

The Internet Protocol Journal

August 2025

Volume 28, Number 2

A Quarterly Technical Publication for
Internet and Intranet Professionals

FROM THE EDITOR

In This Issue

| | |
|-------------------------------|----|
| From the Editor | 1 |
| ShowNet 2024 Highlights | 2 |
| The Root of the DNS..... | 14 |
| Letters to The Editor | 31 |
| In Memoriam: Dave Täht..... | 34 |
| In Memoriam: Fred Baker..... | 35 |
| Fragments..... | 37 |
| Thank You! | 40 |
| Call for Papers..... | 42 |
| Supporters and Sponsors..... | 43 |

You can download IPJ
back issues and find
subscription information at:
www.protocoljournal.org

ISSN 1944-1134

The *TCP/IP Interoperability Conference*—later renamed *Interop*—began as a small workshop in August 1986. It quickly grew in scope to incorporate tutorials, and by 1988 an exhibition network connected 51 exhibitors to each other and to the global Internet. This network was designed and deployed by a group of volunteers, and it became the proving ground for many emerging technologies. In 1994, Interop added Tokyo to its international venues, where 30 years later the conference and exhibition attracts more than 120,000 visitors annually. Following an article in our October 2024 issue describing the history and evolution of the Interop show network, and a second article detailing the Tokyo *ShowNet* in our previous issue, we now bring you the final installment in this series with an article that highlights some of the technology demonstrations performed during the 2024 Interop Tokyo event. The article is by Ryo Nakamura, Haruki Nakamura, Kazuya Okada, and Ryosuke Kato.

The *Domain Name System* (DNS) is one of the core components of the Internet. We have covered many aspects of the DNS over the years, but we have not discussed the *root server system* since an article in Volume 20, No. 2, June 2017. In this issue, Geoff Huston returns to the topic with a detailed tutorial and analysis of today's DNS root server operations.

We always welcome feedback and suggestions on any aspect of this journal. Included in this issue are two Letters to the Editor in response to the IPv6 Transition article in our May 2025 edition. If you'd like to get in touch, send your comments to: ipj@protocoljournal.org.

In late June, I attended the *Internet Governance Forum* (IGF) meeting in Lillestrøm, Norway. Lillestrøm happens to be the place where I attended high school. During my summer breaks, I worked at the nearby *Norwegian Defence Research Establishment* (NDRE), which had one of the first connections to the ARPANET starting in 1973. The IGF exhibition area had a series of posters highlighting the evolution of the Internet in Norway. I was pleased to see that the first poster featured Internet Hall of Fame Inductees Pål Spilling and Yngvar Lundh, my former managers at NDRE. See page 39.

—Ole J. Jacobsen, Editor and Publisher
ole@protocoljournal.org

Technology Highlights of ShowNet 2024

by Ryo Nakamura, Haruki Nakamura, Kazuya Okada, and Ryosuke Kato

Interop Tokyo 2024 was held from June 12 to 14 in the *Makuhari Messe* exhibition halls. With 542 organizations exhibiting and 124,482 visitors attending the exhibition, Interop Tokyo is one of the largest IT shows in Japan. *ShowNet*^[0], the large demonstration network for Interop Tokyo, was also built at the venue. In 2024, the ShowNet comprised approximately 2,300 products and services in more than 20 full-height racks built and operated by 650 engineers including 31 *Network Operations Center* (NOC) team members, 38 volunteer members, and 581 engineers from vendors who contributed their products to ShowNet. These engineers gathered at Makuhari Messe on May 31 and built the network in two weeks. Figure 1 is a picture of the second day of the ShowNet construction in 2024.

Figure 1: A snapshot of ShowNet under construction at Makuhari Messe on June 1, 2024.



The fundamental role of ShowNet is to provide network connectivity to Interop exhibitors and visitors. Furthermore, ShowNet conducts various experiments and demonstrations of new protocols, technologies, and products while serving user traffic.

In 2024, ShowNet featured the following technical topics in each field:

- *Facility*: A high-density *Main Distribution Frame* (MDF) with SN connector-based patch panels^[1].
- *Optical Transport*: Multi-vendor optical transport network with emerging optics such as 400GBASE-ZR+ and XR Optics.
- *Backbone Network*: An SRv6 uSID-based backbone network and *Ethernet VPN* (EVPN) and *Virtual eXtensible Local Area Network* (VXLAN) for access.
- *Data Center and Cloud*: Distributed container clusters and testing lossless networks for *Remote Direct Memory Access* (RDMA) over *Converged Ethernet* (RoCE) traffic.
- *Wireless Network*: Multi-band Wi-Fi access with Wi-Fi6E- and Wi-Fi7-capable Access Points, and multi-vendor *OpenRoaming*^[2].
- *Monitoring*: Integrated monitoring systems with various sensors and user interfaces, and experimentation of how to exploit AI for future monitoring.
- *Security*: Incorporating multiple aspects of protection and hardening such as SASE, ZTNA, EASM, and NGFW.
- *Tester*: Testing upper layers with protocol emulation for routing and penetration tests for security, and demonstrating automating test processes.
- *5G*: Multiple private 5G systems of RAN and cores with multiple vendors, and demonstration of live streaming over the 5G networks.
- *Media-over-IP*: Professional audio and media are now migrating from SDI to IP: demonstrating real-time broadcasting over IP networks.

In this article, we describe four of these topics, namely: *The Backbone Network*, *Optical Transport*, *5G*, and *Media-over-IP*.

The Backbone Network

The backbone network of ShowNet is the core of all the experiments and demonstrations. In 2024, the backbone network was composed of ten routers of nine products listed in Table 1. In addition, two containerized routers, XRd from Cisco Systems and cRPD from Juniper Networks, performed route reflectors for *Border Gateway Protocol* (BGP). With those routers, we built the backbone network based on Segment Routing while conducting SRv6 uSID interoperability tests.

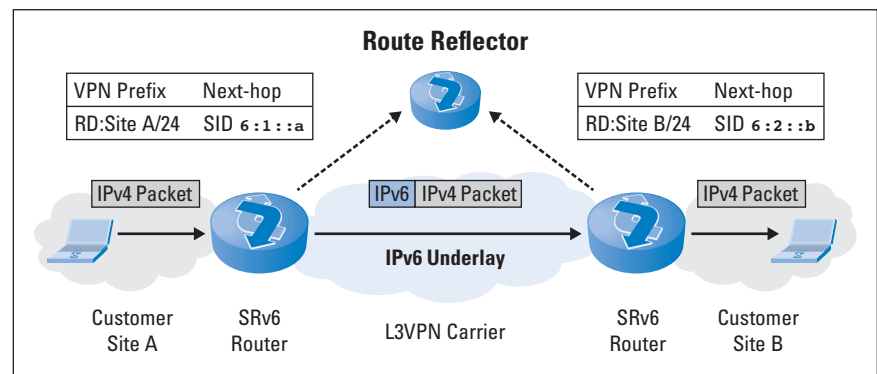
Table 1: Routers composing the backbone network of ShowNet in 2024.

| Vendor | Product |
|---------------------|---------------------------------------|
| Cisco Systems | Cisco 8201-32FH, Cisco 8608, NCS-57B1 |
| Furukawa Electric | FX2 |
| Huawei Technologies | NE8000-M4 |
| Juniper Networks | ACX7348, MX204, MX304, PTX10002-36QDD |

Segment Routing (SR)^[3] is a recent routing and forwarding paradigm that enables source routing. In SR, topological entities are represented by segments; for example, nodes, links, and adjacency. SR nodes control where packets should flow and how packets are processed by embedding a series of segments into a packet. SR has two concrete data-plane implementations: SR-MPLS leveraging *Multi-Protocol Label Switching (MPLS)* labels as *Segment Identifiers (SIDs)* and SRv6 leveraging IPv6 addresses as SIDs. An MPLS label stack encapsulating a packet indicates a SID list in SR-MPLS and IPv6 addresses in a *Segment Routing Header*^[4]—which is a new IPv6 extension header—it also indicates a SID list in the SRv6 data plane.

A major use case of SR is *Layer-3 VPN (L3VPN)*. Figure 2 illustrates a simple example of SRv6-based L3VPN. Two SRv6 routers perform Provider Edge functions for two customer sites, and exchange VPN prefixes via *Multi-Protocol BGP (MP-BGP)*. Note that the next-hops for those VPN prefixes are SRv6 SIDs: the ingress SRv6 router encapsulates packets from the customer site A to site B with IPv6 headers whose destination address is the SID (6:2::b) of the egress SRv6 router.

Figure 2: A simple example of SRv6-based L3VPN.

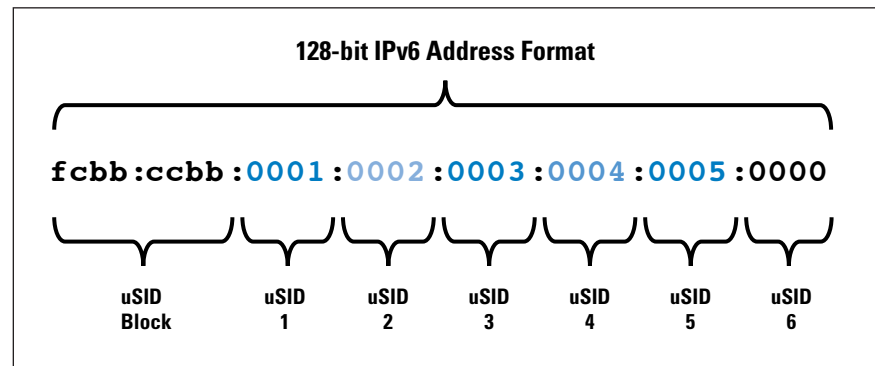


For ShowNet at Interop Tokyo, we have worked on Segment Routing continuously since 2018. In 2018 we conducted a simple and small interoperability test of the SR-MPLS and SRv6 data planes, and in 2019 we demonstrated service chaining over SRv6 with multiple vendors' products. Since 2021, we have deployed SR on the ShowNet backbone networks. The backbone network of ShowNet 2021 was composed of SR-MPLS, and we further conducted a measurement experiment on Internet latency using SR-MPLS-based *Egress Peer Engineering*, which enables steering specific egress traffic to given *External BGP (eBGP)* peers. The results of the experiment were published in a paper^[5] and in an APNIC blog post^[6]. In 2022 and 2023, the ShowNet backbone was fully SRv6-enabled, and IPv4 addresses were eliminated—interfaces of backbone links had no IP addresses configured thanks to IPv6 link-local addresses. Our chronicle with SR was summarized in a presentation at the *Asia Pacific Regional Internet Conference on Operational Technologies (APRICOT) 2024*^[7].

In 2024, a main topic in the backbone network was SRv6 micro SIDs (uSID). uSID, also known as the NEXT-C-SID flavor in^[8], is a mechanism for compressing SID lists in SRv6. A SID in the original SRv6 is a 128-bit IPv6 address, encapsulating packets with multiple SIDs. For example, traffic engineering, involves significant overhead on MTU sizes. uSID encodes multiple SIDs into a 128-bit IPv6 address format to avoid the overhead. Figure 3 illustrates a uSID structure with F3216 format^[9], which implementations must support at present.

The first 32-bit is a uSID block that all routers in an SRv6 domain share. The 16-bit blocks shown in Figure 3 are uSIDs. When an SRv6 node processes the first uSID (**fcbb:ccbb:0001:...**), the node shifts the 80 bits from the second to the last uSID 16 bits to the left and overwrites the first uSID. In other words, the new destination address of the packet is **fcbb:ccbb:0002:0003:0004:0005::**, and the packet is forwarded to the next SRv6 node that has the uSID **0002**. This new packet forwarding mechanism is currently being implemented in router products of multiple vendors, and we confirmed that uSID interoperability between the devices listed in Table 1 was successfully achieved in ShowNet 2024.

Figure 3: A uSID structure with the F3216 format.



The second topic is a demonstration for campus and enterprise networks. The “customers” of ShowNet are exhibitors connecting equipment in their booths to the network. This means that the last hop to the booths consists of several hundred UTP cables spread over the exhibition halls. Accommodating those access circuits becomes a technical demonstration of campus and enterprise networks. This year we built those access networks as L2 and L3VPN with *Ethernet VPN* (EVPN) and *Virtual eXtensible Local Area Network* (VXLAN)^[10] with campus switches from multiple vendors.

VXLAN is an Ethernet-over-IP tunneling protocol, and EVPN is a BGP-based control plane that can construct overlay fabrics using VXLAN as its data plane^[11]. EVPN-VXLAN was originally designed and introduced for data-center use; therefore, switches and routers that were intended primarily for use in data centers supported these protocols in the early days.

Over the years, recent switches for campus and enterprise networks, which are different product lines from those for data centers, have begun to support EVPN-VXLAN for campus use. Adopting Ethernet overlays for campus networks will eliminate (often fragile) spanning-tree protocols and provide scalability and resiliency by using underlying dynamic routing protocols.

The access network in ShowNet 2024 was composed of three routers and eight switches of seven models listed in Table 2. All devices exchanged EVPN routes via route reflectors, constructed a VXLAN fabric, and forwarded user traffic over the fabric. User VLANs could be extended between the switches over the IP underlay. In addition, EVPN can construct L3VPNs using EVPN Type-5 routes^[12]. We also confirmed that the EVPN Type-5 route interoperability works well with these devices.

Table 2: Routers and switches composing the access network with EVPN-VXLAN.

| Vendor | Product |
|---------------------|---------------------------------|
| Cisco Systems | Catalyst 9300, Nexus 93108TC-FX |
| Huawei Technologies | CloudEngine S5732, NE8000 M4 |
| Juniper Networks | EX4400, MX304, SRX4600 |

While SRv6 uSID and EVPN-VXLAN for access were major topics, the demonstrations were not limited to just these two. Other demonstrations and technical challenges were also conducted at the ShowNet backbone network; for example, an experiment of SRv6 over a satellite for disaster recovery, testing *Path Computation Element Protocol* (PCEP), and a total of 2 Tbps external circuits including a capacity of 1.8 Tbps provided by Open APN.

