and solicitations, funding of the R&D itself and then funding post R&D experiments and technology transfer. This third activity, the post R&D phase, is intended to mitigate the post R&D "valley of death" where startups fail despite good and innovative technology.

DNSSEC has been an active topic in DHS since 2004. DHS has been requiring the signing of zones in .gov, and has evidently achieved a 70% signed zone outcome. DHS is also requiring the use of DNSSEC validation in the resolvers used by these federal government agencies. DHS is working on transitioning some of its DNSSEC promotional activities to ISOC's Deploy360 activities. DHS is currently evaluating DANE, and other application developments that can take advantage of DNSSEC.

In the area of securing inter-domain routing DHS has been active in supporting the earlier sBGP work and supporting contributions to the SIDR Working Group of the IETF. Specific projects that have been supported by the DHS include validators, RPKI use tools, CA services and operational tools, such as the BRITE test tool for RPKI. In related routing security activities the DHS has supported the BGPmon program, and supported the RouteViews facilities.

In the area of IPv6, the DHS has been supporting NIST leading the charge for public sector adoption in the US, and the Office of Management of Budget (OMB) and the Whitehouse are requiring federal agencies to deploy both DNSSEC and IPv6.

In other areas the DHS is supporting internet measurement and attack modeling, including the CAIDA ARK work, HOST (Homeland Open Security Technology) to improve some of the current issues with current open security tools, and PREDICT to improve researchers' access to network data.

Routing Topics

NANOG 56 had its fair share of presentations on routing, and inter-domain routing in particular.

Arien Vijn of AMS-IX reported on a proposal to use IPv6 next hop addresses for IPv4 NLRIs in BGP (RFC5549). One of the motivations here is to allow a network to route IPv4 without having to use IPv4-addressed routing infrastructure. Another benefit would be to use a single IGP routing infrastructure rather than operating distinct interior routing systems for each protocol family. AMS-IX has implemented this functionality in Quagga.

There was a report of a survey of inter-domain routing polices from Phillipa Gill of the University of Toronto. Unsurprisingly to some extent more routing policies appear to favour directing traffic to ports that generate revenue (customers) rather than expenditure (upstreams) or, in some fashion, neither (settlement-free peer). Of interest was the reported variance in the treatment of rebroadcasting of routes learned from customers in the surveyed networks.

Matthew Luckie of CAIDA reported on AS Ranking (http://as-rank.caida.org). The approach here is to generate notional "customer cones" of AS's and use the count of direct and indirect customers to derive a ranking across AS's. This is different from the Gao-Rexford upstream/downstream relationship algorithm, in so far as it attempts to rank AS's based on size without reference to any derived routing relationships.

Another routing research presentation was on the topic of joint network and content routing from a graduate student at the Georgia Institute of Technology. I must confess to a fair amount of skepticism of various styles of feedback-based routing. Robust feedback controlled routing systems in IP networks are tough, and while it is possible to display selected cases that show such systems in the best light, that does not necessarily mean that all of the issues in feedback-controlled systems where the load model is elastic have been adequately addressed.

Security is also a dominant theme in today's meetings, and Randy Bush presented on RPKI repository synchronization measurements. As with many measurement exercises it's unclear what the underlying issue might be. It's an open question for me whether its optimal in an architectural sense to match a highly distributed repository publication model with a time critical application with minimal tolerance for missing data that is bound to the operation of real time routing updates in the BGP protocol. I suspect that alternate approaches may reduce some of these stress points. More intriguing for me was a quick presentation by the same presenter on the TCP performance of BGP on a Cisco platform. There was a strong inference in the presentation that the internal scheduler structure of the Cisco kernel (reported in this presentation to be a run-to-completion, non-preempting scheduler) had a very strong negative impact on the performance of this protocol.

A presentation on the state of play with the effort to standardize diverse home networks by Chris Grundemann was for me a clear illustration of the underlying architectural reasons why broadcast LANs are amazingly effective for simple plug and play connection models. Within a broadcast LAN the inherent security domains are embedded into the attached device and its applications rather than into the network. The effort to introduce complex routing structures into the home to create network-based security realms, while still wishing to use a simple plug-an-play model, appears to be inherently contradictory, and this presentation only illustrated these contradictions, and failed to demonstrate that the effort in this IETF working group represented a feasible approach to routing-based frameworks into what essentially remains a plug-and-play domain. If this was intentionally meant to be a critique of the IETF's Home Networking Working Group it appeared to me to be an extremely harsh and scathing critique indeed.

