The ISP Column
A monthly column on things Internet

July 2009

Geoff Huston

The State of SIDR

In the last week of July the IETF held its 75th meeting in Stockholm. Most of the active working groups hold meetings in the week to review progress, debate current issues in their work and possibly consider the next steps to take in meeting their objectives. Sandy Murphy and I are the co-chairs of the Securing Inter-Domain Routing, or SID, Working Group. The work has been underway for some years now, and in this column I'd like to take a look at where the SIDR work has got to and where it is heading.

Understanding the need for Routing Security

While network administrators can do a relatively robust job in protecting their local routing infrastructure by taking reasonable measures in terms of protecting routers and protecting the integrity of the operation of routing protocols within their local network, the task of protecting the integrity and accuracy of the routing state of the entire Internet is a far more challenging task. What we do today, while well-intentioned, is not proving to be resilient against even such unintentional forms of attack in the form of routing information corruption through operational mis-configuration and inadvertent routing leaks, let alone being resilient against determined and novel forms of routing attack.

It's an area where the rewards of mounting a successful attack in the routing system can be very high. It is possible to undertake denial of service, third party traffic inspection and service cloning, and to do so in a manner that can be challenging to detect through deliberate corruption of the information carried in the routing system. For example, by injecting a false route describing a path to an anycast instance of a DNS root server, an attacker can create a network sub-domain where DNS resolution queries are passed to a fake server, who, in turn and manipulate the responses associated with the intended victim, while answering all other queries accurately. Or, more directly, an attacker can create a fake version of an online commerce service, subvert a subdomain of the internet with false routing information and thereby harvest users' access credentials. Obviously the rewards of attack can be high, whether it's a targeted attack against a single service or a coordinated effort to disrupt the operation of large sections of the network.

The potential outcomes of attack on the routing system can include:

- **Blackholing** where false routing advertisements redirect traffic away from its intended destination and instead are directed to a sink point. The outcome of this attack in an effective denial of service attack, where the target service is taken offline. A side-effect may be a rearrangement of traffic flows that could overload some network links.
- **Impersonation** where false routing advertisements redirect traffic away from the intended destination and instead direct traffic to a site that masquerades as the destination service. Commonly this form of masquerading is used to gather otherwise confidential information from users of the original service.
- **Inspection and Alteration** where targeting false routing advertisements cause traffic to an intended destination to be forwarded towards a compromised network segment, where the traffic may be inspected, or even altered before being passed onward to the actual destination.
- **Denial of Service and Network Destabilization** where large scale generation of updates and withdrawals of route objects may trigger routers to suppress routing information through existing route damping responses, which, in turn generates further instability in the routing system.

The attack mechanisms to subvert the inter-domain routing space can include man-in-the middle attacks**,** where false updates or withdrawals are injected into the routing system, or withholding, where legitimate

route updates are withheld from either party. The other form of attack is a remote attack, where TCP reset messages are directed at a BFP speaker in order to disrupt the session.

The Scope of the SIDR work

The scope of activity in securing the inter-domain routing system can encompass a very broad agenda. The potential agenda includes the topic of securing routers from hostile attack and takeover, securing the transport sessions used by the routing protocol against hijacking or disruption, and securing the identity of the parties to each protocol conversation to ensure that the routing protocol conversation is being undertaken with the intended party and not an imposter. However, these aspects of routing security are not unique to only BGP. There are existing mechanisms used to secure host platforms from attack and takeover, and resist attempts to exhaust the system's resources through various forms of DOS attacks, such as SYN flooding, and any public router should certainly adhere to the current best operational practice in terms of defensive measures to secure the system from such forms of hostile attack. There are also existing mechanisms to protect long-held transport sessions that are applicable to BGP, including using the TTL hack to limit the radius of potential disruption, and the use of MD5 protection over the TCP session. Because such activities are already the focus of existing security efforts in the area of standards and in operational procedures, the SIDR Working Group is not attempting to re-invent those particular wheels. Instead, SIDR has focussed its effort in the question that is unique to BGP, namely: *How can a BGP speaker assure itself that the information being passed to it via BGP is authentic?*

The starting assumption here is that the BGP session is protected and the credentials of the remote party have been authenticated in some fashion, so that there is some appropriate level of assurance that BGP speaker is communicating with the party it was configured to communicate with.

So what is left is the content of the BGP session, and securing this is up to SIDR.

The scope of the working group's activities can be seen in its charter, that was defined when the working group was set up.