Optical Transport

The optical transport network in ShowNet multiplexes waves on fibers to optimize fiber use while showcasing products in this area. Furthermore, the optical transport network in 2024 faced challenges, including interoperability tests, and tests with other layers above Layer 2. The topics in 2024 were as follows:

- Using multi-band connections of C-band and L-band.
- Interoperability between 400GBASE-ZR+ transceivers based on OpenZR+.
- 1:N point-to-multipoint connections as defined by the *Open XR Optics Forum*.

The optical transport network in ShowNet 2024 consisted of multiple *Wavelength Division Multiplexing* (WDM) networks. One of the WDM networks used a *Reconfigurable Optical Add-Drop Multiplexer* (ROADM) with C-band and L-band wavelengths, connecting transponders and muxponders with capacities ranging from 400 to 800 Gbps.

This WDM network also provided connections of 100GBASE-LR4 and 400GBASE-FR4 to the backbone routers. In addition, we conducted an interoperability test of 400GBASE-ZR+ transceivers at ShowNet. 400GBASE-ZR+^[13] employs coherent optics that enable configuring and transmitting multiple wavelengths so that they can remove transponders. Different manufacturers provide coherent optics equipped with *Digital Signal Processing* (DSP), and we confirmed that they operated correctly in various combinations. Using this infrastructure, we also tried to transfer wavelengths directly from a carrier through the optical transport network built at ShowNet in collaboration with the carrier.

Another WDM network conducted a test of coherent 100GBASE-ZR in the QSFP28 form factor, which was developed after 400GBASE-ZR+ emerged, with ROADMs using C-band wavelengths, in addition to the interoperability of 400GBASE-ZR+ transceivers. Further, we deployed XR Optics^[14], which enables point-to-multipoint optical connections. Deploying Open XR Optics with a ROADM was the first challenge, and it was successfully completed by strong cooperation with each vendor of the transceiver, transponder, ROADM, and *Erbium-Doped Fiber Amplifier* (EDFA) at ShowNet.

5G

Private 5G networks are wholly owned and operated 5G networks that enable individual companies to possess some radio spectrum for their purposes. In Japan, private 5G networks are recognized as Local 5G. This type of private 5G and local 5G is defined as a *Standalone Non-Public Network* (SNPN) in the 3GPP standards. We have been conducting private 5G experiments in a part of ShowNet with 5G-related vendors and integrators since 2022. This year, we deployed three different private 5G networks with multiple vendors and conducted two demonstrations: live streaming in *Network Operations Center* (NOC) guided tours in the exhibition using the 5G networks to improve participants' experience and provided Internet connectivity to several exhibition booths. In addition, we designed a stable and redundant *Precision Time Protocol* (PTP)^[15] network for the 5G networks. In this demonstration, we constructed three private 5G networks that use licensed n79 spectrums in Japan. The demonstration highlighted the advantages of private 5G networks over mobile carriers' 5G services, including low latency and guaranteed access in licensed areas.

NOC guided tours in the exhibition adopted real-time video streaming with the private 5G systems for this year. On the tours, called the *ShowNet Walking Tours*, a NOC team member gives a talk about design concepts and underlying technologies for every rack. However, the areas around the racks were crowded and noisy during the exhibition, so it was difficult for tour participants to see the equipment that NOC members were describing. Furthermore, technologies and devices introduced during the tour were extensive; therefore, conveying this information clearly to the tour participants through only verbal explanations was also challenging.

To address the uncomfortable situation in the tours, we streamed the voice and live movie of the tour guide describing the racks to attendees' 5G-capable tablets and smartphones. Encoded movies and audio were transported to a decode server located in a ShowNet rack via a 5G system. Then, the decoded movie was mixed with supplemental slides and was presented on the attendees' tablets at the right time. An on-premises streaming server delivered the edited movie and audio to attendees' 5G tablets and smartphones via two different 5G networks. Figures 4 and 5 show a camera recording a tour guide describing a ShowNet rack, and the video is mixed with slides. Tour attendees watch the mixed stream, as shown in Figure 6.

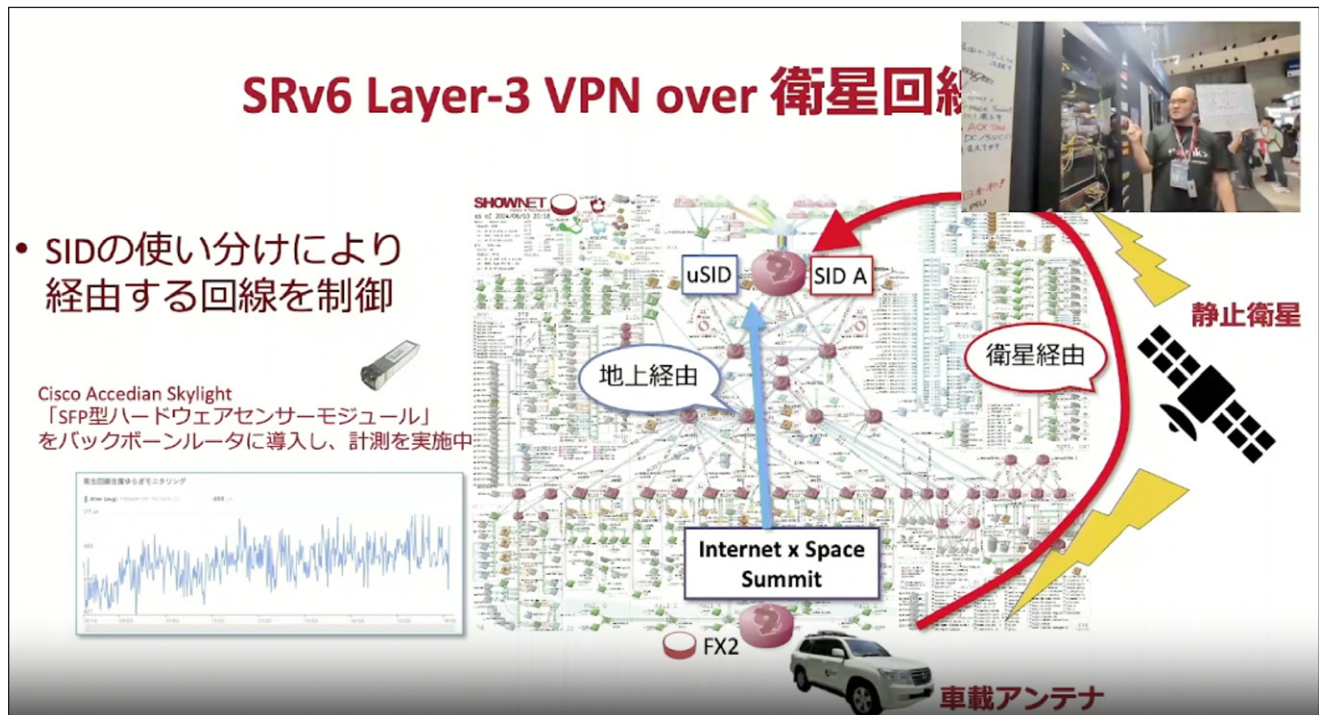
Figure 4: Live broadcasting with private 5G-enabled smartphone cameras.



Figure 5: Mixing the received video image with a slide related to what the NOC member is describing.



Figure 6: Video images delivered to tour attendees' tablets and smartphones.



This demonstration using the 5G systems provided very stable live streaming in the exhibition halls, in contrast to using Wi-Fi. Wi-Fi access was also available, but the Wi-Fi public bands of 2.4 and 5 GHz were already experiencing congestion due to massive numbers of visitors' mobile Wi-Fi devices. Therefore, latency and available network bandwidth were unstable, and it was not easy to provide guaranteed streaming. Wi-Fi 6E, which uses 6-GHz channels, still has not been congested because of the small number of capable devices. However, it is anticipated that this situation will change next year.

Media-over-IP

Professional audio and media are now migrating from *Serial Digital Interface* (SDI) cables, which have low transfer rates and high costs, to Ethernet/IP-based systems for higher transfer rates and lower costs because of the availability of commodity equipment. ShowNet has featured these media-over-IP solutions as one of the main topics since 2022. In 2024, we collaborated with broadcasters pursuing the transition to IP in broadcasting to explore the possibilities of media-over-IP networks and services for broadcasting industries; we attempted to connect and exchange media between the ShowNet booth and three geographically distributed broadcast stations over IP networks.

In the ShowNet booth, we built the *Media Operation Center* (MOC), a broadcast control room for media production and remote operation with IP-based systems. Using this MOC (Figure 7), we demonstrated real-time recording, editing, and broadcasting of a stage (Figure 8) where many sessions were held during the exhibition. This facility also supported live mixing and streaming on the tours with the 5G demonstration described previously.

Media-over-IP technologies are standardized by the *Society of Motion Picture and Television Engineers* (SMPTE)^[16], and its standards are prefixed with SMPTE. For example, the SMPTE ST 2110 series^[17] defines protocols and parameters for professional video, audio, and data-over-IP transport.

From the network viewpoint, that media traffic is *Real-time Transport Protocol* (RTP) streams over IP multicast, and media endpoints speaking the protocols require *Precision Time Protocol* (PTP) to synchronize clocks. Thus, in ShowNet 2024, we built a Layer-3 multicast network with *Open Shortest Path First Version 2* (OSPFv2) and *Protocol Independent Multicast – Sparse Mode* (PIM-SM) for the control room by using Cisco Nexus series and Huawei Cloud Engine switches. These switches are capable of PTP for broadcast profiles (SMPTE ST 2059-2). Furthermore, we configured Layer-2 VPN and Layer-3 VPN connections using VPN devices for media transmission and control between two broadcast stations in Tokyo (30 km away from the venue) and a station in Sapporo (830 km away from the venue) over the Internet. These connections established a remote production environment between the broadcast stations and the MOC booth at ShowNet.

Figure 7: The Media Operation Center at the ShowNet booth.



Figure 8: A stage presentation broadcasted by the media-over-IP systems deployed on ShowNet.



We demonstrated media production with the remote broadcast stations over IP networks during the three-day exhibition. Traffic transferred through the networks included bidirectional uncompressed video streams (SMPTE ST 2110-20, 1080i with 59.94 Hz, up to 1.3 Gbps per stream) and compressed video streams by JPEG-XS (SMPTE ST 2110-22, 1080i with 59.94 Hz, up to 200 Mbps per stream). Additionally, sensors embedded in *Small Form-factor Pluggable* (SFP) modules from Accedian were placed at a ShowNet rack and the broadcast stations to enable active monitoring by *Two-Way Active Measurement Protocol* (TWAMP) measurements. This setup allowed us to observe real-time network performance impacts on media traffic.

During the event, we collaborated with broadcasting industry members to conduct live broadcast and video production of sessions at the exhibition. Eventually, all media transport and equipment operations between the broadcast stations and the Media Operation Center at the ShowNet booth were conducted entirely over IP.

Conclusion

In this article, we introduced technology highlights from ShowNet in 2024. ShowNet covers broader aspects of networking technologies and conducts demonstrations from Layer 1 to Layer 7. Unfortunately, explanations of all the topics discussed in this article are not possible because of the amount of material it would necessitate. So, in this article we covered only four topics: the backbone network, optical transport, 5G, and media-over-IP, and briefly described these technical overviews.

ShowNet is a show in the Interop exhibition; different from ordinary networks, it is an ephemeral network built and operated for just three days. However, we do not let the show network end as just a show. Through conducting various experiments and demonstrations, as described in this article, we aim to encourage network communities in Japan, foster relationships between engineers, and contribute the knowledge and insights gained at ShowNet to society.

Acknowledgments

The design, construction, and demonstration of the ShowNet network were made possible through the collaboration of NOC team members, contributing vendors and their teams, and ShowNet Team members. We want to thank all the people involved in the ShowNet at Interop Tokyo 2024.

References and Further Reading

- [0] Takashi Tomine, Ryo Nakamura, and Ryota Motobayashi, "ShowNet at Interop Tokyo: A Continuously Evolving Demonstration Network," *The Internet Protocol Journal*, Volume 28, No. 1, May 2025.
- [1] SENKO Advance Co., Ltd. SN 1.6mm Standard Connector:
[https://www.senko.com/product/
sn-1-6mm-standard-connector/](https://www.senko.com/product/sn-1-6mm-standard-connector/)

- [2] Wireless Broadband Alliance, OpenRoaming: <https://wballiance.com/openroaming/>
- [3] Clarence Filsfils, Stefano Previdi, Les Ginsberg, Bruno Decraene, Stephane Litkowski, and Rob Shakir, “Segment Routing Architecture,” RFC 8402, July 2018.
- [4] Clarence Filsfils, Darren Dukes, Stefano Previdi, John Leddy, Satoru Matsushima, and Daniel Voyer, “IPv6 Segment Routing Header (SRH),” RFC 8754, March 2020.
- [5] Ryo Nakamura, Kazuki Shimizu, Teppei Kamata, and Cristel Pelsser, “A first measurement with BGP Egress Peer Engineering,” in *Proceedings of 23rd International Conference on Passive and Active Measurement*, PAM 2022, pages 199–215, Springer International Publishing.
- [6] Ryo Nakamura, “Measuring the potential benefit of egress traffic engineering with Segment Routing, *APNIC Blog*, March 10, 2022.
- [7] Teppei Kamata, “Segment Routing Deployments and Demonstrations at Interop Tokyo ShowNet,” Asia Pacific Regional Internet Conference on Operational Technologies (APRICOT) 2024, February 2024.
- [8] Weiqiang Cheng, Clarence Filsfils, Zhenbin Li, Bruno Decraene, and Francois Clad, “Compressed SRv6 Segment List Encoding,” RFC 9800, June 2025.
- [9] Bell Canada, “uSID Address Allocation, How to Assign SRv6 Locators to Network Nodes,” Presentation during IETF srv6 Working Group Meeting at IETF 119, March 2024.
- [10] Mallik Mahalingam, Dinesh Dutt, Kenneth Duda, Puneet Agarwal, Larry Kreeger, T. Sridhar, Mike Bursell, and Chris Wright, “Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks,” RFC 7348, August 2014.
- [11] Ali Sajassi, John Drake, Nabil Bitar, Ravi Shekhar, Jim Uttaro, and Wim Henderickx, “A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN),” RFC 8365, March 2018.
- [12] Jorge Rabadan, Wim Henderickx, John Drake, Wen Lin, and Ali Sajassi, “IP Prefix Advertisement in Ethernet VPN (EVPN),” RFC 9136, October 2021.
- [13] OpenZR+: <https://www.openzrplus.org/>
- [14] Open XR Optics Forum: <https://openxropticsforum.org/>
- [15] IEEE, “1588-2008 IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems,” pages 1–300, 2008.
- [16] SMPTE The home of Media Professionals, Technologists, and Engineers: <https://www.smpte.org/>

