The ISP Column
A column on things Internet

September 2022
Geoff Huston
APNIC

Fragmentation

One of the discussion topics at the recent ICANN 75 meeting was an old favourite of mine, namely the topic of *Internet Fragmentation*. Here, I'd like to explore this topic in a little more detail and look behind the kneejerk response of declaiming fragmentation as bad under any and all circumstances. Perhaps there are more subtleties in this topic than simple judgements of good or bad.

Let's rewind back to the 1980s for a second and look at the environment in the computer industry at the time. This was an environment of mainframe computers and connected peripheral devices. Each computer vendor approached the market with a unique product set of computers and associated peripheral devices, such as data entry devices, printers and similar, that did not necessarily interoperate with those from any other vendor. If you had purchased your computing solution from IBM, or Univac, or from the Digital Equipment Corporation, then the one outcome that was assured was that as a customer you were essentially locked into that vendor's equipment set. Switching vendors was a highly expensive process for many customers and literally everything had to be replaced and vendors appreciated this form of customer lock-in as much as customers did not. Unsurprisingly, customers were losing patience with this situation at the time, and there was conscious effort to force the industry to build computing devices that were based on open standards such that they could interoperate with products from other vendors. In this way customers had some level of assurance that a multi-vendor computer environment was feasible, and that the various devices would interoperate if not seamlessly then at least in a functionally adequate manner.

There were no regulations that enforced such outcomes, but the existence of these industry standards assisted customers to state their desire to procure standards-based products and to guide providers as to the required attributes of their products. The adoption of these standards was a market-based movement where producers and consumers were both motivated to abide by these standards as an act of self-interest. The Internet is an outcome of this same process of voluntary adoption of common technology standards. There are no rules here and there is no enforcement mechanism.

The Internet is the result of a market-based discipline that reflects a consumer preference which itself guides the actions of producers of digital goods and services. Consumers may not necessarily want to avail themselves of every possible service all the time, but the aggregate of these consumer choices is for every service, and every service provider wants to have their service be accessible by every consumer. A *coherent* networking environment exhibits the same behaviours all the time, with consistent access to all servers and displaying the same service outcomes for all consumers. Both producers and consumers can assume universal access capabilities

In contrast, a *fragmented* environment is one where services may not be generally accessible for all. For example, you may not be able to access my web site or may not be able to send me email. A fragmented network may also exhibit collisions, where different consumers may have different outcomes from the same access operation. For example, when I enter the URL for my web site into my browser I see my expected content, yet when you enter precisely the same URL into your browser you are directed to an entirely different site and an entirely different service. Perhaps this is so obvious that it does not need stating, but such fragmentation outcomes are highly disruptive to the utility and value of the network. In many cases such impediments to communication across the network may be temporary and unintended, but when we refer to *fragmentation,* we typically refer to intentional disruptions of service coherence.

Fragmentation comes in many shapes and forms. Many national regimes require their domestic Internet access service providers to maintain network-level blocks to prevent the users from accessing proscribed content or services. Such regulatory blocks may also apply to content publishers and related forms of content distribution networks. While basic consumer protection and more generic cyber-security concerns for consumers are a common factor behind such blocks they are not the only motivations for nation state involvement in directing network fragmentation There are national strategic concerns about foreign dependence for critical domestic network infrastructure as in today's world, with its heightened level of geopolitical tension and its undercurrents of cyber hostility, the Internet has been pushed into a prominent place in such national strategic considerations. The outcomes of such national interventionist actions call into question the basic assumptions of openness, network-wide coherence, and global accessibility in this networked environment.

Such national efforts can go further than blocking of certain sites, content, or resisting the use of foreign-operated infrastructure. Would any nation state go all the way and pull the plug on the Internet? Well, yes! Many nations have tried this approach in the past and may well as well do so again. So self-isolation measures have been used as a means of imposing controls over a restive domestic population in times of civil unrest. Such deliberate self-isolation measures become a means of suppressing the ability to communicate within the national domestic population

There are fragmentation responses to cyber security concerns. Why should a DNS resolver willingly resolve the DNS names of a web site whose only function is to infect end hosts with various forms of malware? Some service providers have made this form of deliberately selective response into a competitive differentiator, such as the QUAD9's open recursive resolver where malware information feeds are used to inform a resolution block list for the DNS service, on the basis that using the services provided by these sites exposes the user to malware, viruses, and other forms of harm. Is such intentional fragmentation, done in the name of protecting the user, a good or bad thing in the larger context of universal accessibility?

This debate has been one that has been going on for decades. The original block lists of IP addresses were a response to the rising torrents of spam. If a network service provider could simply deny all network access to those remote servers that were a consistent source of spam, then they're making the Internet better for users. Right? These block lists were operated by enthusiastic volunteers and the conditions for being entered on a block list, and subsequently being removed were often opaque and arbitrary. Its challenging to ascertain if such fragmentation responses were effective. The torrents of spam appear to continue in an unabated manner and if fragmentation responses were intended to make email abuse harder, then it is hard to point to clear evidence that this has been the case.

