Geoff Huston
July 2016

# IETF 96

The IETF meetings are relatively packed events lasting over a week, and it's just not possible to attend every session. Inevitably each attendee follows their own interests and participates in sessions that are relevant and interesting to them. I do much the same when I attend IETF meetings, and from the various sessions I attended here are a few personal impressions that I took away from the meeting that I would like to share with you.

## IEPG

The IEPG meets the Sunday prior to the IETF meeting for approximately two hours. The topics are varied, but tend to be focused on operational matters of one sort of another.

Randy Bush spoke on "ROA Misconceptions". The validity of a ROA has nothing to do with a BGP announcement, is his assertion. What he is saying is that the valid ROA set may be able to mark a BGP announcement as "valid" or "invalid", while the set of invalid ROAS is ignored in this decision process. The second misconception is that the AS has implicitly allowed a ROA to mention them - this is not the case. The AS in a ROA is an assertion by the prefix holder, not the AS itself.

Davey Song and Shane Kerr spoke of the current status of the YETI project. This was originally mentioned the IEPG meeting in June 2015 attracting some negative reaction as being just another fake root setup. This is a status report, and a request for more DNS traffic. Their intended area of experimentation lies in the size and scope of the root server structure, the key roll, privacy issues, priming queries, IXFR/AXFR and DNSSEC. This also includes the issues of the balance between anycast and DDOS. The major concerns here are given as "external dependency and surveillance risk, and testing new technologies in the root system". Obviously there are few "at scale" recent studies of the root system, particularly so given the extreme negative reactions from the community to the concept of "alternate root" systems over the years. They are running is some 25 Yeti root name servers. They pull a copy of the root zone from F root and sign it with their own keys and redistribute it into the Yeti farm. They are saying that the "response" is 2,134 bytes, but no data on the success/failure rate of this size, nor even what query this response size corresponds to. They need traffic (queries). They are currently operating at < 100 queries per second. Shane presented on the "science" of the Yeti work. One experiment is to use a distinct ZSK to sign distinct distributions - i.e. multiple ZSKs. Another is a 2048 bit ZSK - this is evidently unsurprising. Davey Song commented that: "Frankly speaking we started an experiment with inadequate preparation." At this point the rhetoric and the reality of this work is still somewhat distinct.

Sara Dickinson spoke on DNS Privacy - Implementation and Deployment Status. Background on the use of the DNS as a window on user activity. The initial charter of the IETF in chartering the work was to limit the DPRIVE study to look at the stub to resolver situation, while the larger picture includes recursive/forwarders and auth name servers. RFC7858 for DNS-over-TLS-over-TCP using port 853. RFC7816 accompanies this (QNAME minimization). RFC7626 is the foundation document, as it rebuts the myth that the DNS is "open", so privacy is a tautology. The presentation relies on a standalone TLS and does not look at any particular concerns over the trust anchor structure for the

keys behind the TLS certificates. The presentation painted DNS over TLS over TCP in a very optimistic light, particularly in comparison to DNS over DTLS, the effort to support TLS in a UDP environment. There is some momentum over the emerging pool of DNS recursive resolvers listening on port 853. RIPE NCC is considering joining OARC and NLNET labs in offering an initial service. Francis Dupont mentioned the current ISC work: http://bit.ly/2akebha. There is also of course the Google effort with dns.google.com, that combines DNS over TLS using port 443 with JSON encoding of the DNS response.

Lief Johansen spoke on the current status of Cryptech.is. The motivation is well understood - the project's objective is to produce a readily accessible HSM that is independent on any vendor and built using an open process, yet is trustable. the objective is to diversify the participants, diversify the funding, and diversify the work. The project is now at the point of an alpha version of the HSM board. The talk presented some options on the generation of randomness, and looking at their choice of the noisy diode. The Board uses an ARM CPU and a FPGA for the crypto primitives and the noisy diode random number source. One of the design choices has been FPGA vs CPU. One of the considerations is that fast general purpose CPUs have many extraneous functions and interfaces that increase the attack surface of the processor in an HSM context. the FPGA is certainly a more limited device that allows for a more restricted attack surface. Currently its still not in production and still not fast, but expectations are that given funding this will improve in the coming months. As to the role of HSMs and their utility in a PKI environment, it's still unclear to me the extent to which security pantomime is at work, and the effort to create solid and reliably secure keystores is nowhere near the entirety of a decent picture about digital security, and I have to wonder if this is an instance of lopsided effort that in focusing attention on the HSM distracts from crypto robustness and key usage considerations.