- [17] SMPTE, “ST 2110 Suite of Standards”:
<https://www.smpte.org/standards/st2110>
- [18] David Strom, “The Interop ShowNet,” *The Internet Protocol Journal*, Volume 27, No. 3, October 2024.
- [19] Interop 2024 ShowNet concept:
<https://www.interop.jp/2024/shownet/concept/>
- [20] Interop 2024 ShowNet Brochure:
<https://www.interop.jp/2024/assets/file/arukikata.pdf>
- [21] Interop 2024 ShowNet map:
<https://www.interop.jp/2024/assets/file/e-web.pdf>
- [22] ShowNet map icons:
<https://github.com/interop-tokyo-shownet/shownet-icons>
- [23] “Behind the Scenes - Interop Tokyo 2019 ShowNet,” Interop Tokyo YouTube video:
<https://www.youtube.com/watch?v=X-JhPs1T7sc>

RYO NAKAMURA received his Ph.D. degree in Information Science and Technology from the University of Tokyo, Tokyo, Japan, in 2017. He is currently an Associate Professor at the Information Technology Center, the University of Tokyo, where he operates the university’s campus network. His research interests include networking in operating systems, network virtualization, and network operations. Since 2009, he has been involved in Interop Tokyo ShowNet as a ShowNet team member until 2011, and as a member of the NOC team from 2012 to the present. He has been primarily responsible for the backbone network of ShowNet, and he led demonstrations of SDN-related technologies from 2013 to 2017. He can be reached at:
ryo@interop-tokyo.net

HARUKI NAKAMURA received a Master’s degree from Keio University Graduate School of Media Design in 2019. He started his career as a Solutions Engineer at Cisco Systems G.K., focusing on data-center networking and computing technologies. Since 2022, he has expanded his responsibilities to include IP Media Networking and joined Interop Tokyo ShowNet as a contributor. In 2024, he served as a ShowNet NOC member in the Media-over-IP Working Group, where he led initiatives to collaborate with contributors and broadcasting companies on proof-of-concept projects for the next-generation Media-over-IP system. He can be reached at:
hanakamu@interop-tokyo.net

KAZUYA OKADA received a PhD degree in computer science from the Nara Institute of Science and Technology (NAIST), Japan, in 2014. He is currently a Principal Researcher at the InfoTech (Research Division of Information Technology), Toyota Motor Corporation, Japan. His research interests include cyber resilience and cybersecurity for connected vehicles. He has been a NOC team member of ShowNet since 2011. He can be reached at: **okada@interop-tokyo.net**

RYOSUKE KATO has been employed by BroadBand Tower, a Japanese data-center company, since 2013. His role involves the investigation of essential network technologies for data centers, with a focus on IP closed network services between data centers and optical transmission technologies. Additionally, he has been an active member of the NOC team for ShowNet since 2017. He contributed to the demonstration of data-center network interconnection technologies from 2017 to 2019 and was involved in optical transmission technology from 2021 to 2024. He can be reached at: **kato@interop-tokyo.net**

The Root of the DNS

by Geoff Huston, APNIC

The *Domain Name System* (DNS) of the Internet is a remarkably simple system. You send queries into this system via a call to the name resolution library of your local host, and you get answers back. If you peek into the DNS system you'll see exactly the same simplicity: The DNS resolver that receives your query may not know the answer, so it, in turn, will send queries deeper into the system and collect the answers. This query/response process is the same, applied recursively. Simple.

However, the DNS is simple in the same way that Chess or Go are simple. They are all constrained environments governed by a small set of rigid rules, but they all generate surprising complexity in their operation.

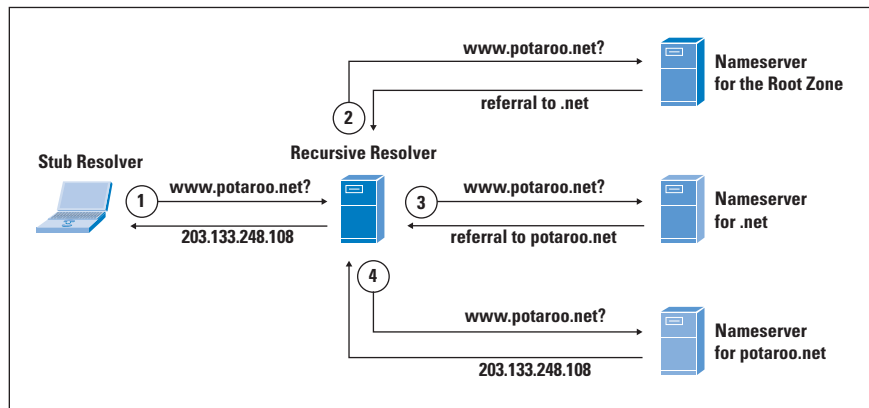
The Root Zone

The DNS is not a dictionary of any natural language, although these days when we use DNS names in our written and spoken communications we might be excused from getting the two concepts confused! The DNS is a hierarchical namespace. Individual domain names are constructed using an ordered sequence of labels. This ordered sequence of labels serves numerous functions, but perhaps most usefully it can be used as an implicit procedure to translate a domain name into an associated attribute value through the DNS name resolution protocol.

For example, I operate a web server that is accessed using the DNS name **www.potaroo.net**. If you direct your browser to load the contents of this DNS name, your system first needs to resolve this DNS name to an IP address, so that your browser knows where to send the IP packets to perform a transaction with my server. But how does the system know which nameserver is authoritative for the zone that includes the name **www.potaroo.net**?

This point is where the structure of the namespace is used to discover the nameserver. In this case, the DNS resolver will query a *root server* to resolve the name. As this name is not defined within the *Root Zone* (the zone that is served by the root servers), the response from any root server to such a query will be a *referral* response. In this example, this response is a redirection that lists the set of nameservers that are authoritative for the **.net** zone. Ask any of these **.net** nameservers for this same DNS name and again you will get back a redirection response, consisting of the list of nameservers that are authoritative for the **potaroo.net** zone. Ask any of these **potaroo.net** nameservers for the same name, **www.potaroo.net**, and you will receive the IP address you are looking for (Figure 1).

Figure 1: Name Resolution in the DNS.



Every DNS name is resolved in the same way. The name itself defines the order of name resolution processing, and it defines the path to be followed through the distributed database that leads to the answer you seek.

In this entire process, there is one starting point for every DNS resolution operation: the *Root Zone*.

Some criticize any exceptional consideration given to the root zone of the DNS; they think it is just another DNS zone, like any other. It is a set of authoritative servers that receive queries and answer them, like any other zone. There is no magic in the root zone, and all this attention on the root zone as *special* in some way is entirely unwarranted.

However, I think this view understates the criticality of the root zone in the DNS. The DNS is a massive, distributed database. Indeed, it is so massive that there is no single static map that identifies every authoritative source of information and the collection of data points about which it is authoritative. Instead, we use a process of dynamic discovery, where the resolution of a DNS name is first directed to locating the authoritative server that has the data relating to the name we want resolved, and then querying this server for the data. The beauty of this system is that these discovery queries and the ultimate query are precisely the same query in every case.

But everyone has to start somewhere. A DNS recursive resolver does not know all the DNS authoritative servers in advance, and it never will. But it does know one thing: It knows the IP address of at least one of the root servers in its provided configuration. From this starting point everything can be constructed in real time. The resolver can ask a root server for the names and IP addresses of all other root servers (the so-called *priming query*), and it can store that answer in a local cache. When the resolver is given a name to resolve, it can then start with a query to a root server to find the next point in the name delegation hierarchy and go on from there in a recursive manner.

If this description illustrates how the DNS actually works, then it is pretty obvious that the entire DNS system would have melted down years ago. What makes this approach viable is *local caching*. A DNS resolver stores the answers in a local cache and uses this locally held information to answer subsequent queries for the life of the cached entry. So perhaps a more refined statement of the role of the root servers is that every DNS resolution operation starts with a query to the cached state of the root zone. If the local cache cannot answer the query, then a root server must be queried.

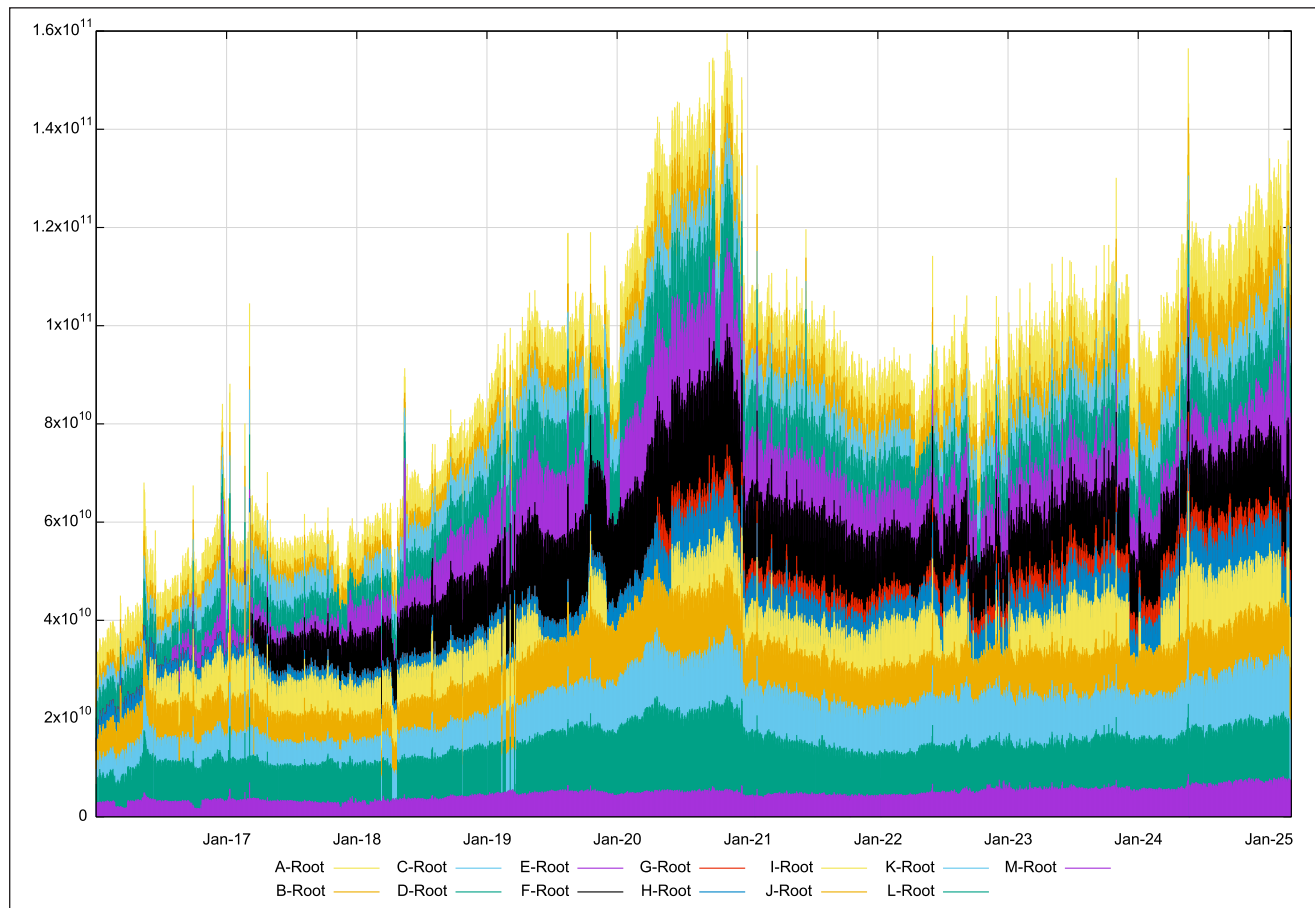
However, behind this statement lurks an uncomfortable observation: If all of the root servers are inaccessible, then the entire DNS ceases to function. This is perhaps a dramatic overstatement in some respects, as there would be no sudden collapse of the DNS and the Internet along with it. In the hypothetical situation where all the instances of the root servers were inaccessible, then DNS resolvers would continue to work using locally cached information. However, as these cached entries time out, they would be discarded from these local resolvers (as they could not be refreshed by re-querying the root servers). The light of the DNS would slowly fade to black bit by bit as these cached entries time out and are removed. The DNS root zone is the master lookup for every other zone. That's why it deserves particular attention. For that reason, the DNS root zone is uniquely different from every other zone.

Root zone servers are not used for every DNS lookup because of local caching. The theory is that the root servers will only see queries as a result of cache misses in resolvers. With a relatively small root zone and a relatively small set of DNS recursive resolvers, the root zone query load should be small. Even as the Internet expands its user base the query load at the root servers does not necessarily rise in direct proportion. It is the number of DNS resolvers that supposedly determines root server query load if we believe in this model of the function of the root in the DNS.

However, the model does not appear to hold up under operational experience. Figure 2 shows the total volume of queries per day recorded by the root servers since January 2016.

Over the period from 2016 to 2020, the volume of queries seen by the collection of root servers tripled. The query volume decreased in 2021 and stabilised over 2022. It is likely that changes to the behaviour of the Chrome browser may explain this abrupt change. Chrome used to probe the local DNS environment by making a sequence of queries to non-existent names (so-called *Chromeoids*) upon startup, and because the query names referred to undelegated top-level domains, these queries were a significant component of the queries seen at the root servers. Changing this behaviour in Chrome at the end of 2020 appears to have resulted in a dramatic change to the DNS query profile as seen by the root servers. However, over 2023 and 2024 the aggregate volume of queries seen by the root servers resumed its upward trend, rising by 40% from some 90 billion queries per day at the start of 2023 to more than 130 billion queries per day at the start of 2025.

Figure 2: Root Service Queries per Day – from ^[1].



What are we doing in response to this trend in the growth of queries to the root zone? How are we ensuring that the root zone service can continue to grow in capacity in response to this resumption in the growth of query rates?