DNS Topics

The NANOG DNS Threats track was very interesting. Joe Abley of ICANN described the structure of the cryptographic material at the root zone of the DNS. Conventionally a hash of the Key-Signing-Key (KSK) of a DNS zone is published as a DS resource record (RR) in the parent zone. Of course the root does not have a parent zone, so in this case the hash of the root zone's KSK is published at https://data.iana.org/root-anchors/root-anchors.xml. Joe's report noted that since July 2012 the volume of downloads of this file was recorded as some 1,500 to 6,000 downloads per month. However it was noted that in September 2012 the volume of downloads of this file was 353 million! It was noted that this escalation in volume commenced on September 19, and all of these additional retrievals have the user agent string of "CFNetwork/609 Darwn/13.0.0", which corresponds to iOS 6 activations and upgrades. This points to an entirely new world of DNSSEC where its not just a select group of operators of DNS recursive resolvers that perform DNSSEC validation, but a world where hundreds of millions of devices each perform DNSSEC validation of DNS queries directly. The implications here is that when (and it is evidently "when" and not "if") it's time to perform a key rollover of the root zone's KSK then the set of relying parties that will need to track this key rollover now includes a large world of devices that are operating in a totally un-administered mode. The presentation pointed to two internet draft documents: draft-jabley-validator-bootstrap and draft-jabley-dnssec-trust-anchor. I'd like to quote from his slidepack: "we live in hope".

The other theme in this track was the proliferation of the DNS reflection attack. This is commonly seen as a query for the ANY pseudo-RR, where the response is expected to be far larger than the query. Because this is a UDP transaction the DNS response will be sent to the source address contained in the original query, without any effort to validate that the source address was the actual originator of the query. This form of attack creates a DDOS attack, and had evidently been observed as an active attack vector for many months. A common form of response is to perform query rate limiting of the triggering queues. This form of rate limiting can be found at http://www.redbarn.org/dns/ratelimits. Other forms of mitigation were also considered, including the use of anycast nameservers, to limit the scope of the attack to a single anycast instance.

There was also a presentation on DNSSEC practices in the upper zones of the DNS by Ed Lewis of Neustar. He reports that of the 306 zones currently at the "top level" of the DNS some 27% are DNSSEC-signed, and some 70% have no DNSSEC. Perhaps unsurprisingly, he also reported that adoption at the top of the DNS tree is greater than in the lower portions of the tree. It seems that within the upper levels of the DNS there is a certain degree of "set and forget" going on with DNSSEC, with slightly less than half of the domains still using the older RSA-SHA1 algorithm to generate the cryptographic material. The issue of key lengths as related by Ed is, as he puts it, a good illustration of the uncritical reception that is often associated with RFC publication. RFC4641 suggested to use 1024 bits for the zone signing key and 2048 bits for the key signing key. However these are only suggestions and another important consideration is the size of the DNSSEC query responses because of the issues with fragmentation, truncation and failover to TCP in terms of performance and robustness. However irrespective of these considerations Ed observed that 90% of the signed zones simple use 1024 and 2048 but keys respectively. On the other hand key lifetimes are creeping upward and the RFC guidelines are no longer being followed so assiduously in this respect.

I presented a lightning talk on an exercise to count the number of visible DNSSEC-validating resolvers, and the proportion of users who appear to use exclusively DNSSEC-validating resolvers.

Operational Topics

Job Snijders presented on the NLNOG ring. This form of cooperative endeavour in sharing a set of resources among a collection of operators under a common provision of mutual trust is something that was more a feature of the Internet of the early 1990's than today's operational Internet. The question "does everyone else see what I am seeing?" and "how does everyone else see me?" are still very important diagnostic questions when debugging operational incidents and this form of approach that allows an operator to gather a collection of remote views remains truly useful!