Daniel Karrenberg spoke briefly about Atlas with an introduction to the Atlas measurement approach. The presentation showed "something happened" in a number of cases, but as to exactly what happened and why is not directly addressed in the report - largely an awareness presentation.

Sabrina Tanamal spoke of IANA Registry Updates. The presentation noted a subtle change to the AS number registry, a subtle change to the use of terminology of "global" in the IPv6 registry. POC details are also being updated.

Jordi Palet spoke of his recent IPv6 Deployment Survey. The survey polled ISPs about the address plans and technology for IPv6.

I spoke on recent work to report on IPv6 Performance. I'll report on this in a seperate article so I won't go into detail here

Dan York spoke briefly on DNSSEC Algorithm Agility. This is a followup of the slides from the DNSSEC Workship at ICANN 54 on the moves to get away from RSA to ECDSA. This presentation questions where and how there are points of resistance to cyrpto algorithm agility.

All the slides from the IEPG meeting are at http://www.iepg.org/2016-07-16-ietf96

## Transport Services Area

Discussion on IP Stack Evolution. It appears that it's hard to deliberately evolve the stack, but it is evolving anyway! The changes are more casual rather than deliberate. This is one effort to try and impose some order and direction on the evolution. Workshops, BOFs, reports have been the result so far. So far these efforts have all provided commentaries on stack evolution efforts, but they tend to be after the event rather than acting as an impetus or setting a particular direction. There is also an increasing level of frustration with this space. It's not that there are no more ideas about how to innovate with the stack, and no shortage of proposals to tune the transport protocols to react in productive ways to various networks. That's not the source of frustration. The frustration lies in the

level of middleware in today's Internet that attempts to manipulate the control parameters of transport sessions. This then imposes some constraints about how the ends can use TCP, as if they stray too far from a narrow set of conventional behaviours middleware will intervene and disrupt the session.

There was a report on the QUIC work. (QUIC is one solution to the stack stasis problem that works by removing the transport protocol from network visibility and using UDP as the visible end-to-end transport protocol.) The user-space transport in QUIC is then a shared state between the endpoints with no intervention by network middleware. It was reported that there are some interesting observed network anomalies, such as large scale (>100 packets in a single flow) packet reordering. They have seen small packets that appear to race ahead of large packets, but not necessarily the opposite. They are looking at time-based loss detection to be tolerant of reordering - i.e. time based not sequence number based. They also see rapid (<15 second) NAT rebinding, and note that port rebinding is more common than IP address rebinding. This fast rebinding is currently isolated to a small fraction of networks. In response they have adopted a HIP-like connection identifier, and route based on this connection identifier at the server side. Such connections are then more resilient to NAT rebinding, as the connection identifier is tolerant of changes in the IP address and UDP port fields. They have noticed that packets are switched to the wrong server. The see sudden blackholing of packets partway in a connection. They assert QUIC works 93% of the time, and in other cases its UDP rate limited, UDP is blocked, or QUIC loses the UDP/TCP race. They observed that QUIC rate limitation has decreased by 2/3.

MPTCP developments. Apple's Siri, Gigapath in Korea and several startups apparently use MPTCP. They claim its mostly smooth with a 94% success rate, although deployment issues were simplified to some extent due to small scale controlled environments. However, there are still many unanswered questions, such as the timing of opening of sub flows, and there is a need to understand how to control the API to switch the flows across heterogeneous paths. What are the possible impacts if MPTCP were to be used on a large scale? How would this edge-controlled multi-path system interface load balancing flows with network-based Equal Cost Multi-Path (ECMP) load balancing? In this case the ECMP flow hashing would mean that the client presented as multiple IP addressed end points would infer the use of multiple head end servers. At this stage it looks as if MPTCP is a solution that is looking for the 'right' problem.