Digression – The Economics of the DNS

In conventional markets, when a good is consumed, the consumer pays the producer a fee for the consumption of that good. As long as the fee covers the cost of production of the good, increasing consumption generates increasing revenue that can cover the costs associated with expanding the means of production of the good. Obviously, that's a very simplistic view of the operation of markets, but the key assumption is that greater consumption generates more revenue for producers, which, in turn, allows producers to produce greater volumes of the good. The essential assumption is that there is an underlying market-based discipline associated with the production and consumption of the good.

This assumption breaks down in the DNS, and in the root zone servers in particular. DNS queries are essentially unfunded. Like many Internet users, I have an *Internet Service Provider* (ISP), and I pay an access fee for its service.

Typically, an ISP operates a DNS recursive resolver for its clients, and my access fee contributes to my ISP's costs in running this resolver service. However, it's a fixed access fee, not a metered fee, so I contribute the same sum to the running of this shared resolver whether I submit one DNS query per day or one million!

As well as the costs in operating this resolver service, does the ISP incur any other cost in operating a DNS service to resolve my queries? No! All of the authoritative nameservers that are queried by my ISP's resolver are not funded by my ISP. More generally, all DNS queries in the public Internet are not directly funded by the querier!

Obviously, there are costs associated with operation of authoritative nameservers, and, for the most part, these costs are met by the "owners" of the zones that are served by these nameservers. There are various funding models for authoritative nameservers, ranging from metered costs per answered query, flat-rate costs, and even free services in some circumstances. But the essential aspect of this service is that authoritative nameservers do not derive revenue from the entities that query them. If there is a revenue stream, it comes from the DNS zone administrators who are paying for the nameservers to serve their zone.

I did note that this fact holds "for the most part," and there is one very notable exception here, namely the root zone. The twelve entities who provide the nameservers for the root zone do so as a collection of independent, largely autonomous volunteers who meet their own costs.

This situation is in many ways a curious relic of an earlier Internet that had a spirit of cooperative enterprise in many of its endeavours, but at the scale where each *Root Service Operator* is operating a service platform capable of responding to an average query load of some 10 billion queries per day, then it is no slight donation of effort and resources to a common-good outcome. Such a core of altruism in the centre of a market-driven frenzy of activity that operates today's digital world is unusual to see.

Given the criticality of the role that these operators collectively undertake, and the observation that directly or indirectly we are all beholden to the outcomes of these efforts to maintain a functional namespace for the Internet, then perhaps, odd as it may be, this situation is better than many of the alternatives.

In a market economy, a monopoly supplier of a critical resource is able to extract a monopoly rental from all others, while customers cannot seek relief through competitive offerings because of the very nature of the monopoly. Today's world looks to market regulators and the associated public regulatory frameworks to protect markets from such forms of abuse. But in the Root Service function we find a service that is both universal across the entire collection of individual public regimes and a collective monopoly.

A self-imposition by these operators of a freely offered service is perhaps not the only possible response to counter such risks of potential abuse of role. So far, however, the ethos of these twelve root service operators has proved to be an adequate and sufficient measure.

But perhaps it's now time to consider the outstanding question, namely "How are we ensuring that the root-zone service can continue to grow in capacity in response to this resumption in the growth of query rates?", and now factor in the apparent need to escalate the level of resources that are in effect donated to this service by this small collection of operators.

Root Zone Scaling

The original model of authoritative servers in the DNS was based on the concept of *unicast* routing. A server name had a single IP address, and this single server was located at a single point in the network. Augmenting server capacity entailed using a larger server and adding network capacity. However, such a model does not address the issues of a single point of vulnerability, nor does it provide an optimal service for distant clients.

More Servers

The DNS approach to this problem is to use multiple nameserver records. A DNS resolver was expected to retry its query with a different server if its original query did not elicit a response. That way, a collection of servers could provide a framework of mutual backup. To address the concept of optimal choice, DNS resolvers were expected to maintain a record of the query/response delay for each of the root servers and prefer to direct the majority of their queries to the fastest server.

Why not use multiple address records for a single common server name? The two approaches (multiple server names and multiple address records for a name) look similar. Once a resolver has assembled a collection of IP addresses that represent the nameservers for a domain, then it seems to me that a resolver could be justified for treating the list of IP addresses consistently, irrespective of whether the list was assembled from multiple IP addresses associated with a single name, or from multiple names. The use of multiple names allows for the use of multiple paths through the DNS to resolve these names of the nameservers that can remove a potential single point of failure, although I wonder as to the true benefit of using a set of nameserver names within a common single DNS zone as compared to using a single name with multiple IPv4 and IPv6 *Resource Records*, particularly when the bulk of DNS zones are provisioned with 2 or 4 nameservers, so there are typically 2 or 4 IPv4 and IPv6 addresses. I suspect that the use of multiple names is a policy compliance outcome rather than a true effort to provision nameservers with resilience through diversity.

If we want to increase the capacity of the root zone, then why not just add more nameserver names to the root zone?

What's so special about this zone's use of 13 named nameservers and a total of 26 IP addresses? For the root zone, the scaling issue with multiple nameservers is the question of completeness and the size of the nameserver response to the *priming query*. The question here is: If a resolver asks for the nameservers of the root zone, should the resolver necessarily be informed of all such servers in the response? The size of the response will increase with the number of servers, and the size of the response may exceed the default maximal DNS over a *User Datagram Protocol* (UDP) payload size of 512 bytes.

The choice of the number of server names for the root zone, 13, was based on the calculation that this was the largest list of a server list that could fit into a DNS response that was under 512 bytes in size. This choice assumed that only the IPv4 address records were being used in the response. With the addition of the IPv6 AAAA records, the response size has expanded. The size of the priming response for the root zone with 13 dual-stack authoritative servers is 823 bytes, or 1,097 bytes if the *Domain Name System Security Extensions* (DNSSEC) signature is included, and slightly larger if DNS cookies are added.

In today's DNS environment, if the query does not include an *Extension Mechanisms for DNS* (EDNS)(0)^[2] indication that they can accept a DNS response over UDP larger than 512 bytes, then the root servers will provide a partial response in any case, usually listing all 13 names, but truncating the list of addresses of these services in the *Additional Section* of the response to fit with a 512-byte payload.

Past experiments have been conducted with more than 13 nameservers at the apex of a DNS-like name system (such as the *Yeti*^[3] project, of some 5–8 years ago), and while it is technically feasible to do so, some vexing questions remain, such as how to select new root service operators, what is a safe ceiling of the number of such services, and how would it impact the stability and coherence of the name system.

Until we have much broader levels of adoption of query name minimisation than we appear to have today, root servers are privy to the myriad of domain names that users are querying. Such data is effectively a real-time view into the activity in the Internet through this meta-data query stream. If we opened up the root service to a broader set of operators, would a temptation to monetise this unique and highly valuable data stream prove overwhelming? In this space is it even possible to enforce constraints that would preclude any such activity?

So far, we appear to have avoided such difficult questions by leaving the number of root nameservers constant and scaling the root service in other ways.

If we can't, or don't want, to just keep on adding more root servers to the nameserver set in the root zone, then what are the other scaling options for serving the root zone?

More Service Platforms

The first set of responses to these scaling issues was in building root servers that have greater network capacity and greater processing throughput. But with just 13 servers to work with, this capacity was never going to scale at the pace of the Internet. We needed something more.

The next scaling step was the conversion from unicast to *anycast*^[32–37] services. There may be 26 unique IP addresses for root servers (13 in IPv4 and 13 in IPv6), but each of these service operators now uses anycast to replicate the root service in different locations. The current number of root server sites is described at root-servers.org (Table 1). Now the routing system is used to optimise the choice of the “closest” location for each root server.

Table 1: Anycast Site Counts for Root Servers, March 2025^[4].

| Root | A | B | C | D | E | F | G | H | I | J | K | L | M | Total |
|-------|----|---|----|-----|-----|-----|---|----|----|-----|-----|-----|----|-------|
| Sites | 59 | 6 | 13 | 220 | 328 | 359 | 6 | 12 | 85 | 148 | 131 | 123 | 23 | 1,513 |

The root server system has embraced anycast, some parts more enthusiastically than others. Currently a total of 1,513 sites have one or more instances of root servers. Some 24 months earlier, in January 2023, the root server site count was 1,396, so that’s an 8% increase in the number of sites in a little over two years.

The number of authoritative server instances is larger than the number of sites, as it is common these days to use multiple server engines within a site and use some form of query distribution at the front end to distribute the incoming query load across multiple back-end engines at each site. Today, the total of root server system instances is 1,907.

Even this form of expanding the distributed service may not be enough in the longer term. We are seeing the resumption of the growth profile last seen in 2016–2020. With a 25% compound annual query growth rate, in four years we may need double the root service capacity from the current levels, and in a further four years we’ll need to double it again. Exponential growth is a very harsh master.

Can this anycast model of replicated root servers expand indefinitely? Or should we look elsewhere for scaling solutions?

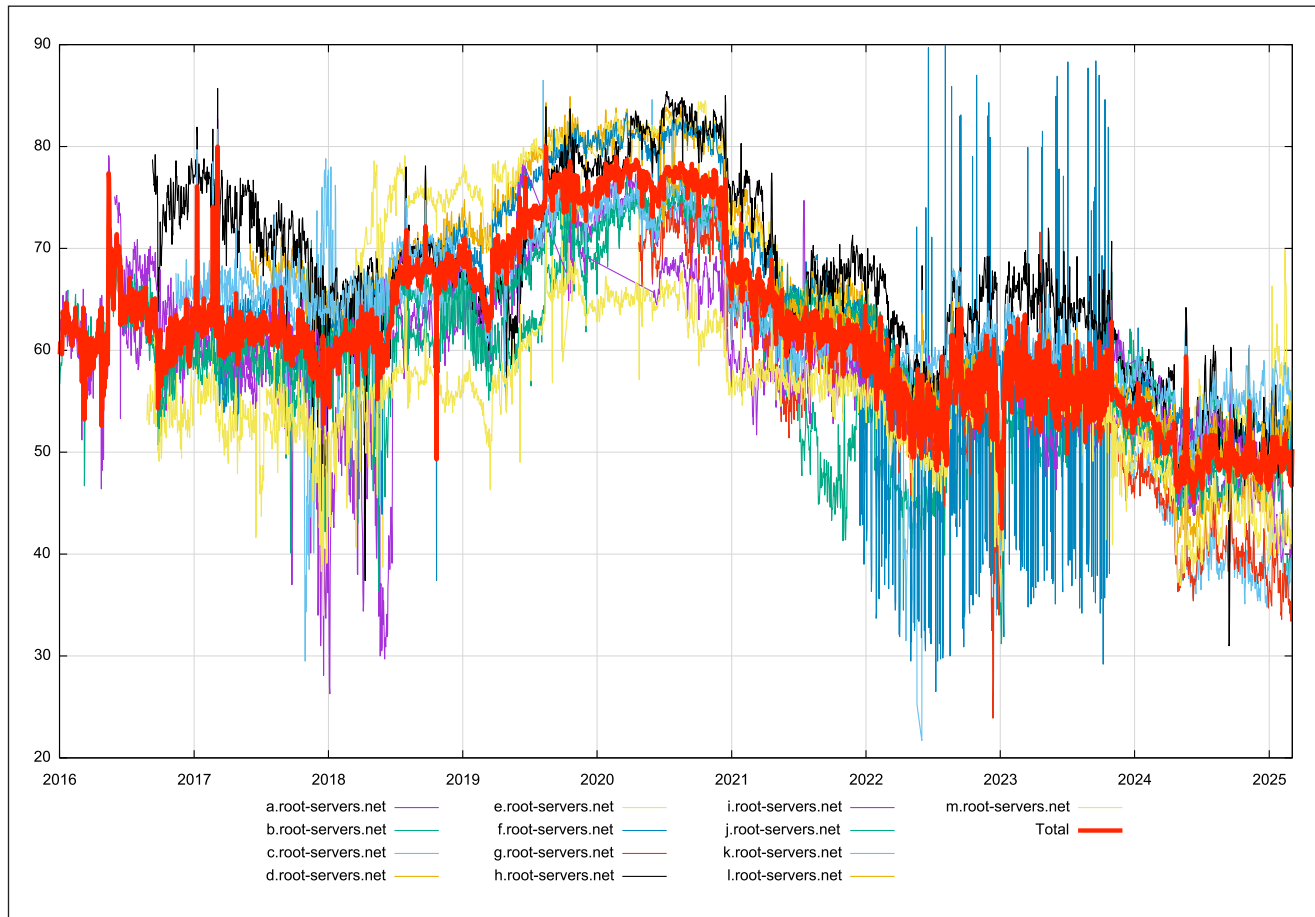
Query Deflection for Negative Responses

There have been many studies of the root service and the behaviour of the DNS over the past few decades. If the root servers were meant simply to respond to the cache misses of DNS resolvers, then whatever is happening at the root is not entirely consistent with such a model of behaviour. Indeed, it’s not clear what is going on at the root!

It has been reported that the majority of queries to the root servers result in NXDOMAIN (“non-existent domain”) error responses.

In looking at the published response code data, it appears that some 50% of root zone queries result in NXDOMAIN responses (Figure 3). The NXDOMAIN response rate was as high as 75% in 2020, and dropped presumably when the default behaviour of the Chrome browser in using Chromeoids changed. In theory these queries are all cache misses at the recursive resolver level, so the problem is that the DNS is not all that effective in handling cases where the name itself does not exist.

Figure 3: Proportion of Root Zone NXDOMAIN Responses per Day ^[1].



If we want to reduce the query pressure on the root servers, one possible approach is to alter the way DNS resolvers handle queries for non-existent names, and in particular names where the top-level label in the queried name is not delegated in the root zone. How else can we deflect these queries away from the root server system?

One such approach is described in RFC 8198^[5], “Aggressive NSEC Caching.” When a top-level label does not exist in a DNSSEC-signed zone and the query has the EDNS(0) DNSSEC “OK” flag enabled, the NXDOMAIN response from a root server includes a signed NSEC record that gives the two labels that exist in the root zone that “surrounds” the non-existent label.

