December 2022
Geoff Huston

Is Secured Routing a Market Failure?

The Internet represents a threshold moment for the communications realm in many ways. It altered the immediate end client of the network service from humans to computers. It changed the communications model from synchronized end-to-end service to asynchronous, and from virtual circuits to packet switching. At the same time, there were a set of sweeping changes in the public communications framework, changing the general characterization of the set of service operators in this space from state-operated monopolies, or if not state-operated then at a minimum state-sanctioned, to enterprise actors operating within a largely deregulated market economy.

These were very fundamental changes. In particular, the implication of the deregulation of this activity was that market disciplines become the arbiter of change rather than regulatory fiat. Change occurs when the service operators who adopt such changes achieve greater returns from their invested capital, or, at a minimum offer the enticing prospect of such greater returns. In a market-based activity, changes do not generally occur when there is no market-based incentive for actors to undertake such changes.

The protracted transition to IPv6 is an interesting case in point. The intent of the IPv6 design was to change as little as possible from the design of IPv4. At its best the transition to IPv6 was intentionally invisible, in that applications and users of such applications could operate seamlessly over IPv4 or IPv6 to the point that applications and users need not even be aware which IP protocol was being used for any particular session.

The result was that there was little in the way of natural market incentives to drive the adoption of IPv6 other than the imminent exhaustion of the so-called *free pools* of IPv4 address space. For as long as existing service providers hesitated to adopt IPv6, presumably because they had an adequate supply of IPv4 addresses for their immediate operational requirements, new market entrants are forced to secure adequate supplies of IPv4 addresses. The increased costs associated with purchasing such IPv4 addresses on the open market formed an added barrier to market entry. In addition, the immediate adoption of IPv6 by incumbents provided no direct business or competitive advantage, as such a move does not improve revenue margins, nor reduce operating costs. Without a natural market-based incentive to integrate IPv6 into their service offering, the adoption of IPv6 became a protracted exercise.

Using a similar lens of market dynamics and looking at the level to which market-based incentives exist for actors, is the adoption of routing security heading in the same direction of market failure?

Why do this at all?

Firstly, let's understand why routing security has been perceived to be important.

Since the very first days of the Internet, we have worked on the principle that if you always send a packet to the "right" IP address you can always believe the response. For services that do not require any further client or server authentication, such as the DNS or the original port 80 web services, then this is a critical assumption. Most of the time that assumption is a reasonable one. But sometimes such a level of trust in the integrity of the network to deliver a packet to its intended destination is just naïve.

Packets can be misdirected. Packets can be intercepted. Responses can be synthesized. If we are talking about the operation of the DNS, for example, then the result is that a user can be unwittingly directed to a fake service that steals the user's credentials. We've seen such forms of interception attacks used by criminals to steal money from users and used by state-based actors to inspect their citizens' email archives, and of course many other forms of abuse.

Such misdirection of packets is achieved by the injection of false reachability into the routing system, or a routing attack. The elegance of this form of attack is that nothing on the user's system needs to be attacked or altered. The problem is not with the end system or its suite of applications. The system is working exactly as intended. The problem lies within the network. The problem is that it's the network's routing system that is the gullible component. It's the routing system that cannot tell the difference between a routing update containing authentic information and a faked one.

The combination of a naïvely trusting routing system and naïvely trusting applications is very discomforting. If we had been able to prevent the corruption of routing information, and better yet, able to ensure that packets addressed to an IP destination address reached the device that is attached to the network with that IP address, then we would've been able to eliminate a large class of vulnerabilities in the networked space. If you operate a network then attempting to reduce the naïvety of trust within the routing system would appear to be a natural response, and this has been an active topic for network operators and their various service providers for more than three decades.

Efforts to Secure Routing

One of the early efforts in this space was *routing registries*. The concept is that each network describes what it intends to announce into the routing system, and any associated policies it will apply to routing information as entries in a routing registry so that anyone else can use this information to construct a filter. If I inadvertently leak routes, then anyone else that is using these filters will catch and discard these unintended announcements. If every network operator used this routing registry, then I could audit the entries and look for my address prefixes to ensure that all registered claims associated with these prefixes matched the permissions that I have given to networks to announce my routes.