## DNSOPS

Shutting down the DNSEXT Working Group did not mean that the work in extending DNS functionality stopped at the same time. It just meant that the agenda for the DNSOPS Working Group now also carries the DNSEXT agenda, so the workload in this working group is now somewhat intense.

Thre are a number of recent published RFCs: chain query is now RFC7901, DNS cookies is RFC7873, EDNS0 client subnet is RFC7871. There are also a number of documents in IESG evaluation, including DNSSEC roadblock avoidance, maintain DS, and the NXDomain cut. Also waiting is the isp-ip6rdns and no-response-issue documents. Near future work includes resolver pinning, refuse any, key tag, 2317 bis, and attrleaf. Also work on wireformat is in a call for working group adoption.

Paul Hoffman presented on a terminology update to RFC7719. It's a little scary that the DNS now has its own considerable lexicon of terms that need to be used with precision!

Warren Kumari gave an update NSEC aggressive use. If has been observed that conventional NSEC responses cover a range of possible queries, and a recursive resolver could cache the NSEC response, including the RRSIG values and simply replay the response for all queries that fall into the covered range for the lifetime of the NSEC TTL records. This would have a dramatic impact on the amount of query traffic presented to root servers were all recursive resolvers in a position to perform such caching of NSEC responses.

> The Cloudflare synthesised NSEC approach raises an interesting question about the value of NSEC3. The "problem" of NSEC was that it enabled comprehensive zone enumeration. For a zone manager that was in effect trying to sell unused names, access to the inventory of unused names was an anathema to them. So the answer was to hash the names and use the hash ordering as a new form of authenticated denial of existence, AND thus NSEC3 was born. It turns out that NSEC3 is pretty much a wasted of effort. Not only is the NSEC3 hashing function susceptible to basic efforts to decrypt the hashed values (http://bit.ly/2acmvuP), but it's possible to achieve the same intended outcome of preventing the enumeration of zone contents by a far simpler approach to generation of NSEC records, as Cloudflare has shown.

Sara Dickinson reported on progress with DPRIVE work, principally DNS over TLS over TCP implementation status. Knot is working on the TLS feature. Another option is to use DNSDIST on the server side, but this has a number of shortfalls. GetDNS has a daemon mode as a local stub resolver to forward over TLS.

Ray Bellis reported on the session signal draft (new work). The background here is that EDNS is per message and stateless. RFC7828 (tcp keepalive) fudges around this, and the realization was that what was missing was session signalling. This overlapped with dnssd-push. So: what is a "session"? They propose that its a bidirectional long lived state but include ordering (i.e. excludes UDP) It uses a new SESSION OpCode to signal session options. Each session option is explicitly acknowledged.

DNS with Multiple QTypes/Responses. The observation they are using is that DNS queries are often "bunched" with multiple queries, as there are logically linked. The inference is that a server may well be told the set of related queries and pre-provision the set of related answers. So the server can bundle a set of related answers from a trigger query. Uses the EXTRA EDNS0 option in the query and this loads the Additional section of the response with these multiple additional responses. Interestingly, this gratuitous bundle must be DNSSEC-signed.

A variant of this is Multiple Qtypes, where there is a single QNAME with multiple QTypes. this is not the first effort to put multiple queries in a single query msg, and no doubt will not be the last, On the other hand even in the case of A and quad A this could be a big win on the recursive resolver load.

The BULK option is interesting – it's like loading the auth server with regex patterns and match query to response - it reminds me of the NAPTR work in a strange way.

Special Use Names- the chairs asked for a problem statement - there is some convergence but people are getting tired about this topic.

## MAPRG

A number of research groups meet during the IETF week. I attended MAPRG, concerned with active and passive measurements. I was particular struck by a report from Philipp Richter on address use. He used source addresses collected by Akamai to generate a "map" of addresses, and then took successive snapshots of address use to compile a time-series map. What was evident was that there is a distinct "signature" of various forms of DHCP address pool management. They saw some 1.24B active IPv4 addresses, which is 42% of the routed space. This corresponds to 6.5M active /24's, which is 59% of the routed total. Of these addresses some 44% is unresponsive to ICMP probes. It was also observed

that over 52 weeks some 25% of the active address pool changed! It was a useful commentary to the address policy community's discussions on "address utilization".