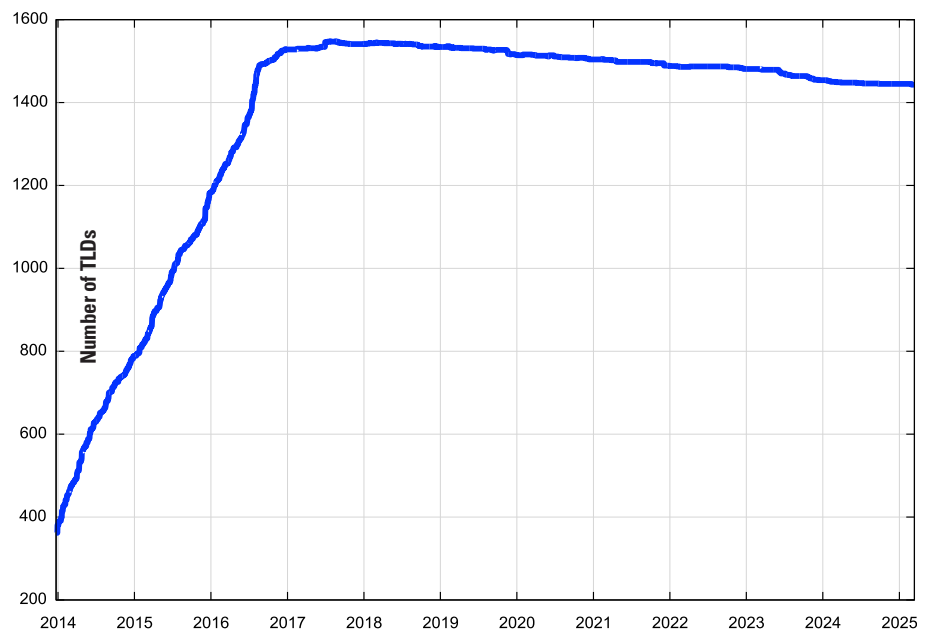
NSEC records say more than “this label is not in this zone.” It says that no label that is lexicographically between these two labels exists. If the recursive resolver caches this NSEC record, then it can use this same cached record to respond to all subsequent queries for names in this label range, in the same way that it conventionally uses “positive” cached records.

If a recursive resolver cached both the 1,443 top-level delegated labels and the 1,444 NSEC records in the root zone, then the resolver would not need to pass any queries to a root server for the lifetime of the cached entries. If all recursive resolvers performed this form of NSEC caching of the root zone, then the query volumes seen at the root from recursive resolvers would fall significantly for non-existent labels.

How Many TLDs Are in the Root Zone?

There were 1,443 *Top-Level Domains* (TLDs) in the root zone of the DNS in March 2025. It has not always been this size. The root zone started with a small set of generic labels, and in the late 1980’s expanded to include the set of two-letter country codes. There were some tentative steps to augment the number of generic top-level domain names, and then in the 2010s *The Internet Corporation for Assigned Names and Numbers* (ICANN) embarked on a larger program of generic TLD expansion. Figure 4 shows the daily count of TLDs in the root zone since 2014.

Figure 4: Daily Count of Root Zone TLDs.



What was surprising to me was that TLDs are not necessarily permanent. The largest TLD count occurred in August 2017, with 1,547 TLDs, and since then the number of TLDs has been declining.

Aggressive use of NSEC caching in recursive resolvers appears to play a contributory role in helping us scale the root zone. *Bind* supports this function as of release 9.12, *Unbound* supports it as of release 1.7.0, and *Knot* resolver supports it as of version 2.0.0. But the queries at the root zone keep growing despite the declining proportion of queries, resulting in an NXDOMAIN response. While this measure may have dampened the relative growth of queries for non-existent names seen at the root servers, to some extent it has not significantly affected the overall problem of the growth of queries directed to the root servers; other factors appear to be causing it.

I'd characterise the situation as aggressive NSEC caching representing a tactical response to root zone scaling concerns, as distinct from a strategic response. The technique is still dependent on the root server infrastructure, and it uses a query-based method of promulgating the contents of the root zone. Nothing really changes in the root service model. What NSEC caching does is allow the resolver to make full use of the information in the NSEC response.

Root Zone Mirroring

Another scaling option is to jump completely out of the query/response model where recursive resolvers incrementally learn the contents of the root zone query-by-query and simply load the entire root zone into their local cache and refresh this local copy with a period of several hours or even a day or so. The idea here is that if a recursive resolver is loaded with a copy of the root zone, then it can operate autonomously with respect to the root servers for the period of validity of the local copy of the root zone contents. It will send no further queries to the root servers.

The procedures to follow to load a local root zone are well documented in RFC 8806^[6], and I should also note here the existence of the *LocalRoot*^[7] service that apparently offers DNS NOTIFY messages when the root zone changes. The root zone is not a big data set. A signed, uncompressed plaintext copy of the root zone as of March 14, 2025, is 2.2 MB in size.

However, this approach has its potential drawbacks. How do you know that the zone you might have received via some form of zone transfer or other is the current genuine root zone? Yes, the zone is signed, but not every element in the zone is signed (NS records for delegated zones are unsigned). The client is left with the task of performing a validation of every digital signature in the zone, and at present there are some 1,444 *Resource Record Digital Signature* (RRSIG) records in the root zone. Even then the client cannot confirm that its local copy of the root zone is complete and authentic because of the unsigned NS delegation records in the root zone.

The IETF published RFC 8976^[8], the specification of a message-digest record for DNS zones, in February 2021. This RFC defines the *Message Digest for DNS Zones* (ZONEMD) record.

What's a Message Digest?

A *Message Digest* is a form of a condensed digital signature of a digital artefact. If the digital artefact has changed in any way, the digest will necessarily change in value as well. If a receiver of this artefact is given the data object and its digest value, then the receiver can be assured, to some extent, that the contents of the file have been unaltered since the digest was generated.

These digital signatures are typically generated using a *Cryptographic Hash Function*. These functions have several useful properties. They are normally a fixed-length output function, so that the resulting value is a fixed size, irrespective of the size of the data for which the hash has been generated.

They constitute a *unidirectional* function, in that knowledge of the hash function value will not provide any assistance in trying to recreate the original data. They are *deterministic*, in that the same hash function applied to the same data will always produce the same hash value. Any form of change to the data should generate a different hash value. Hash functions do not necessarily produce a unique value for each possible data collection, but it should be exhaustively challenging (unfeasible) to synthesise or discover a data set that produces a given hash value (*preimage resistance*), and equally challenging to find or generate two different data sets that have the same hash function value (*collision resistance*).

In other words, an adversary, malicious or otherwise, cannot replace or modify the data set without changing its digest value. Thus, if two data sets have the same digest, one can be relatively confident that they are identical. Second pre-image resistance prevents an attacker from crafting a data set with the same hash as a document the attacker cannot control. Collision resistance prevents an attacker from creating two distinct documents with the same hash.

The root zone includes a ZONEMD record, signed with the *Zone Signing Key* of the root zone. When a client receives the root zone it should look for this record, validate the *Resource Record Digital Signature* (RRSIG) of the ZONEMD record in the same way that it DNSSEC-validates any other RRSIG entry in the root zone, and then compare the value of this record with a locally calculated message digest value of the local copy of the root zone. If the digest values match, then the client has a high level of assurance that this copy of the root zone is authentic and has not been altered in any way.

The dates in the DNSSEC signatures can indicate some level of currency of the data, but further assurance at a finer level of granularity than the built-in key validity dates that the local copy of the root zone data is indeed the current value of the root zone is a little more challenging in this context. DNSSEC does not provide any explicit concept of revocation of prior versions of data, so all “snapshots” of the root zone within the DNSSEC key validity times are equally valid for a client.

The root zone uses a two-week signature validity period (Figure 5).

Figure 5: Root Zone Start of Authority (SOA) Signature.

```
. 86400 IN SOA a.root-servers.net. nstld.verisign-grs.com.
2025031303 1800 900 604800 86400

. 86400 IN RRSIG SOA 8 0 86400 20250326200000 20250313190000
26470 . nYhmvV[...]Ng==
```

This approach of a whole-of-zone signature has some real utility in terms of the distribution of the root zone to DNS resolvers and thereby reduces the dependency on the continuous availability and responsiveness of the root zone servers. The use of the ZONEMD record allows any client to use a local copy of the root zone irrespective of the way in which the zone file was obtained. Within the limits of the authenticated currency of the zone file, as already noted, any party can redistribute a copy of the root zone, and clients of such a redistributed zone can answer queries using this data with some level of confidence that the responses so generated are authentic. It would be useful to augment the existing in-band root zone retrieval using *Authoritative Transfer* (AXFR) with a simple memorable web-retrieval object, such as https://1.2.3.4/root_zone.txt, for example, to allow the zone distribution function to be undertaken by *Content Distribution Networks* (CDNs) as well as by DNS servers.

Resolvers that elect to use a locally managed copy of the root zone can use the ZONEMD record to verify the authenticity of a received root zone. Resolver implementations that perform this verification using ZONEMD include *Unbound* (from v1.13.23) and *PowerDNS Recursor* (from v4.7.04) and *Bind* (v9.17.13).

Notification mechanisms that could prompt a resolver to work from a new copy of the root zone are not addressed in this ZONEMD framework. To me that's the last piece of the framework that could promote every recursive resolver into a peer root server. We've tried numerous approaches to scalable distribution mechanisms over the years. There is the structured *push* mechanism, where clients sign up to a distributor and the distributor pushes updated copies of the data to them. Routing protocols use this mechanism. There also is the *pull* approach, where the client probes its feed point to see if the data has changed and pulls a new copy if it has changed. This mechanism has some scaling issues in that aggressive probing by clients may overwhelm the distributor. We've also seen hybrid approaches where a change indication signal is pushed to the client, and it is up to the client to determine when to pull the new data.

This model of local root zone distribution has the potential to change the nature of the DNS root service, unlike NSEC caching. If there is one thing that we've learned to do astonishingly well in recent times it is distribution of content.

Indeed, we've concentrated on this activity to such an extent that it appears that the entire Internet is nothing more than a small set of CDNs. If the root zone is signed in its entirety with zone signatures that allow a recursive resolver to confirm its validity and currency and is submitted into these distribution systems as just another digital object, then the CDN infrastructure is perfectly capable of feeding this zone to the entire collection of recursive resolvers with ease. Perhaps if we changed the management regime of the root zone to generate a new zone file every 24 hours according to a strict schedule, we could eliminate the entire notification superstructure. Each iteration of the root zone contents would be published 2 hours in advance and it would be valid for a period of precisely 48 hours, for example. At that point the root zone could be served by the existing millions of recursive resolvers rather than the twelve operators and some 2,000 server instances we use today. That's a thousand-fold increase in the capacity of the root system, and at the same time it eliminates the general reliance on a narrow neck of incremental queries being directed to the 12 root server operators that underpin today's DNS.

Futures

We operate the root service in its current framework because it represents a set of compromises that have been functionally adequate so far. That is to say the predominate query-based approach to root zone distribution hasn't visibly collapsed in a screaming heap of broken DNS yet! And it will probably continue to operate in a robust manner for many years to come.

But we don't have to continue relying on this query-based approach just because it hasn't broken so far. Our need to further scale this function is ongoing, and it makes a lot of sense to take a broader view of available options and the just-in-time delivery process used by the DNS incremental query name-resolution algorithm.

We have some choices as to how the root service can evolve and scale.

With Aggressive NSEC Caching we can have recursive resolvers make better use of signed NSEC records and we appear to have staved off some of the more pressing immediate issues about further scaling of the root system. But that's probably not enough.

We can either wait for the DNS system to collapse and then try to salvage the DNS from the broken mess, or perhaps we could explore some alternatives now. For example, we could look at how we can break out of a query-based incremental root content promulgation model and view the root zone as just another content "blob" in the larger ecosystem of content distribution. If we can cost-efficiently load every recursive resolver with a current copy of the root zone, and these days that's not even a remotely challenging target, then perhaps we can put aside the issues of how to scale the root server system to serve ever greater volumes of queries to ever more demanding clients, and perhaps also provide an alternate answer to the continual questions about the politics and finances relating to root servers and their operation.

The reason why content distribution networks have revolutionised the Internet in recent years is that pre-provisioning at the edge makes for a faster, cheaper, and more scalable network in the current context of abundant computing and storage capabilities. If we are prepared to allow this same thinking to intrude into the way we provision the DNS, I suspect we could realise similar benefits for the DNS as well.

Disclaimer

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

References and Further Reading

- [0] ICANN Root Server System Advisory Committee (RSSAC), “RSSAC Advisory on Measurements of the Root Server System,” RSSAC002, June 1, 2016.
- [1] A collection of collected RSSAC002 data:
<https://github.com/rssac-caucus/RSSAC002-data>
- [2] Joao Damas, Michael Graff, and Paul Vixie, “Extension Mechanisms for DNS (EDNS(0)),” RFC 6891, April 2013.
- [3] The Yeti Project: <https://yeti-dns.org/>
- [4] <https://root-servers.org/>
- [5] Kazunori Fujiwara, Akira Kato, and Warren Kumari, “Aggressive Use of DNSSEC-Validated Cache,” RFC 8198, July 2017.
- [6] Warren Kumari and Paul Hoffman, “Running a Root Server Local to a Resolver,” RFC 8806, June 2020.
- [7] *LocalRoot*—Serve Yourself the Root:
<https://localroot.isi.edu/>
- [8] Duane Wessels, Piet Barber, Matt Weinberg, Warren Kumari, and Wes Hardaker, “Message Digest for DNS Zones,” RFC 8976, February 2021.
- [9] Jon Postel, “Internet Name Server,” IEN 61, October 1978.
- [10] Paul Mockapetris, “Domain names: Concepts and facilities,” RFC 882, November 1983.
- [11] Paul Mockapetris, “Domain names: Implementation specification,” RFC 883, November 1983.
- [12] Zi Hu, Liang Zhu, John Heidemann, Allison Mankin, Duane Wessels, and Paul Hoffman, “Specification for DNS over Transport Layer Security (TLS),” RFC 7858, May 2016.
- [13] Christian Huitema, Sara Dickinson, and Allison Mankin, “DNS over Dedicated QUIC Connections,” RFC 9250, May 2022.
- [14] Paul Hoffman and Patrick McManus, “DNS Queries over HTTPS (DoH),” RFC 8484, October 2018.
- [15] Eric Kinnear, Patrick McManus, Tommy Pauly, Tanya Verma, and Christopher A. Wood, “Oblivious DNS over HTTPS,” RFC 9230, June 2022.