The concept of a single routing registry used by all operators that contain the complete set of anticipated route advertisements has always proved to be an elusive goal. The number of distinct routing registries quickly proliferated, as many major Tier 1 transit providers adopted their own routing registry. Not only did the number of registries quickly expand, but it also happened without an eye to mutual consistency. Different routing information was loaded into different routing registries.

The registries were also overloaded with contingency routing information. If a provider had made backup arrangements in the event that its primary transit paths had failed, then these backup routes were also described in the routing registry. And unless the routing registry contents were assiduously maintained, which they generally were not, historical information jostled with current information. Network operators who were using this information to construct filters to apply to routing information were facing an extremely challenging task. The filter construction process had to somehow resolve contradictions if two or more routing registries were used as inputs, as well as cope with extremely large filter constructs.

Describing routing policies in this environment is also extremely challenging. The language to describe such policies (the Routing Policy Specification Language, or RPSL) never really gained traction in the network operations community, so the use of routing registries was often confined to simple route origination constructs.

There were two further issues with the routing registry model that also impacted on its acceptance. The first is that the authority models were missing information submission constraints that would enable the registry to only accept registry entries relating to the routing of address prefixes and Autonomous Systems (ASes) that the submitter had direct control over. This was not limited to the lack of entry control, but also had no mechanism to audit existing routing registry content to ensure that existing entries were still

compliant with such a relationship of direct control. If IP addresses or ASes changed hands, then the registry needed to reflect that change of control and discard the historical information. The implication of this missing access control was that the authenticity of the registry data was questionable. This lead directly to the second issue. The data in the registry was not protected with any digital signature that conveyed an authority to make assertions relating to routing and routing policy. If the problem was unverifiable trust in the routing system, a response that featured unverifiable trust in the routing registries was hardly progress. We now had two problems.

Perhaps in spite of these shortcomings, routing registries enjoyed some level of success, but only in some locales and some communities. As an Internet-wide tool it was unable to gain a critical mass of adoption, even though it was the only available routing security mechanism for many years.

In the early 2000's we returned to this issue and looked directly at the routing protocol itself. In this case we are looking at the Border Gateway Protocol (BGP), and the question was re-phrased in the context of the operation of this protocol. How could a BGP speaker assure itself that the routing information presented to it was authentic? This can be broken into two parts: the injection of the *origination* of the address prefix into the routing system and the integrity of *propagation* of this routing information through the routing system. To achieve this assurance, we used a conventional approach of public key cryptography and an associated inter-locking key framework, a *public key infrastructure* (PKI).

There were, however, some issues raised by these choices.

The router vendors in the IETF had taken a hostile view on the topic of the IETF making further changes to the BGP protocol, as this represented for these vendors incremental cost for standards compliance without the opportunity for additional revenue from their customers. As some of these vendors had a highly influential position in the IETF, these concerns were expressed as an overriding constraint placed on the technical specification of a secured routing framework as part of the IETF's development work. The result was that this IETF work in securing the BGP protocol was unable to propose changes to the BGP protocol. To claim that this was a short-sighted and self-interested perspective on the part of the router vendors (Cisco and Juniper) who, at the time, dominated the IETF decision making in the area of routing that turned this work from something that had a prospect of being viable into an assured failure is perhaps understating the longer-term ramifications of this adopted position.

The choice of a PKI was not without its implications, as the applicable standards in this space are based on X.509 public key certificates. X.509 certificates are not well suited to the routing space. Each BGP speaker holds the entirety of the routing information all the time. What this means is that we need to use these certificates in a different way. We don't want to use them selectively on a per-transaction basis. What we want is for every relying party in the system to hold the entirety of the issued certificate set all the time.

X.509 was not designed for such a use case and very quickly the design effort found itself inventing new credential distribution mechanisms to support this objective. Normally, if the objective is to allow a receiver of a routing update to validate the information contained in a routing update, the approach would be to attach the credentials to the routing update, but this implies a change to the BGP protocol, which was not a permitted option within the design criteria.