There are also geopolitical motivations lurking behind this conversation about Internet fragmentation. Should a nation state be obliged to open up its network to allow domestic consumers unconstrained access to the services provided by the largely US-based clique of big tech enterprises? As we seen in the decades of free trade discussions in the WTO this is a highly contentious topic, both in the traditional realm of global free trade in physical goods and in the digital trade of online goods and services. Is a professed reluctance to adopt rule-based mechanisms of unconstrained universal access just a pretext that allows the incumbent large scale service providers some level of political support to resist others entering the market and competing with these incumbent entities at both a corporate and a national level. Is the debate about the perils of fragmentation really a debate about the responses by some nation states to try and resist further escalation of digital exploitation by foreign entities?

This discussion emerges in many contexts in the Internet environment. It can be found in Internet Governance Forums, in internal conferences on global communications, and in global trade negotiations, to name a few. Today, Internet fragmentation, or constraints applied to Internet accessibility, is as relevant to the global economy as constraints on the trade of physical goods. The conversation about fragmentation also surfaces in forums that are focused on common Internet infrastructure, notably in the administration on the Internet's naming and addressing schemes. In this case the fragmentation discussion is about defining norms and safeguards that apply to every service provider in operating this common infrastructure, to enable uses to experience a single seamless Internet on top of this diverse collection of individual component services. The assignation of a DNS name or an IP address prefix to be used by an entity is intentionally an Internet-wide unique assignation. Fragmentation in this context of common infrastructure directly confronts the assumptions about uniqueness and global scope.

Let's sharpen the focus and further look at this topic of fragmentation through the lens of the Internet's naming and addressing infrastructure.

Is the Internet's address space already fragmented?

It sounds like such an easy question. And of course, the answer is not straightforward.

To start with we can observe that with IPV4 the address space contains some 4.3 billion unique 32-bit addresses, but if we believe the various estimates of the size of the Internet there are some five billion human users of this network and perhaps some 30 to 40 billion devices of various forms that are connected to it. Clearly, not every user exclusively uses an access device that has a unique stable IPv4 address, and certainly not every device has such a form of connection. The technical answer to this over population lies in the widespread use of address sharing. The architectural answer lies in the adoption of a client/server architecture for the Internet, where clients communicate with servers and not with other non-local clients.

Is this use of IPv4 address space fragmented from the perspective of the original peer-to-peer Internet? Well, yes, it's a fractured and fragmented space.

There is, however, a corollary to this question. Does this matter to users? Here the answer is clearly no! The commercial imperatives and the security impurities of the Internet are served by this client/server architecture. As long as service delivery points have stable addresses, then to the client user it still looks like a coherent and connected network. And we continually experiment in network mechanisms where even stability of service delivery addresses is not a strict prerequisite.

How about IPv6? Does the adoption of IPv6 addressing solve the fragmentation that we see in the IPv4 space? Perhaps not.

Despite its massively larger total address space, not every client and not every service is available over IPv6. Today only one third of the Internet's user population are IPv6 capable. If universal coherent connectivity is the desired outcome, then the inability to reach some two thirds of users clearly represents a significant fracture in the assertion of coherent universal connectivity.

What if we constraint our question to refer to just the IPv6 Internet. Surely that's a coherent space? Not really. IPv6 addressing is not stable for end clients. Deliberately so. It doesn't take long to understand that that presenting the same client IPv6 interface identifier value, even across changes to the device's external connectivity, such as from fixed to mobile access, becomes a significant privacy leak. These days most IPv6 clients use privacy addresses, deliberately obfuscating the low order 64-bit interface identifier in order to counter such tracking. There is also the observation that IPv6 does not necessarily operate as a fully interconnected subnet. You may have an IPv6-enable device on an IPv6-capable access network and I might be similarly connected. But you may not be able to send me an IPv6 packet. We might both the regarded as clients and there may be a firewall on the client-to-client path that rejects incoming connection requests over TCP. Or you might be routed using a highly specific /64 route and my access network might drop any prefix more specific address prefix that is small than a /48, for example. Or the IPv6 network simply does not connect our respective IPv6-capable access networks over an IPv6 path.

The IPv6 network is fragmented from an address perspective. But that does not necessarily compromise the ability of clients to access services on the network, at least not a present. The reason is that the dual stack nature of the intended IPv6 transition plan means that, for the present at any rate, any breakage in connectivity between clients and servers in IPv6 is seamlessly pathed up using IPv4. The service environment has managed to seamlessly bridge around such problems and largely ignore them.

How has the service environment managed to respond to the fragmentation in the address space? By relying on a coherent and unfragmented name space. So if the question is the level of fragmentation in the infrastructure of the internet, then the focussed question is: How's the name space going with respect to fragmentation pressures?

Is the Internet's name space already fragmented?