Tommy Pauly of Apple reported on Dual Stack IPv6 preference, where the more recent versionsof iOS and macOS attempt to bias the address choice in favour of IPv6. They observed a connection ratio of 87% using IPv6 on Dual Stack hostnames.

## 6MAN

> Sometimes I have to wonder about the modes of behavior in IETF Working Groups. Assembling some 1,500 interested individuals into the same location three times a year is not an insignificant exercise, and the expectation is that this would be an opportunity to hash out some of the matters that otherwise take endless rounds of email exchanges on the Working Group mail lists. So it seems curious to me to bring all of these folk into the same room to hear a continual incantation from the WG Chairs top "take it to the list". That's broken in my opinion! I bring this up because that particular incantation was evident in the 6MAN Working Group meeting, concerned with the maintenance of the IPv6 standard specification.

Some topics appear to oscillate between two states without clear resolution. One of these is the treatment of the 64 bit interface identifier part of the IPv6 address. Originally this was specified as a "stable" value which was suggested that it could be automatically generated by the interface's 48 bit 802 MAC address when appropriate. It was subsequently pointed out that this leaks a lot of information when a host moves across networks, and the concept of a "Privacy Address" was introduced. These addresses are randomly generated by a host and change periodically. So now there are proposals for "Stable" Interface Addresses once more! It strikes me that this is a case of bit twiddling, and It's hard to take this seriously! Considering that an entire 64 bit field is used for interface identifiers it's a pity that we appear to have no clear idea how best to use them, and instead there is an endless parade of drafts proposing some new way of treating these bits. It's hard to take IPv6 seriously when you observe the elves muck around with the IPv6 address architecture in this manner!

The IPv6 as a full Internet Standard debate continues - at this stage the observation is that the decades long hacking by committee has left a warped and sometime mutually inconsistent legacy of tweaks. This effort to standardize exposes these internal contradictions without understanding how to resolve them. See Tim Chown's slides (http://bit.ly/2afJAjX) for a laundry list of the overt inconsistencies in the current IPv6 specification set. Some documents are not in the set to be advanced (RFC 4941 on Privacy Addresses, for example) so what is being proposed as a standard is a subset of the "complete" specification.

What is evident is that the specification for IPv6 is not clear in every respect, and well understood in every respect. Any "standard" specification will inevitably make some choices between viable alternatives and there is no real assurance that the choices made here and now are the same as choices that may be made in the future or in other venues with different relative ranking of trade-offs. So the first message that I took away from this session is that the IPv6 specification is not yet stable, complete and well understood.

## Other Notes

### KSK Roll

Yes, its happening, and Matt Larson of ICANN explained the current intentions of this work. More details can be found at https://www.icann.org/resources/pages/ksk-rollover.

### Plenary Session

Ross Callon gave what he called a "retirement rant" on the topic of too many standards and the resultant devaluation of IETF work. "The IETF needs to finds a way to avoid frivolous standards" is his plea. It seems like a valiant effort, but one that will have little impact. The current ethos of the IETF is better phrased by quoting Mao-Tsung: 百花齊放, 百家爭鳴. It is certainly true that the IETF does not perform any critical selection process any more, and whether its VPN tunneling approaches, virtual circuits, IPv6 transition mechanisms, or any other active area, the IETF now appears to allow many parallel non-interoperable approaches to emerge. The OSI protocol suite was roundly criticized at the time for allowing mutually incompatible Connection-oriented and Connectionless transport protocols to coexist in the protocol specification, yet the IETF has managed to perform similar feats of incompatibility in many more areas. The Internet is composed of many moving parts, constructed and operated by many different players. It works because these components interoperate, and this interoperability is the result of careful and considered work in producing and maintaining a consistent and coherent standard specifications. It seems that this is a fragile state, and the pressures to allow every thought to become a standard specification is devaluing the entire corpus of IETF work over time.

## 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 building the Internet within the Australian academic and research sector in the early 1990's. 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 and chaired a number of IETF Working Groups. He has worked as an Internet researcher, as an ISP systems architect and a network operator at various times.

*www.potaroo.net*

## Disclaimer