- [16] Stephane Bortzmeyer, “DNS Query Name Minimisation to Improve Privacy,” RFC 7816, March 2016.
- [17] Roy Arends, Rob Austein, Matt Larson, Dan Massey, and Scott Rose, “DNS Security Introduction and Requirements,” RFC 4033, March 2005.
- [18] Ben Schwartz, Mike Bishop, and Erik Nygren, “Service Binding and Parameter Specification via the DNS (SVCB and HTTPS Resource Records),” RFC 9460, November 2023.
- [19] Ben Schwartz, “Service Binding Mapping for DNS Servers,” RFC 9461, November 2023.
- [20] Carlo Contavalli, Wilmer van der Gaast, David C. Lawrence, and Warren Kumari, “Client Subnet in DNS Queries,” RFC 7871, May 2016.
- [21] Miek Gieben, “DNSSEC: The Protocol, Deployment, and a Bit of Development,” *The Internet Protocol Journal*, Volume 7, No. 2, June 2004.
- [22] Richard Barnes, “Let the Names Speak for Themselves: Improving Domain Name Authentication with DNSSEC and DANE,” *The Internet Protocol Journal*, Volume 15, No.1, March 2012.
- [23] M. Stuart Lynn, “A Unique Root,” *The Internet Protocol Journal*, Volume 4, No. 3, September 2001.
- [24] Geoff Huston, “A Question of DNS Protocols,” *The Internet Protocol Journal*, Volume 17, No. 1, September 2014.
- [25] Geoff Huston, “Scaling the Root,” *The Internet Protocol Journal*, Volume 18, No. 1, March 2015.
- [26] Geoff Huston, “What’s in a DNS Name?” *The Internet Protocol Journal*, Volume 19, No. 1, March 2016.
- [27] Geoff Huston, “The Root of the DNS,” *The Internet Protocol Journal*, Volume 20, No. 2, June 2017.
- [28] Geoff Huston, “DNS Privacy and the IETF,” *The Internet Protocol Journal*, Volume 22, No. 2, July 2019.
- [29] Geoff Huston, “DNS Trends,” *The Internet Protocol Journal*, Volume 24, No. 1, March 2021.
- [30] Geoff Huston, “DNS Evolution,” *The Internet Protocol Journal*, Volume 27, No. 2, July 2024.
- [31] Burton Kaliski Jr., “Minimized DNS Resolution: Into the Penumbra,” *The Internet Protocol Journal*, Volume 25, No. 3, December 2022.
- [32] Craig Partridge, Trevor Mendez, and Walter Milliken, “Host Anycasting Service,” RFC 1546, November 1993.
- [33] David B. Johnson and Steve Deering, “Reserved IPv6 Subnet Anycast Addresses,” RFC 2526, March 1999.
- [34] Dino Farinacci and Yiqun Cai, “Anycast-RP Using Protocol Independent Multicast (PIM),” RFC 4610, August 2006.

- [35] Joe Abley and Kurt Eric Lindqvist, “Operation of Anycast Services,” RFC 4786, December 2006.
- [36] Danny McPherson, Dave Oran, Dave Thaler, and Eric Osterweil, “Architectural Considerations of IP Anycast,” RFC 7094, January 2014.
- [37] Sebastian Kiesel and Reinaldo Penno, “Port Control Protocol (PCP) Anycast Addresses,” RFC 7723, January 2016.

GEOFF HUSTON AM, 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 1990s. He is author of numerous Internet-related books, and was a member of the Internet Architecture Board from 1999 until 2005. He served on the Board of Trustees of the Internet Society from 1992 until 2001. At various times Geoff has worked as an Internet researcher, an ISP systems architect, and a network operator. E-mail: gih@apnic.net

Our Privacy Policy

The *General Data Protection Regulation* (GDPR) is a regulation for data protection and privacy for all individual citizens of the *European Union* (EU) and the *European Economic Area* (EEA). Its implementation in May 2018 led many organizations worldwide to post or update privacy statements regarding how they handle information collected in the course of business. Such statements tend to be long and include carefully crafted legal language. We realize that we may need to provide similar language on our website and in the printed edition, but until such a statement has been developed here is an explanation of how we use any information you have supplied relating to your subscription:

- The mailing list for *The Internet Protocol Journal* (IPJ) is entirely “opt in.” We never have and never will use mailing lists from other organizations for any purpose.
- You may unsubscribe at any time using our online subscription system or by contacting us via e-mail. We will honor any request to remove your name and contact information from our database.
- We will use your contact information only to communicate with you about your subscription; for example, to inform you that a new issue is available, that your subscription needs to be renewed, or that your printed copy has been returned to us as undeliverable by the postal authorities.
- We will never use your contact information for any other purpose or provide the subscription list to any third party other than for the purpose of distributing IPJ by post or by electronic means.
- If you make a donation in support of the journal, your name will be listed on our website and in print unless you tell us otherwise.

Letters to the Editor

As a long-time subscriber to the *Internet Protocol Journal*, I have always found the articles to be timely, extensively researched, and presented with clarity. Because of these qualities I've often used IPJ articles as reading assignments to my students in several of the courses that I teach. They give students a depth that is not too technical but yet contains enough of the necessary technical details that help them better understand the technology while also grasping the impact of the technology in the context of how it is being used today while looking ahead to the future.

I especially enjoy reading and sometimes sharing articles by Geoff Huston. Geoff has a real knack for presenting technical details within the larger scope of “how we got here and where we are going” that always makes his articles a pleasure to read.

However, Geoff's article on “The IPv6 Transition” in Volume 28, No. 1, May 2025 was especially beneficial. Students often ask me why, after all these years, IPv6 still plays “second fiddle” to IPv4 and is not more widely adopted. Geoff's article does a superb job of explaining the numerous reasons behind the slow transition to IPv6. And his analysis of how today's Internet is moving away from a strict address-based architecture offers an excellent assessment as to what lies ahead in the future for IPv6.

And the timeliness of Geoff's article could not have been better: after arriving in my email inbox that afternoon I had enough time to add some of his observations to my class discussion the very next morning!

Thanks, Geoff, and keep up the good work!

Regards,

—Dr. Mark Ciampa
Professor, Western Kentucky University
Bowling Green, KY
mark.ciampa@wku.edu

The Author responds:

Thank you Mark for your kind words. The Internet has not followed a path driven as much by market pressures as it is by technical evolution, and the outcomes are often surprising. This topic was the main theme of my article. One thing is sure, however, that the pressures to innovate will continue, and tomorrow will be as surprising as today!

Kind regards,

—Geoff Huston
gih@apnic.net

Hi Ole,

As always, I enjoy reading Geoff's articles in IPJ and I appreciate Geoff's continuing to write and your willingness to publish his material. What I so appreciate, Ole and Geoff, about IPJ is the *context* which you provide that gives me the larger frame/larger picture into which to place much of what I do. And even when I don't specifically need the context—I don't do anything with cellular networks for example—I find the articles fun reading regardless.

My “two bits” on Geoff's “IPv6 Transition” article: I logged into my first network-attached computer in 1981, as a student. I configured my first IP router in 1991, and I have spent my career to date supporting IT infrastructure (compute/network/storage) for academic or other non-profit research institutes, mostly in the life sciences. My perspective over the years, as I have sat in seminars at Interop or read news or attended internal meetings about IPv4 address exhaustion and the need for IPv6:

- Adding IPv6 support to our network would take money, staff hours, and training time, not only from network engineers but also from desktop, server, and storage system engineers; smells like a lot of effort to me.
- I find it difficult to prioritize distant risks over immediate priorities.
- Our user base has yet to ask for access to an IPv6-only resource (such a request would affect how we prioritize, but I have not seen even one).
- I suspect that any institution that wants to make a resource broadly available will invest significant effort into making it available via IPv4 because there are so many IPv4 users out there.

As a result, I have yet to configure any device to support IPv6. I am not opposed to IPv6 ... but since I still get up at 2am when my [pager](#) phone buzzes, in response to our network malfunctioning, I have—reasonably I propose—allocated my time to other priorities.

—Stuart Kendrick

Allen Institute, Seattle, WA USA
stuartk@alleninstitute.org

The Author responds:

Hi Stuart. We share a similar vintage, as I first logged into a computer as a student in 1976 (A Sperry Rand Univac mainframe), although building a network connection to the Internet for all Australian Universities would take a further 13 years, when the project that I was leading, AARNet, had managed to complete its initial mission.

In the early 1990's when the IETF was debating the approach to be used for the "next generation" IP protocol, there were many general approaches. One approach, originally called "SIP" was intended to change the IPv4 design as little as possible. It lengthened the address fields to 128 bits, but not much else changed. Other approaches described a more radical set of design changes. SIP won the day, and IPv6 is, to all intents and purposes, just IPv4 with bigger address fields.

In other words, IPv6 was not intrinsically "better" than IPv4 for any particular use case. It wasn't intrinsically faster, nor more secure, nor more agile. It just had bigger address fields. The result was that deploying IPv6 did not provide a network operator with a compelling competitive product. If a network operator already had secured ample pools of IPv4 addresses, then it was not necessarily impacted by IPv4 address scarcity, and the case for incurring the cost of deploying IPv6 in a dual-stack network scenario was extremely challenging to make. A direct result of this situation is the protracted transition to IPv6 for many parts of the Internet, an issue I explored in this article.

Frankly, it does not really make much sense to comment on the IPv6 design as "right" or "wrong." It represented the common wisdom of the IETF at that time. However, we did fail to predict just how long the dual-stack transition was going to take, or even if this transition would ever come to an end. Some 30 years after we started down this path I suspect we are no closer to answering these two very fundamental questions.

Regards,

—*Geoff Huston*
gih@apnic.net

Check your Subscription Details!

Make sure that both your postal and e-mail addresses are up-to-date since these are the only methods by which we can contact you. If you see the words "Invalid E-mail" on your printed copy, this means that we have been unable to contact you through the e-mail address on file. If this is the case, please contact us at **ipj@protocoljournal.org** with your new information. The subscription portal is located here:
<https://www.ipjsubscription.org/>

In Memoriam



Dave Täht

Dave Täht, formerly known as Michael David Täht (August 11, 1965 – April 1, 2025) was our friend, colleague, and mentor at LibreQoS. To the rest of the world, Dave was an American network engineer, musician, lecturer, asteroid exploration advocate, Internet activist, and much more.^[1]

The fruits of Dave’s work are everywhere. Most people will never notice—a testament to his engineering. Dave’s work on creating algorithms like *Flow Queueing with Controlled Delay*^[2] and *Common Applications Kept Enhanced* (CAKE)^[3] was instrumental, and now it’s part of Linux, OpenWrt, and Starlink; mainstream networking equipment vendors like MikroTik use it as well.

Dave and Jim Gettys spearheaded the networking industry’s effort to eliminate bufferbloat, latency, and jitter on today’s interactive Internet, where bandwidth matters less.^[4] Around 2010, Dave was semi-retired in Nicaragua—and Jim in the USA. They independently came to the realization that Voice over IP and videoconferencing were suffering from the same issues: *lag* and *jitter*, caused by the proverbial time difference between Dave speaking on one continent and Jim hearing his voice on another. Jim coined the term “bufferbloat” to describe the culprit: the extensive and ever-increasing size of buffering on network devices. They started **Bufferbloat.net** and began solving the problem.

Besides his work on solving bufferbloat, Dave also spent years in Nicaragua trying to find ways to bring the Internet (and power, lighting, food, medicine, and books) as an outgrowth of Nicholas Negroponte’s *One Laptop Per Child Project*^[5].

Dave was also known for his little ditties, songs he liked to play while presenting at conferences or podcasts. He wrote “One First Landing,” for example, to cheer up people at SpaceX when they were not doing well with landing their rocket, and he was forced to rewrite it when they started to land their rockets successfully.^[6] And he was happy to do it!

For the last couple of years Dave lived on a boat in Half Moon Bay, a small city on the California coast, south of San Francisco; Dave always liked to sail.

Ad astra per aspera, Dave, you are an astronaut now!

—Robert, Herbert, and Frank - LibreQoS

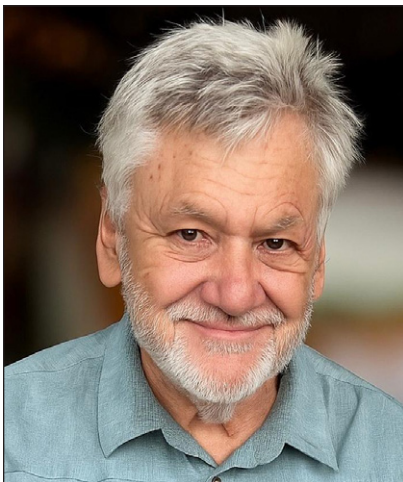
LibreQoS is an Open Source project founded by Robert Chacón (and quickly joined by Dave Täht). LibreQoS provides a drop-in middlebox for *Internet Service Providers* (ISPs), applying CAKE to all of the ISP’s users, as well as an array of network monitoring tools.

References

- [1] Wikipedia article on Dave Täht, May 2025:
https://en.wikipedia.org/wiki/Dave_T%C3%A4ht
- [2] Toke Hoeiland-Joergensen, Paul McKenney, Dave Täht, Jim Gettys, and Eric Dumazet, “The Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm,” RFC 8290, January 2018.
- [3] Toke Høiland-Jørgensen, Dave Täht, and Jonathan Morton, “Piece of CAKE: A Comprehensive Queue Management Solution for Home Gateways.” <https://arxiv.org/abs/1804.07617>
- [4] Jim Gettys, “The Blind Men and the Elephant,” Jim Gettys’ ramblings on random topics, and occasional rants, February 11, 2018. <https://gettys.wordpress.com/2018/02/11/the-blind-men-and-the-elephant/>
- [5] Wikipedia article on One Laptop per Child:
https://en.wikipedia.org/wiki/One_Laptop_per_Child
- [6] YouTube video, “One First Landing (Thank you SpaceX for a Wonderful Year!).”
<https://www.youtube.com/watch?v=wjurORG-v-I>

“In my own mind, I like to think of him as the person who added the most effective capacity to the Internet.”