The use of the definite article in that question is perhaps misleading. While the IP address is an intrinsic part of an IP packet, the name system is an application construct. It's applications that essentially define the common name space. A set of cooperating applications could certainly define their own space which was different to the namespace defined by the Domain Name System. We saw this in the fragmented network landscape of the 1980's, where each network realm of connectivity used its own name space, and it's possible that this will happen again.

While it is not an intrinsic part of base Internet technology, we have all agreed to use a single namespace and a single name resolution mechanism as a matter of common convention. The reason for this convention is simple. Users want to use a communications network to pass references to each other about accessible resources. For this to happen we need a common name space. Without such commonality in our network's name space, we simply cannot communicate. Such name coherence within the Internet is in everybody's interest and market disciplines become the most effective way to ensure that the name space remains unified and cohesive. From this perspective defining a rule set and enforcement mechanisms for an unfragmented name space are unnecessary.

If it's in everybody's interest to work within the context of a single interoperable framework within the Internet, then we are we looking at fragmentation? What could possibly motivate an actor to break away from this common environment?

An answer to this question lies in the efforts to achieve competitive differentiation. How do service providers differentiate their service to obtain market share? Price alone often creates perverse incentives for centralisation, particularly when the service is susceptible to economies of scale. In such an environment, new entrants to the market often need to differentiate themselves from the incumbents by using factors other than just the price of the service. Such competitive efforts may result in a level of deviation from the common norms of the technology and the service profile and these behaviours man not easily interoperate with the installed base. In other words, fragmentation may be a deliberate action by a new market entrant to differentiate themselves from the services provided by the current incumbents.

In other cases, it's a reaction to the implicit barriers to participation being set by the incumbents. With reference to domain names and access to the so-called "top level" name space an applicant either needs to be a recognised country or be willing to pay the not inconsiderable costs associated with an application for a generic top-level name, and also be willing to wait for an indeterminate amount of time. If these preconditions are not to your liking, then the alternative is to head into the world of name fragmentation and use a name drawn from an alternative name framework, whether its unstoppable names, TOR, Etherium names or the GNU Name system. Names drawn from these alternative frameworks may be more accessible but suffer from the issue that few users run applications that recognise these alternative name frameworks, and more accessible names must be traded off against vastly more limited utility through limited adoption.

Why would any service provider use a name drawn from one of these fragmented spaces? Surely its limited adoption would be a powerful deterrent? The answer lies in the marginal cost for the name applicant. As long as the service provider already has a name and the marginal cost of registering the same or a similar name in this alternative space is extremely low for the provider, then there is no downside in participating in this alternate name space. The hope for the alternative name provider here us that as more entities make a similar decision the alternative name space gathers usage momentum and if that continues then at some point it would gain sufficient market presence to cross the line from fragmented challenger to mainstream incumbent.

However, this is not an easy outcome to achieve. The incumbent name space, the Domain Name System (DNS) never really assume that it was one of many potential name spaces. The DNS carries no distinguishing label with it that says "use the DNS resolution protocol with the IANA root as the seed of the query to resolve this name" There is no "super" top level domain that maps sets of domain names to one resolution realm or another. There is not even an ability to uniquely assign a given DNS name into one resolution space or another, and the likelihood of collisions where the same name can be resolved in multiple ways with multiple different outcomes is very high.

We've had to go to some extraordinary lengths to resist fragmentation in some cases. The efforts to pull Unicode into the name system to support non-Latin scripts in the name space has been a challenging exercise. The Unicode character set is everything ASCII is not. Its ambiguous, in that there are multiple ways to encode the same visual appearance, it's not predictable, in that the same glyph can be presented in multiple ways, it does not have a clear upper/lower case distinction and the DNS itself is not 8-bit clean. The choices were to adopt a new DNS environment for each script collection, or somehow fold Unicode into the constrained ASCII character set. The first solution was clearly going to present fragmentation issues for the name infrastructure, while the latter solution, mapping Unicode into an ASCII representation, enrols application behaviour into the DNS, something the DNS had tried hard to avoid until then. The "solution" of IDNs is far from ideal. It adds various vulnerabilities into the user presentation of the application environment., by virtue of permitting visually identical name strings to be represented by distinct Unicode glyph sequences which in turn encode to distinct domain names. But the main factor in favour of this approach is that is keeps the DNS as a cohesive and unfragmented space.

So far, the effort to resist fragmentation of the name space has been largely successful. So far, we've been able to operate the Internet in a mainstream of the name space that provides the essential attributes of universality and cohesion. But there is a cost of this outcome. The incumbents in the provision of name infrastructure are few in number and large in size, and the longer-term trend is fewer and larger. The price of this outcome is centrality. The price is an ever-increasing level of dominance by incumbents and an increased level of resistance to any and all forms of evolutionary change. Obviously, there are no clear answers as to which is the preferred outcome here.

Disclaimer

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

Author

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

www.potaroo.net