There was a panel on Network-centric performance management which I must admit did not do that much for me. I was left with the distinct impression that the story about the state of network management today is that we are still doing little more that banging the rocks together, but these days we are calling it "rocket science!" The panel on traffic accounting practices struck me in a similar way.

The presentation on Ethernet speeds showed just how tightly supply and demand are tracking these days. Current high end Ethernet speeds in production equipment run at 40Gbps, and the question in IEEE circles is what speed to concentrate on next. The presentation indicated that 400Gbps was the next objective, using a base of 25Mbps signaling and combining 16 parallel lanes into a 400Gbps channel. Options include 10 lanes of 40Gps or 8 lanes of 56Gps. Admittedly this is a conservative target, but one that attempts to create a product within relatively aggressive time and const targets. However, there is some skepticism about whether this incremental approach is efficient, in so far as the industry demand for 1Tbps speeds is now well accepted. The challenges for Tbps speeds are well appreciated, and the estimates in this presentation are that it will be more likely 2017 or later before these speeds are standardized and available in equipment.

The last presentation I'd like to highlight here was by Lee Howard on the total cost of ownership of CGNs. It was an interesting exercise in trying to quantify the capex, opex and externalities associated with addressing IPv4 address scarcity through the deployment of NAT technology within the carrier's infrastructure. His starting point is an estimate of a capex of $90K per 10,000 users, depreciated over 5 years, and an opex estimate of 10K p.a.. The process then attempts to estimate the level of "brokenness" using as a starting point the internet draft draft-donley-nat444-impacts, and from this estimates that one third of these 10,000 users will be impacted in some way, which will, in turn, generate 914 support calls and 914 users who terminate the service. His model writes off the cost of these calls at $20 per call over the first year, and none thereafter. The terminating users represet lost revenue, which he estimates at $400 per user per year, based on industry filings in the US. This lost revenue represents a business loss of some $365,000 p.a., which is by far the largest component of the costs here. This model produces a $2M total cost over 5 years, or $40 per user per year. According to the annual reports of the ISPs the ARPA per user is, on average, $400 per year, and the margin is some $140 per user per year. Over 5 years each customer will generate $700 of profit, and the thought experiment goes that if each IPv4 address currently owned by the business is worth $10, then its in the business' interests to retain the address and retain the customer. But if the price per address reaches upwards of $40 per address then its an interesting decision whether its in the business' best interests to sell the address and immediately realise the asset, and switch the customer to a CGN service. This presentation postulates that at a market price of $71 and higher there is a strong business case to place the customer behind a CGN and realize the value of the released IPv4 address.

It's a thought provoking analysis, but I wonder if it includes the entire business case relating to CGNs. CGNs generate a rich data seam. Every web site a customer visits, every advertisement they click on, every email, every network transaction in effect, is exposed as a binding in the CGN, so the CGN log represents a high volume, but potentially valuable, asset. It is often observed that the various "rewards cards" used by the retail industry make sense only in the context that the use of the reward card allows retailers to track the user. A typical rewards transaction is of the order of 1 "point" per dollar of spend, which generally equates to the value of tracking at around 0.5 cents in the dollar. If the average expenditure per user in online purchases is some $8,000 per year, then this equates to a tracking value generated by the CGN log of some $40 per year. It would be interesting to plug this into Lee's model to observe the outcomes in terms of the cost and opportunities that CGNs bring the to the ISP's table. I also wonder if the CGN creates a distinct "two-tier" market for content distribution, where the rack space located in the interior of the CGN provides a clear and coherent view of each customer and the externally NATTed view becomes one prone to proxy and translation-induced distortions. I suspect there will be more in this area as the industry continued to grapple with the now inevitable transition from a simple edge model to a combined edge and infrastructure model of NAT deployment.

The next NANOG meeting will be at Orlando, Florida, February 2 through 4, 2013.

Disclaimer

The views expressed are the author's and not those of APNIC, unless APNIC is specifically identified as the author of the communication. APNIC will not be legally responsible in contract, tort or otherwise for any statement made in this publication.

About the Author

Geoff Huston B.Sc., M.Sc., 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 has been active in the Internet Engineering Task Force for many years.

www.potaroo.net