What was designed in place of direct support of information flooding within BGP was a mechanism of on-demand pull information flooding. Each client of the routing system, or every network, has to synchronize its local cached copy of the information published by every network, and do so within an unspecified time tolerance that is no slower than the time scale of BGP route propagation. If every network publishes its information using a locally operated publication point, then every network as a client has to poll the publication points of every network continuously. This is an N-squared scaling problem and with today's 75,000 individual networks that would imply a system local of 5.6B individual synchronisation operations. If the time scale of BGP propagation is of the order of tens of seconds, then that's a daily total query load of 48T such operations if synchronization uses a 10-second poll interval.

This system load could be reduced if we used aggregated publication points, which is the operational preference for the Resource Public Key Infrastructure (RPKI) credential dissemination framework, but with this structure we've created a small number of highly critical points of vulnerability where failure to access these servers can cause operational problems for the routing fabric of the entire network. This mechanism of on-demand pull for information flooding simply does not scale to the size of today's Internet if every network operator is active as both a client and a publisher of these X.509 credentials.

On the other hand, partial adoption has its own problems. The issue with partial adoption is that signing systems can sign what's valid but cannot sign what's invalid. When a client is presented with unsigned data, it cannot clearly determine whether it's unsigned because its falsified information that could never have a valid digital signature, or because it's authentic information that is just not signed. Systems that rely on positive attestations work when the client can assume that the entire realm of valid data is validly signed. In scenarios of partial adoption of signing the client is no better off.

There is the also the issue of integrity of the propagation of routing updates. This has been addressed in this framework by adding digital signatures to protect the AS Path attribute of a BGP update. In the BGPSEC protocol each eBGP speaker signs the update with the private key associated with the Autonomous System Number (ASN), signing across the already signed AS Path and the ASN of the network to which this update is being passed. The issue here is that this model does not accommodate piecemeal partial deployment.

For this form of protection to be viable, the domain of connection is limited to a fully connected subset of networks, and once the update is passed to a non-BGPSEC speaker the path protection credentials are no longer useful. Even if the update subsequently reaches a BGPSEC-aware network, the AS Path signature credentials cannot be restored as the interlocking signing chain is broken. If every network operator was running BGPSEC then any receiver of a routing update could use the AS Path signature structure to assure itself that the AS Path update accurately reflects the propagation of the update through the inter-AS network. However, in a scenario of piecemeal partial deployment, such assurances, even when they can be validated, are of limited use. Whenever an update is propagated across a non-BGPSEC network, all AS Path signatures for all updates are voided.

Without AS path protection, protection of origination is an easily circumvented security mechanism, and by itself probably does not even warrant a characterization as a *security mechanism*. At its best, the validation of route origination credentials when used by itself can be used to detect and filter a sub-class of *routing leaks* (inadvertent errors by a network operator that unintentionally announce routes into the inter-domain routing system that may, or may not, have an impact on packet forwarding and may cause instability in the routing system).

If the efficacy of this entire routing security framework hinges upon the universal adoption of an AS Path assurance mechanism, namely BGPSEC, then until such a state is reached the incremental benefits that accrue to both the individual adopting network and the routing system in general are marginal, if any, while the cost of integrating these mechanisms into the individual network's routing infrastructure are constant. In other words, to make this point in possibly overly blunt terms, the point in time at which there will be a useful benefit to the individual adopter is when the last network operator adopts this technology, not the first. This is the precise opposite of an *early adopter benefit*. This is a classic case of a *late adopter benefit* or *deferred benefit*.

There is a final more generic issue when reviewing efforts to address routing security, namely that *routing integrity* is not the same as *forwarding integrity*. Even if we are able to assure ourselves that we can align the information propagated within the routing protocol to the current state of the network in terms of topology and forwarding, this is still only part of a larger picture. The issue here is that even if the entirety of information being passed to routers in the network is authentic, what the router does with this information, how the router makes forwarding decisions for packets and how the packets actually reach

their intended destination by the cumulative results of the sequence of such individual forwarding decisions is beyond the scope of the routing system.

Even if the routing system is fully secured there is still ample opportunity to interfere with the network's packet forwarding operation by tampering with the forwarding behaviour on individual routers.

All these are not necessarily totally prohibitive issues, but they do impose additional operational burdens on those networks that choose to adopt this technology, as well as being a source of greater levels of risk. In other words, this is a source of additional cost for network operators, while at the same time, the extent of realizable benefit is very limited.