—Karl Auerbach



Frederick Juergens Baker

The Internet has lost a generous long-time contributor. Frederick Juergens Baker (February 28, 1952 – June 18, 2025)^[1] was one of the original members of the *Internet Systems Corporation* (ISC) Board of Directors, appointed at ISC’s incorporation in 1994.

Fred had a long career in the communications industry, working for Control Data Corporation, Vitalink Communications, Advanced Computer Communications, and for 22 years, at Cisco Systems.

After retiring from Cisco, Fred worked as a contractor, notably for the Internet Society and ISC. In addition to serving on the ISC BOD, in 2017 he joined the *Root Server System Advisory Committee* (RSSAC) of the *Internet Corporation for Assigned Names and Numbers* (ICANN), representing ISC. He served as co-chair of RSSAC from October 2018 to December 2019, and as chair from January 2020 through December 2022.

Fred volunteered a lot of his time to working with the *Internet Engineering Task Force* (IETF), the body that develops standards for the Internet. He chaired numerous IETF working groups, including several that specified the *Management Information Bases* (MIB) used to manage network bridges and popular telecommunications links, and the IPv6 Operations working group.

He served as IETF chair from 1996 to 2001, and he served on the Internet Architecture Board from 1996 through 2002. Fred co-authored or edited at least 60 *Request for Comments* (RFC) documents ^[2,3] on Internet protocols, and contributed to others. The subjects covered include network management, *Open Shortest Path First* (OSPF) and *Routing Information Protocol* (RIPv2) routing, *Quality of Service* (using both the *Integrated Services* and *Differentiated Services* models), *Lawful Interception*, precedence-based services on the Internet, and others.

In addition, Fred served as a member of the Board of Trustees of the Internet Society from 2002 through 2008, and as its chair from 2002 through 2006. He was a member of the *Technical Advisory Council* of the US *Federal Communications Commission* in 2004. He worked as a liaison to other standards organizations such as the ITU-T. In 2009–2010, he served as chair of the *RFC Series Oversight Committee*.

He represented IETF on the *National Institute of Standards and Technology Smart Grid Interoperability Panel* and the *Architecture Committee* from 2008 through 2013, and was Cisco's representative to the *Broadband Internet Technical Advisory Group* (BITAG). He also holds several patents.

Fred was committed to the collaborative, consensus-driven process of creating open standards for the Internet, and he demonstrated his commitment throughout his long career with years of active volunteering. Besides his leadership roles, he also welcomed and mentored new participants in the IETF.

Fred was a wonderful guy, an Internet luminary, and a great friend to ISC over the course of decades of board membership, as well as representing ISC at RSSAC as chair and in many other roles in the IETF, ICANN, and ISOC worlds. We all will dearly miss him.

We extend our deepest condolences to Fred's family.

—Jeff Osborn, ISC
jsosborn@isc.org

References

- [1] Wikipedia article on Fred Baker:
[https://en.wikipedia.org/wiki/Fred_Baker_\(engineer\)](https://en.wikipedia.org/wiki/Fred_Baker_(engineer))
- [2] RFCs authored or co-authored by Fred Baker:
<https://datatracker.ietf.org/doc/search?name=&sort=&rfcs=on&activedrafts=on&by=author&author=fred+baker>
- [3] Datatracker Profile for Fred Baker:
<https://datatracker.ietf.org/person/fredbaker.ietf@gmail.com>
- [4] In Memory of Fred Baker, *Ever Loved*:
<https://everloved.com/life-of/frederick-baker/>

IETF-developed MLS set to be used on 100s of Millions of Mobile Devices

Less than two years after *Messaging Layer Security* (MLS) was published as an RFC^[0], it is poised to be deployed on Android phones and Apple iPhones and other devices, thanks to newly updated RCS specifications, enabling interoperable encryption between different platform providers for the first time.

The GSM Association^[1] announced that the latest *Rich Communications Services* (RCS) standard includes end-to-end encryption based on the MLS protocol. RCS enhances traditional SMS messaging by offering a suite of service capabilities, including group chat, file transfers, typing notifications, and more. Key stakeholders for RCS implementation include device manufacturers, telecommunications operators, and business service providers.

MLS, developed by the IETF *Messaging Layer Security Working Group*^[2], provides unsurpassed security and privacy for users of group communications applications. Using MLS, participants always know which other members of a group will receive the messages they send, and the validity of new participants joining a group is verified by all the other participants. During its development^[3] in the IETF, MLS underwent formal security analysis and industry review. It currently supports multiple cipher suites, and makes it straightforward to add quantum attack resistant cipher suites in the future^[4].

The open processes and “running code” that are hallmarks of the IETF, mean that MLS is already proven to be efficient at Internet scale, working efficiently with groups that have thousands of participants. MLS is already available from, and implemented and deployed by a wide range of companies and organizations^[5]. This includes real-time platforms such as Webex, Wire, and Discord, as well as in devices such as drones.

MLS is also extensible, meaning it can be easily updated in a number of ways. Work is continuing in the MLS Working Group in a number of areas and the IETF *More Instant Messaging Interoperability* (mimi)^[6] working group is looking to build on MLS as they aim to specify the minimal set of mechanisms required to make Internet messaging services interoperable.

[0] Richard Barnes, Benjamin Beurdouche, Raphael Robert, Jon Millican, Emad Omara, and Katriel Cohn-Gordon, “The Messaging Layer Security (MLS) Protocol,” RFC 9420, July 2023.

[1] Tom Van Pelt, GSMA, “RCS Encryption: A Leap Towards Secure and Interoperable Messaging,”
<https://www.gsma.com/newsroom/article/rcs-encryption-a-leap-towards-secure-and-interoperable-messaging/>

[2] IETF Messaging Layer Security Working Group:
<https://datatracker.ietf.org/wg/mls/about/>

- [3] Nick Sullivan and Sean Turner, “Messaging Layer Security: Secure and Usable End-to-End Encryption,” *IETF Blog*, March 29, 2023.
- [4] Rohan Mahy and Richard Barnes, “ML-KEM and Hybrid Cipher Suites for Messaging Layer Security,” Internet-Draft, Work in Progress, **draft-mahy-mls-pq-00**, March 2025.
- [5] “Support for MLS,” *IETF Blog*, July 18, 2023.
<https://www.ietf.org/blog/support-for-mls-2023/>
- [6] mimi Working Group:
<https://datatracker.ietf.org/group/mimi/about/>

ICANN and ISOC Joint Report on 20 Years of IGF

The *Internet Corporation for Assigned Names and Numbers* (ICANN) and the *Internet Society* (ISOC) recently released a report that offers a substantive look at the global impact of the *Internet Governance Forum* (IGF). It demonstrates how coordination—rather than control—has driven tangible progress in the Internet’s resilience, reach, and trust. Structured not as a retrospective but as a practical record of outcomes, the report draws from two decades of work across infrastructure, access, security, and policy. It offers grounded evidence of what coordination has made possible and what could be lost if support for multistakeholder cooperation erodes. The key insights of the report are as follows:

- *Infrastructure and Access*: In Africa, *Internet Exchange Points* (IXPs) more than doubled in over a decade. In countries like Kenya and Nigeria, this growth helped localize traffic, cutting the delay in data travel (latency) from around 200–600 milliseconds to 2–10 milliseconds, and saving millions annually in international connectivity costs. The IGF enabled the sharing of best practices that directly contributed to this expansion.
- *Multilingual Access*: Nearly 4.4 million domain names are now registered in non-Latin scripts. Through IGF-hosted sessions and stakeholder coalitions, *Internationalized Domain Names* (IDNs) and *Universal Acceptance* (UA) have gained critical momentum. In 2025, more than 50 global events marked *UA Day*, promoting linguistic access across the Internet ecosystem.
- *Security and Resilience*: Today, 93% of top-level domains are secured using *Domain Name System Security Extensions* (DNSSEC), which protect against forged DNS responses. In parallel, over 1,000 networks have adopted the *Mutually Agreed Norms for Routing Security* (MANRS), a global initiative to reduce routing attacks. The IGF has catalyzed awareness, collaboration, and implementation of these safeguards.
- *Local Engagement and Policy Influence*: More than 180 *National and Regional IGFs* (NRI) now form a decentralized backbone of year-round Internet governance dialogue. Initiatives like *Youth IGFs* and the *IGF Parliamentary Track* are shaping national and international policy—including formal declarations on digital trust, user rights, and multistakeholder governance.

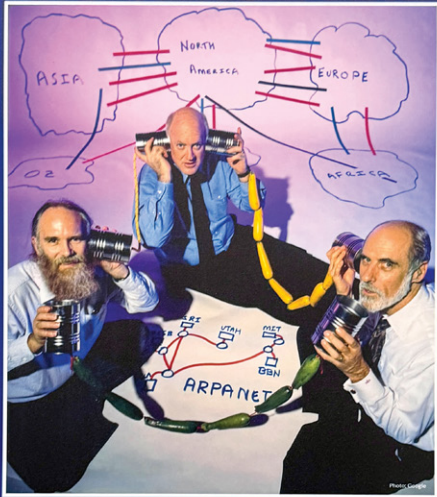
- *Community-Centric Innovation*: From the Arctic to the Andes, community networks have grown through IGF platforms and the *Dynamic Coalition on Community Connectivity* (DC3). These grass-roots efforts now inform regulatory change, including *International Telecommunication Union* (ITU) resolutions and national endorsements, and have helped close connectivity gaps in underserved regions.
- *Global Coordination Platform*: The IGF has evolved from an annual convening into a living ecosystem. It bridges technical and policy domains, connects local with global perspectives, and enables distributed but aligned Internet governance. That structure is now both a model and a necessity.



The report launches ahead of the 20-year review of the *World Summit on the Information Society* (WSIS+20)—a pivotal moment that will shape the next phase of global digital cooperation. It serves as both a record of achievement and a warning: coordination works but is not self-sustaining. The Internet's openness, security, and interoperability depend on it. If that cooperation falters, the conditions that have made the Internet thrive may not hold. The full report is available here:

https://www.internetsociety.org/wp-content/uploads/2025/06/20-Years-of-IGF_EN.pdf

Poster featuring Internet Pioneers Pål Spilling and Yngvar Lundh displayed at IGF 2025 in Lillestrøm, Norway.

1973



Pål Spilling, Yngvar Lundh and Dag Belsnes took part in the development of the TCP/IP protocol and established with Vint Cerf and his team the first connection to the Internet (Arpanet) at Kjeller – visit us on Tuesday and Thursday.

Thank You!

Publication of IPJ is made possible by organizations and individuals around the world dedicated to the design, growth, evolution, and operation of the global Internet and private networks built on the Internet Protocol. The following individuals have provided support to IPJ. You can join them by visiting <http://tinyurl.com/IPJ-donate>

| | | | | |
|----------------------------|------------------------|--------------------------|-------------------------|-------------------------|
| Kjetil Aas | Ilia Bromberg | Freek Dijkstra | Radu Cristian Gheorghiu | Nils Johansson |
| Fabrizio Accatino | Lukasz Bromirski | Geert Van Dijk | Greg Giessow | Brian Johnson |
| Michael Achola | Václav Brožík | David Dillow | John Gilbert | Curtis Johnson |
| Martin Adkins | Christophe Brun | Richard Dodsworth | Serge Van Ginderachter | Don Johnson |
| Melchior Aelmans | Gareth Bryan | Ernesto Doelling | Greg Goddard | Richard Johnson |
| Christopher Affleck | Ron Buchalski | Michael Dolan | Tiago Goncalves | Jim Johnston |
| Scott Aitken | Paul Buchanan | Eugene Doroniuk | Ron Goodheart | Jose Enrique Diaz Jolly |
| Jacobus Akkerhuis | Stefan Buckmann | Michael Dragone | Octavio Alfageme | Jonatan Jonasson |
| Antonio Cuñat Alario | Caner Budakoglu | Joshua Dreier | Gorostiaga | Daniel Jones |
| William Allaire | Darrell Budic | Lutz Drink | Barry Greene | Gary Jones |
| Nicola Altan | BugWorks | Aaron Dudek | Jeffrey Greene | Jerry Jones |
| Shane Amante | Scott Burleigh | Dmitriy Dudko | Richard Gregor | Michael Jones |
| Marcelo do Amaral | Chad Burnham | Andrew Dul | Martijn Groenleer | Amar Joshi |
| Matteo D'Ambrosio | Randy Bush | Joan Marc Riera | Geert Jan de Groot | Javier Juan |
| Selva Anandavel | Colin Butcher | Duocastella | Ólafur Guðmundsson | David Jump |
| Jens Andersson | Jon Harald Bøvre | Pedro Duque | Christopher Guemez | Anders Marius Jørgensen |
| Danish Ansari | Olivier Cahagne | Holger Durer | Rafael Leon Guerrero | Merike Kao |
| Finn Arildsen | Antoine Camerllo | Karlheinz Dölger | Gulf Coast Shots | Andrew Kaiser |
| Tim Armstrong | Tracy Camp | Mark Eanes | Galen Guyer | Vladislav Kalinovsky |
| Richard Artes | Brian Candler | Andrew Edwards | Sheryll de Guzman | Naoki Kambe |
| Michael Aschwanden | Fabio Caneparo | Peter Robert Egli | Rex Hale | Akbar Kara |
| David Atkins | Roberto Canonico | George Ehlers | Jason Hall | Christos Karayiannis |
| Jac Backus | David Cardwell | Peter Eisses | James Hamilton | Daniel Karrenberg |
| Jaime Badua | Richard Carrara | Torbjörn Eklöv | Darow Han | David Kekar |
| Bent Bagger | John Cavanaugh | Jacobus Gerrit Elsenaar | Handy Networks LLC | Stuart Kendrick |
| Eric Baker | Lj Cemerias | Y Ertur | Stephen Hanna | Robert Kent |
| Fred Baker† | Dave Chapman | ERNW GmbH | Martin Hannigan | Thomas Kernen |
| Santosh Balagopalan | Stefanos Charchalakos | ESdatCo | John Hardin | Jithin Kesavan |
| William Baltas | Molly Cheam | Steve Esquivel | David Harper | Jubal Kessler |
| David Bandinelli | Christof Chen | Jay Etchings | Edward Hauser | Shan Ali Khan |
| A C Barber | Pierluigi Checchi | Mikhail Evstiounin | David Hauweele | Nabeel Khatri |
| Benjamin Barkin-Wilkins | Greg Chisholm | Bill Fenner | Marilyn Hay | Dae Young Kim |
| Ryan Barnes | David Chosrova | Paul Ferguson | Headcrafts SRLS | William W. H. Kimandu |
| Feras Batainah | Marcin Cieslak | Ricardo Ferreira | Hidde van der Heide | John King |
| Michael Bazarewsky | Lauris Cikovskis | Kent Fichtner | Johan Helsingius | Russell Kirk |
| Robert Beckett | Brad Clark | Ulrich N Fierz | Robert Hinden | Gary Klesk |
| David Belson | Narelle Clark | Armin Fisslthaler | Michael Hippert | Anthony Klopp |
| Richard Bennett | Horst Clausen | Michael Fiumano | Damien Holloway | Henry Kluge |
| Matthew Best | James Cliver | The Flirble Organisation | Alain Van Hoof | Michael Kluk |
| Hidde Beumer | Guido Coenders | Jean-Pierre Forcioli | Edward Hotard | Andrew Koch |
| Pier Paolo Biagi | Robert Collet | Gary Ford | Bill Huber | Ia Kochiashvili |
| Arturo Bianchi | Joseph Connolly | Susan Forney† | Hagen Hultzs | Carsten Koempe |
| John Bigrow | Steve Corbató | Christopher Forsyth | Kauto Huopio | Richard Koene |
| Orvar Ari Bjarnason | Brian Courtney | Andrew Fox | Asbjørn Højmark | Alexader Kogan |
| Tyson Blanchard | Beth and Steve Crocker | Craig Fox | Kevin Iddles | Matthijs Koot |
| Axel Boeger | Dave Crocker | Fausto Franceschini | Mika Ilvesmaki | Antonin Kral |
| Keith Bogart | Kevin Croes | Erik Fredriksson | Karsten Iwen | Robert Krejčí |
| Mirko Bonadei | John Curran | Valerie Fronczak | Joseph Jackson | John Kristoff |
| Roberto Bonalumi | Sergio Danelli | Tomislav Futivic | David Jaffe | Terje Krogdahl |
| Lolke Boonstra | André Danthine† | Laurence Gagliani | Ashford Jaggernauth | Bobby Krupczak |
| Cente Cornelis Boot | Morgan Davis | Edward Gallagher | Thomas Jalkanen | Murray Kucherawy |
| Julie Bottorff Photography | Jeff Day | Andrew Gallo | Jozef Janitor | Warren Kumari |
| Gerry Boudreaux | Nicholas Dean | Chris Gamboni | Martijn Jansen | George Kuo |
| Leen de Braal | Fernando Saldana | Xosé Bravo Garcia | John Jarvis | Dirk Kurfuerst |
| Stephen Bradley | Del Castillo | Osvaldo Gazzaniga | Dennis Jennings | Mathias Körber |
| Kevin Breit | Rodolfo Delgado-Bueno | Kevin Gee | Edward Jennings | Darrell Lack |
| Thomas Bridge | Julien Dhallenne | Rodney Gehrke | Aart Jochem | Andrew Lamb |