Cost and Benefit

If there is increased cost to be borne by network operators to operate a secured routing framework in their network, then is there a benefit and who accrues such benefit?

It could be argued that ultimately users benefit from a secured routing system, and therefore the cost should be directly accrued to users through increased fees, but the case to do so is very weak. Normally this would be a consumer preference. In a competitive market for goods and services sold to consumers, each consumer would have some form of choice and would express an individual preference as to service and cost through their purchasing decisions.

If a consumer wished to access a service that had integrated these mechanisms into their routing system, then then the consumer could express their preference for such a service through the purchasing decisions. However, routing security is a highly esoteric and obscure attribute of any network, and it is extremely improbable that a provider's use of this technology would impact a consumer decision. Even the integration of IPv6 into goods and services sold to consumers has apparently failed as a visible differentiator that would command a price premium and compared to IPv6 routing security functions are easily an order of magnitude more obscure and more challenging for many consumers to place into a meaningful context of incremental benefit and cost. If differentiating the offerings of services based on their capability to support IPv6 has proved challenging in the consumer market, then one could reasonably expect differentiation based on the network's use of routing security mechanisms to be an impossible proposition. A network provider that had integrated these routing security mechanisms into their network would finding it extremely challenging to present this attribute as an advantageous product differentiator. In other words, it's really hard to get consumers to pay the costs of adding routing security to the network.

But perhaps this view is overly simplistic. Individual users do not directly craft IP packets and pass then into the network and then try to determine their fate. We as human users of the network just don't do that. So why should we expect these same individuals to make purchasing decisions based on the subtle differences in these obscure behaviours? That's the role of applications. So, let's change the focus of working through the issues of cost and benefit and look at applications in the context of routing security.

It's a reasonable assumption that applications would directly benefit from a secured routing system. When the routing system directs traffic to the wrong destination then this opens up a set of consequent issues. There is the problem of *passing off* where a fake server is substituted in place of the intended server. This server can provide incorrect responses (such as a DNS server that provides incorrect answers) or steal the user's login credentials (such as a fake web server), or more insidious attacks then enable *data exfiltration*.

The issue here is that applications have been facing this issue since the inception of the Internet. They cannot assume that the network is operating with forwarding integrity, and they cannot assume that the traffic exchange is free from onlookers. Routing security cannot address the entirety of the potential issues associated with forwarding integrity, and if the onlooker is on the forwarding path already then routing security is of no help whatsoever. Not only are applications unwilling to wait for a secured routing

environment, neither will secured routing address the full range of issues they face with respect to the integrity of application operation in any case. Applications have turned to using their own solutions that provide remote end authentication and session encryption.

Would a secured routing system annul the motivation for applications to perform authentication and encryption? This would appear not to be the case. There are many potential traffic observation and interception points within the Internet even without the potential of adding more via deliberate manipulation of the routing system, and the benefits of authentication and encryption are largely unaltered for applications and application services, whether or not the routing system itself is secured. So, the answer to this question is an emphatic "No".

Applications and services still need to equip themselves with protections to ensure the integrity of the service irrespective of the state of routing security. Applications should protect their network traffic and authenticate the identity of the server through the use of Transport Layer Security (TLS). If applications rely on the integrity of the Domain Name System (DNS) (and most do, directly or indirectly) then the relevant domain names should be signed with the DNS Security framework (DNSSEC), and all DNS clients should perform DNSSEC validation, to prevent unauthorised efforts to manipulate DNS information. Applications need to perform these steps irrespective of the presence of a secured routing framework.

As well as masquerading and unauthorized inspection, a routing attack can also result in denial of service where the routing manipulation prevents clients from accessing the service at all. We have seen the application environment already adjust to this form of threat with the increasing use of anycast services to both diffuse a broad-based attack and limit the damage from a single point attack. Routing security does not directly address other forms of denial-of-service attack, notably various forms of overwhelming the service with either high volumes of spurious traffic or high volumes of service requests, although once again anycast can be used to mitigate the severity of these attacks in many instances

The conclusion here is that the application space has elected to create defences from various form of misdirection attack using application-level mechanisms that do not rely on a secured routing system. From the perspective of the application, a secured routing system falls into the *nice to have, but not absolutely necessary* category. Ultimately, it's up to the application to defend itself against such efforts to subvert its operation.

Does this mean that applications are beneficiaries of the network's effort to deploy secured routing? Well, yes, to some extent.

Does this mean that applications would be motivated to fund networks to deploy secured routing? Well, no, not at all. Not in this specific case, and apparently not in general terms either. The larger set of tensions between content and carriage are playing out in favour of the content and services folk, and it's likely that little will change this position in the near or medium future.

We are left with the conclusions that there are non-trivial costs associated with the deployment and operation of a secure routing framework, that this framework has some serious functional shortcomings, and there is no natural market-based incentive that will drive such a deployment in any case.

Is Regulation the Answer?

In economic terms, the question is how the incremental expenditure associated with the provision of a secured routing system is to be recouped. Would it be a competitive differentiator between networks, where a network that offers a secured routing environment be able to charge a premium for the service? As we've explored, this proposition is highly dubious. In terms of competitive incremental benefit, the case to deploy secure routing from a self-centred perspective is quite weak. It appears that the case to support secure routing is more in common with the case to support a public good. The actions of each actor in deploying secure routing enhances the common public routing space. The implication of this

observation is that any market-based mechanisms would not necessarily provide strong incentives for deployment on the part of each individual actor.

If this outcome is generally considered to be desirable in terms of the public interest, but unachievable in the context of a market-based response, then market intervention in the form of regulation is one option as a response. If all actors are required to deploy secure routing technology, then the issue of competitive disadvantage is largely addressed. No individual network actor is able to avoid the cost of deployment of secure routing measures while all networks obtain the indirect benefits arising through the deployment of secure routing measures by others.

However, there is a consideration that undermines the case for regulatory intervention at this point in time. The technology being used to support secure routing is just not ready, and there is a valid concern that this is not a situation that will change in the foreseeable future. We have already considered the challenges in the deployment model of AS Path protection. A robust routing security framework requires protection of the AS Path attribute, and without such protection is it still possible to mount hostile routing attacks that any secure origination-based framework simple cannot detect.

However, the problems with securing the AS Path have already been noted as presenting some very formidable challenges, not only with the technology of the BGPSEC protocol but with the deployment models of this type of chained credentials that are intolerant of piecemeal partial adoption. There also remains a fundamental question as to whether the particular technology choices used in this RPKI-based framework to distribute credential has efficient scaling properties. There are related questions of the robustness of the management of the key information in the PKI, and the ability to recover from situations of key compromise.

There is also the more general issue that protecting the integrity of the routing protocol is not the same as protecting the integrity of packet forwarding. Is the intention of a regulatory imposition an outcome of a secured routing protocol or an imposition placed upon the network's operators to support some secured form of packet forwarding? And if the former has been giving us trouble in determining how to design and deploy such a framework where does that leave the far more ambitious objective of secured packet forwarding?

Using public regulation to mandate the use of an immature technology for unclear reasons that introduces additional costs and risks without absolutely clear user benefit tends to undermine the essential role of the public regulatory regime.

Also, the case for regulatory imposition is not so effective if the realm of regulatory control is not matched to the scope of the issues that such controls were intended to address. At a national level the incremental costs of national requirement to support secure routing are borne by those networks that operate within the regulated environment, yet the risks associated with corruption of routing still remain outside the scope of the national or regional environment. It has been a conventional response to such situations to use the general adoption of a common set of regulations such as, for example the ITU-T's convention of common adoption of regulatory recommendations. This is not a clear policy option for the Internet and its routing environment, as many nation states have expressed a strong preference for its evolution to be guided by private investment and market-based disciplines and at the same time eschew any form of common Internet regulation or the adoption of dictates from international treaty organisations.

On the whole, such a deregulated market-based stance has proven to be effective in allowing the entire Internet industry to focus on what users will devote their time, attention and money on. It has acted as an incentive to allow the industry to rapidly assimilate new technologies where there are clear early adopter advantages that would otherwise present challenges to a highly regulated framework. For example, the Internet industry responded rapidly to the fundamental changes introduced by the transformation of the mobile industry from dedicated voice to data, which has presented significant challenges in terms of service models, scaling, application behaviour and security and integrity. Similar fundamental challenges are expected with the next wave of digitization and automation that passes within

the collective term "the Internet of Things", and there is a general sense of optimism that the Internet and its collection of actors will similarly adapt to such challenges.