| | | | | |
|---------------------------|---------------------------|-----------------------|--------------------------|-------------------------|
| Richard Lamb | Bart Jan Menkveld | Derrell Piper | Timothy Schwab | Sandro Tumini |
| Yan Landriault | Sean Mentzer | Rob Pirnie | Roger Schwartz | Angelo Turetta |
| Edwin Lang | Eduard Metz | Jorge Ivan Pincay | SeenThere | Brian William Turnbow |
| Sig Lange | William Mills | Ponce | Scott Seifel | Michael Turzanski |
| Markus Langenmair | David Millsom | Marc Vives Piza | Paul Selkirk | Phil Tweedie |
| Fred Langham | Desiree Miloshevic | Victoria Poncini | Andre Serralheiro | Steve Ulrich |
| Tracy LaQuey Parker | Joost van der Minnen | Blahoslav Popela | Yury Shefer | Unitek Engineering AG |
| Christian de Larrinaga | Thomas Mino | Andrew Potter | Yaron Sheffer | John Urbanek |
| Alex Latzko | Rob Minshall | Ian Potts | Doron Shikmoni | Martin Urwaleck |
| Jose Antonio Lazaro | Wijnand Modderman- | Eduard Llull Pou | Tj Shumway | Bart Vanautgaerden |
| Lazaro | Lenstra | Tim Pozar | Jeffrey Sicuranza | Betsy Vanderpool |
| Antonio Leding | Mohammad Moghaddas | David Preston | Thorsten Sideboard | Surendran Vangadasalam |
| Rick van Leeuwen | Charles Monson | David Raistrick | Greipur Sigurdsson | Ramnath Vasudha |
| Simon Leinen | Andrea Montefusco | Priyan R Rajeevan | Fillipe Cajaiba da Silva | Randy Veasley |
| Anton van der Leun | Fernando Montenegro | Balaji Rajendran | Andrew Simmons | Philip Venables |
| Robert Lewis | Roberto Montoya | Paul Rathbone | Pradeep Singh | Buddy Venne |
| Christian Liberale | Joel Moore | William Rawlings | Henry Sinnreich | Alejandro Vennera |
| Martin Lillepuu | Joseph Moran | Mujtiba Raza Rizvi | Geoff Sisson | Luca Ventura |
| Roger Lindholm | John More | Bill Reid | John Sisson | Scott Vermillion |
| Link Light Networks | Maurizio Moroni | Petr Rejhon | Helge Skrivervik | Tom Vest |
| Art de Llanos | Brian Mort | Robert Remenyi | Terry Slattery | Peter Villemoes |
| Mike Lochocki | Soenke Mumm | Rodrigo Ribeiro | Darren Sleeth | Vista Global Coaching & |
| Chris and Janet Lonvick | Tariq Mustafa | Glenn Ricart | Richard Smit | Consulting |
| Mario Lopez | Stuart Nadin | Justin Richards | Bob Smith | Dario Vitali |
| Sergio Loreti | Michel Nakhla | Rafael Riera | Courtney Smith | Marc Vives |
| Eric Louie | Mazdak Rajabi Nasab | Mark Risinger | Eric Smith | Rüdiger Volk |
| Adam Loveless | Krishna Natarajan | Fernando Robayo | Mark Smith | Jeffrey Wagner |
| Josh Lowe | Naveen Nathan | Michael Roberts | Tim Sneddon | Don Wahl |
| Guillermo a Loyola | Ryan Nelson | Gregory Robinson | Craig Snell | Michael L Wahrman |
| Hannes Lubich | Darryl Newman | Ron Rockrohr | Job Snijders | Lakhinder Walia |
| Dan Lynch [†] | Mai Nguyen | Graziano G Rodegari | Ronald Solano | Laurence Walker |
| David MacDuffie | Thomas Nikolajsen | Carlos Rodrigues | Asit Som | Randy Watts |
| Sanya Madan | Paul Nikolich | Magnus Romedahl | Ignacio Soto Campos | Andrew Webster |
| Miroslav Madić | Travis Northrup | Lex Van Roon | Evandro Sousa | Jd Wegner |
| Alexis Madriz | Marijana Novakovic | Marshall Rose | Peter Spekrijse | Tim Weil |
| Carl Malamud | David Oates | Alessandra Rosi | Thayumanavan Sridhar | Westmoreland |
| Jonathan Maldonado | Ovidiu Obersterescu | David Ross | Paul Stancik | Engineering Inc. |
| Michael Malik | Jim Oplotnik | William Ross | Ralf Stempfner | Rick Wesson |
| Tarmo Mamers | Tim O'Brien | Boudhayan | Matthew Stenberg | Peter Whimp |
| Yogesh Mangar | Mike O'Connor | Roychowdhury | Martin Štěpánek | Russ White |
| John Mann | Mike O'Dell | Carlos Rubio | Adrian Stevens | Jurrien Wijlhuizen |
| Bill Manning [†] | John O'Neill | Rainer Rudigier | Clinton Stevens | Joseph Williams |
| Diego Mansilla | Carl Ötne | Timo Ruiters | John Streck | Derick Winkworth |
| Harold March | Packet Consulting Limited | RustedMusic | Martin Streule | Pindar Wong |
| Vincent Marchand | Carlos Astor Araujo | Babak Saberi | David Strom | Brian Woods |
| Normando Marcolongo | Palmeira | George Sadowsky | Colin Strutt | Makarand Yerawadekar |
| Gabriel Marroquin | Gordon Palmer | Scott Sandefur | Viktor Sudakov | Phillip Yialeloglou |
| David Martin | Alexis Panagopoulos | Sachin Sapkal | Kathleen Summers | Janko Zavernik |
| Jim Martin | Gaurav Panwar | Arturas Satkovskis | Edward-W. Suor | Bernd Zeimetz |
| Ruben Tripiana Martin | Chris Parker | PS Saunders | Vincent Surillo | Muhammad Ziad |
| Timothy Martin | Alex Parkinson | Richard Savoy | Terence Charles Sweetser | Ziayuddin |
| Carles Mateu | Craig Partridge | John Sayer | T2Group | Tom Zingale |
| Juan Jose Marin Martinez | Manuel Uruena Pascual | Phil Scarr | Roman Tarasov | Matteo Zovi |
| Ioan Maxim | Ricardo Patara | Gianpaolo Scassellati | David Theese | Jose Zumalave |
| David Mazel | Dipesh Patel | Elizabeth Scheid | Rabbi Rob and | Romeo Zwart |
| Miles McCredie | Dan Paynter | Jeroen Van Ingen | Lauren Thomas | 廖明沂. |
| Gavin McCullagh | Leif-Eric Pedersen | Schenau | Douglas Thompson | |
| Brian McCullough | Rui Sao Pedro | Carsten Scherb | Kerry Thompson | |
| Joe McEachern | Juan Pena | Ernest Schirmer | Lorin J Thompson | |
| Alexander McKenzie | Luis Javier Perez | Benson Schliesser | Jerome Tissieres | |
| Jay McMaster | Chris Perkins | Philip Schneek | Fabrizio Tivano | |
| Mark Mc Nicholas | Michael Petry | James Schneider | Peter Tomsu Fine Art | |
| Olaf Mehlberg | Alexander Peuchert | Peter Schoo | Photography | |
| Carsten Melberg | David Phelan | Dan Schrenk | Joseph Toste | |
| Kevin Menezes | Harald Pilz | Richard Schultz | Rey Tucker | |

Call for Papers

The *Internet Protocol Journal* (IPJ) is a quarterly technical publication containing tutorial articles (“What is...?”) as well as implementation/operation articles (“How to...”). The journal provides articles about all aspects of Internet technology. IPJ is not intended to promote any specific products or services, but rather is intended to serve as an informational and educational resource for engineering professionals involved in the design, development, and operation of public and private internets and intranets. In addition to feature-length articles, IPJ contains technical updates, book reviews, announcements, opinion columns, and letters to the Editor. Topics include but are not limited to:

- Access and infrastructure technologies such as: Wi-Fi, Gigabit Ethernet, SONET, xDSL, cable, fiber optics, satellite, and mobile wireless.
- Transport and interconnection functions such as: switching, routing, tunneling, protocol transition, multicast, and performance.
- Network management, administration, and security issues, including: authentication, privacy, encryption, monitoring, firewalls, troubleshooting, and mapping.
- Value-added systems and services such as: Virtual Private Networks, resource location, caching, client/server systems, distributed systems, cloud computing, and quality of service.
- Application and end-user issues such as: E-mail, Web authoring, server technologies and systems, electronic commerce, and application management.
- Legal, policy, regulatory and governance topics such as: copyright, content control, content liability, settlement charges, resource allocation, and trademark disputes in the context of internetworking.

IPJ will pay a stipend of US\$1000 for published, feature-length articles. For further information regarding article submissions, please contact Ole J. Jacobsen, Editor and Publisher. Ole can be reached at ole@protocoljournal.org or olejacobsen@me.com

The Internet Protocol Journal is published under the “CC BY-NC-ND” Creative Commons Licence. Quotation with attribution encouraged.

This publication is distributed on an “as-is” basis, without warranty of any kind either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement. This publication could contain technical inaccuracies or typographical errors. Later issues may modify or update information provided in this issue. Neither the publisher nor any contributor shall have any liability to any person for any loss or damage caused directly or indirectly by the information contained herein.

Follow us on X and Facebook

 @protocoljournal

 <https://www.facebook.com/newipj>

Supporters and Sponsors

Supporters



Internet
Society



Diamond Sponsors

Your logo here!

Ruby Sponsors



Sapphire Sponsors



Emerald Sponsors



Corporate Subscriptions



For more information about sponsorship, please contact sponsor@protocoljournal.org

The Internet Protocol Journal
Link Fulfillment
7650 Marathon Dr., Suite E
Livermore, CA 94550

CHANGE SERVICE REQUESTED

The Internet Protocol Journal

Ole J. Jacobsen, Editor and Publisher

Editorial Advisory Board

Dr. Vint Cerf, VP and Chief Internet Evangelist
Google Inc, USA

John Crain, Senior Vice President and Chief Technology Officer
Internet Corporation for Assigned Names and Numbers

Dr. Steve Crocker, CEO and Co-Founder
Shinkuro, Inc.

Dr. Jon Crowcroft, Marconi Professor of Communications Systems
University of Cambridge, England

Geoff Huston, Chief Scientist
Asia Pacific Network Information Centre, Australia

Dr. Cullen Jennings, Cisco Fellow
Cisco Systems, Inc.

Merike Kaeo, Founder and vCISO
Double Shot Security

Olaf Kolkman, Principal – Internet Technology, Policy, and Advocacy
The Internet Society

Dr. Jun Murai, Founder, WIDE Project
Distinguished Professor, Keio University
Co-Director, Keio University Cyber Civilization Research Center, Japan

The Internet Protocol Journal is published quarterly and supported by the Internet Society and other organizations and individuals around the world dedicated to the design, growth, evolution, and operation of the global Internet and private networks built on the Internet Protocol.

Email: ipj@protocoljournal.org
Web: www.protocoljournal.org

The title "The Internet Protocol Journal" is a trademark of Cisco Systems, Inc. and/or its affiliates ("Cisco"), used under license. All other trademarks mentioned in this document or website are the property of their respective owners.

Printed in the USA on recycled paper.