However, this does not mean that market-based responses are always effective. The protracted transition to adopt IPv6 is a case in point where the market signals that were intended to propel the change in platform technology were interpreted in diverse ways. Many early Internet platform operators appear to have secured adequate pools of IPv4 addresses so that they have been able to determine their own schedule for adoption of IPv6. Later market entrants have been constrained by the increasing scarcity and associated price premium in the address trading markets of IPv4 addresses, and while their need to adopt an IPv6-only platform is clearly evident, the lack of sufficient widespread use of IPv6 elsewhere makes such an option unviable. These actors have been forced to adopt IPv6 as a measure to alleviate the pressure on their current pools of IPv4 addresses. The lack of clear common market signals in this space has caused considerable market confusion and lead to many structural inefficiencies in attempting to support continued scaling in the network while operating across two distinct network protocol platforms.

The original projections of a complete transition to an all-IPv6 network in less than five years have had to be discarded. The reality is that after nearly three decades, the industry is no closer to understanding when this transition will be complete, if ever. The picture of adoption of a secure routing framework has some disturbing similarities to that of IPv6 adoption.

As well as the direct cost of supporting a secure routing framework there is also the consideration of a systemic cost via an implicit change to the Internet's routing model. The highly decentralised model of routing that is used today is also extremely resilient. There is no point of overall control, or even a small collection of control points. Each network is a peer routing entity, and the network's ability to learn and advertise routes to its peers is not regulated or controlled by any third party. Introducing a security framework adds some control over exactly what routing information can be learned and advertised, but the security framework itself is then part of the routing system's control framework. The security credentials use a hierarchical PKI, and this places the Regional Internet Registries (RIRs) at the apex of this PKI framework as putative trust anchors for the entire PKI. Were the RIR's certificate publication systems to be compromised in any way, or to be taken offline for all or some clients, then the foundation of the secured routing system for the entire Internet is compromised.

This illustrates one of the basic design tensions in the routing system, in that a fully decentralized distributed system is highly resilient to some forms of hostile attack due to the absence of one, or a small number of, control points that are a feature of a centrally orchestrated system. However, such decentralized systems have strong elements of mutual trust, and individual actors can readily subvert the integrity of this system. Introducing points of orchestration allows for the use of a reference point whereby the integrity of information can be validated, but at the cost of introducing critical control points that become a potential target of hostile attack.

Who should operate such control points for the Internet's routing system? In what legal jurisdiction should they be operated from? How should their operation be funded? What is the decision process that should be used to address such questions? In a purely private communications domain such answers are entirely obvious: the operator of the private domain assumes such responsibilities. When we apply a similar test to a public communications platform then it is challenging to invoke a control regime that entirely eschews all forms of public engagement and public policy.

The nature of the Internet today could be characterized as a grand experiment in how to sustain a robust, functional and responsive global public communications platform based predominately on the drivers that exist in a deregulated space, drive largely by private sector investment and engagement. The achievements over the past two decades have been transformational for most of the world's population one way or another. However, that does not mean that all of the issues have been adequately addressed and the issue of how to collectively respond to these basic questions of mutual trust, authenticity and integrity in the communications system remain as largely open questions in today's Internet.

The open issues in the effort of how to design and sustain a secure routing infrastructure reflect these larger issues, and it is perhaps naïve to think that these secure routing issues can be addressed independently of the larger questions about how to sustain an open, decentralised, and vigorous and trustable public communications platform for the entire world.

Conclusions

Is secured routing a market failure today?

I would have to offer the personal view that the answer is a clear "Yes", it's a market failure.

It has failed in this respect because it has failed in so many other respects. It's a failure in technology, in that it fails to offer a scalable and robust solution. It's a failure in supporting a realistic deployment model of piecemeal partial adoption. It's a failure in terms of being able to incentivize deployment by offering tangible incremental benefits to early adopters. It's a failure in terms of the alignment of cost and benefit. It's a failure in the trust model that is being imposed upon relying parties. It's a failure in terms of the overall objective, in so far that securing the routing system is not the ultimate objective here. The real assurance here is that of assuring the sender that the hop-by-hop forwarding path that will be taken by the packet will reach its intended destination.

But mostly it's a failure because it does not deliver. Security solutions that offer only a thin veneer of the appearance of improvement while offering little in the way of improved defence against determined attack are perhaps worse than a placebo. They become nothing more than a costly distraction. Unfortunately, the currently proposed routing security framework that applies a detached PKI and signed authorities to route origination while being unable to offer a practical and deployable approach to secure route propagation falls into this category of vapid pseudo-solutions.

Is this fixable?

In my view some of these aspects of failure are, to some extent, fixable, but from my perspective the larger issues remain uncomfortably elusive.

We could contemplate dropping the entire parallel credential distribution infrastructure and just use BGP to perform the timely flooding of credentials along with updates. We could contemplate the re-design of the RPKI certificate structure and use registration agents to reduce the length of validation paths, while reducing the number of points of unqualified trust to a single apex point.

But these measures to not address the entire spectrum of failure. Some of these are extremely challenging topics, not the least of which is the visible assurance of the integrity of the end-to-end forwarding paths on the Internet. Other areas are issues of defining an acceptable balance between the extremes of absolutism of security at all levels and complete mutual trust and dependency.

Does it matter if it's not fixed?

Perhaps asking wheter all these component failures of a secured routing framework are fixable is the wrong question. The more relevant question is: *Does it really matter if we don't fix them all?*

Who is dependent on us fixing all of these issues? What valuable application or use model has been patiently sitting off to one side for some decades waiting for us to deliver a secured routing framework? As far as I am aware there is no such application or network use model. Nobody is awaiting the delivery of a completely secured routing system to address some critical, and as yet unmet, dependency.

Instead, we have seen applications make progress by assuming that end-to-end forwarding paths may lead a packet to passed to the wrong destination some of the time, and it is up to the application to identify this situation and drop the connection. These days our reliance on TLS is close to complete, and within the TLS framework the authentication of the server is a critical element. It transforms the common assumption from "*If I send a packet to the correct address then I can believe whatever I get back in response*" to the assumption that "*If I send a packet to the intended server and I can assure myself that this remote server has proved its identity to me, then I can believe what the server sends me back in response.*"

But there is still a troubling weakness here. TLS certificates that perform this identity verification work on domain names, and the assumption here is that the DNS is also capable of defending itself. In the case of the DNS, we need an essential assurance that the contents of DNS responses are authentic. This is precisely the intention of DNSSEC. If all DNS responses were signed and all forms of DNS name resolution performed DNSSEC validation then the DNS would also be able to defend itself against many forms of insidious attack, including routing attacks.

What's the real question here?

The question of whether routing security is a failure, market or otherwise, is perhaps the wrong question to ask. The real question is all about the current state of the adoption of technologies to support application-level integrity of operation. Here the story is far more reassuring. The available data relating to the adoption of DNSSEC and the use of TLS paints to a more positive alignment of cost and benefit, and the trajectory of adoption seems to be more optimistic.

The Internet is not a network-centric model where credulous applications blindly accept all network behaviours as authentic and trustable. The paradigm shift for the Internet was to strip out as much as possible from the network and allow each application to define the terms of engagement and trust in the network. Assisting applications to improve their ability to do exactly this puts the objective of a secured routing infrastructure in a different light. Yes, some level of predictability that most packets are delivered to their intended destination good to have. But in the same way that packets can be lost, reordered and mangled in various interesting ways, packets can be misdirected, and that should not lead to instant catastrophic failure. Applications need to be tolerant of some level of errant behaviour by networks in their handling of packets and should not expect networks to always maintain stringent levels of adherence to absolute levels of integrity of packet handling. Prudent applications should treat all networks as untrusted.

What does that imply for efforts to improve the assurance in integrity of routing? For me it's a balance between the marginal improvements that such efforts can offer, and the incremental cost of their application in terms of actual cost of adoption and the implicit cost of exposure of new forms of vulnerability and fragility. It's not a case of absolute security at all times at any cost, but neither is it a case that no security framework at all is acceptable. It's not a balance that we can describe in forms of defined and possibly regulated behaviours. But, oddly enough, it's a dynamic balance that markets can achieve in terms of factoring in costs, risks and opportunities into service offerings and the costs and risks that consumers of these services are willing to take on.

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