> SIDR Charter
>
> The basic security questions that can be posed regarding routing information are whether the originating Autonomous System is authorized to advertise an address prefix by the holder of that prefix, whether the originating AS is accurately identified by the originating Autonomous System Number in the advertisement, and the validity of both the address prefix and the Autonomous System Number. A related question concerns the level of trust than can be ascribed to attributes of a route object in terms of their authenticity, including consideration of the AS Path attribute.
>
> The Routing Protocol Security Group (RPSEC) has been chartered to document the security requirements for routing systems, and, in particular, to produce a document on BGP security requirements.
>
> The scope of work in the SIDR working group is to formulate an extensible architecture for an interdomain routing security framework. This framework must be capable of supporting incremental additions of functional components. The SIDR working group will develop security mechanisms which fulfill those requirements which have been agreed on by the RPSEC working group. In developing these mechanisms, the SIDR working group will take practical deployability into consideration.
>
> The scope of work will include describing the use of certification objects for supporting the distribution of authorization and authentication information. Both hierarchic and distributed non-hierarchic trust systems are intended to be supported within this framework. The intended support of both forms of trust models is to allow for the use of this framework for routing security in diverse routing environments that have different underlying trust characteristics.
>
> The scope of work is limited to inter-domain router-to-router protocols only, for both unicast and multicast systems.
>
> The SIDR working group is charged with the following tasks:
> - Document an extensible inter-domain routing security architecture
> - Document the use of certification objects within this secure routing architecture

- Document specific routing functionality modules within this architecture that are designed to address specific secure routing requirements as they are determined by the RPSEC Working Group

The intended approach was that one working group (the RPSEC working group) would work through the objectives of a security framework and come up with a set of "requirements", and the SIDR working group would then work on the definition of mechanisms that would meet these requirements.

The area where requirements could be voiced was described in a RPSEC working document, and it identified the information contained in the BGP Update message as the critical component to secure. The first requirement was that a BGP speaker should be able to validate that the originating AS was a valid AS, that the prefix itself was valid, and that the prefix holder had authorized the AS to originate an advertisement for the prefix into the inter-domain routing system. The second requirement concerned the transitive component of the Update, namely the AS Path, and the requirement was for a BGP speaker to be able to have some level of assurance that the AS Path represented a valid transit path through the network.

While the SIDR WG has been able to work on the origination requirement as a clearly stated requirement, the AS Path security requirement is a little less clear. How stringent should the AS Path validation test be? Does the level of Path validation require that the BGP update has indeed traversed the AS's noted in the AS Path, and that each AS attested that it advertised what it believed to be a viable path to the prefix to its successor AS on the AS Path? Or is the level of validation one that allows a BGP speaker to conclude that the AS Path represents a feasible transit path through the inter-AS network without necessarily validating that the BGP Update has traversed this path in reverse order? Given this level of uncertainty over the level of requirement for AS Path validation, the SIDR WG has, to date, looked exclusively at the problem of route origination.

Securing Origination

In order to meet the origination requirements a security framework needs to be able to allow BGP speakers to be able to confirm that a given address prefix and a given AS are indeed valid addresses.

A simple assertion of the form "these are my addresses" is not terribly effective, given that the problem is that its just as possible to assert a lie, and there is no obvious way to resolve a clash in assertions between two or more parties. One way for a relying party to resolve this is to refer back to the address distribution structure, and challenge such assertions with the question: "Who supplied you with this address?" If this sequence of challenges can be traced back to someone the relying party is prepared to trust as being the original source of all addresses, then the resultant framework can be used to secure attestations about addresses and their use.

Over the years there have been a number of proposals about how to create this interlocking hierarchy of assertions. Two approaches appear to offer the greatest promise. One is to use DNSSEC and place these assertions into the DNS using a dedicated resource record, and use the interlocking key structure of DNS to re-create the address allocation hierarchy. The problem with the DNS approach lies in the need to create a DNS hierarchy that does not necessarily conform to DNS zone delegation boundaries .The other approach lies in adoption of X.509 public key certificates, and the creation of a certificate hierarchy that mirrors the address distribution structure, where a "node" in hierarchy uses its private key to sign over the public keys of all of its immediate subordinate "nodes" in the hierarchy to whom it has allocated or assigned addresses, and its own public key is signed by the private key of its immediate superior, from whom its own address pool was allocated or assigned.

Resource Certificates are X.509 certificates that contain an extension that lists a collection of IP resources (IPv4 addresses, IPv6 addresses and AS Numbers). These certificates attest that the certificate’s issuer has granted to the subject a unique “right-of-use” of the associated set of IP resources by virtue of a resource allocation action. This concept mirrors the resource allocation framework, where the certificate provides a means of third-party validation of assertions related to resource allocations. By coupling the issuance of a certificate by a parent Certification Authority (CA) to the corresponding resource allocation, a test of the certificate's validity can be interpreted as validation of the associated resource allocation. Signing operations which descend from that certificate can therefore be held to be testable, under the corresponding hierarchy of allocation. The intent of the Resource Public Key Infrastructure (RPKI) is to construct a robust hierarchy of X.509 certificates that allows relying parties to validate assertions about IP addresses and AS Numbers, and their use. The RPKI allows a relying party to determine if an address is valid to use in the context of the public Internet, and is able to validate assertions relating to the current "right-of-use" holder of an AS number or IP address.

The application of these certificates is in the generation of Route Origination Attestations, or ROAs. A ROA is an authority, issued by a prefix holder that authorizes a nominated AS to originate a route into the routing system.

So if a BGP speaker receives a route for, say, 203.0.113.0/24 originated by AS 64500, then the BGP speaker would accept this as a valid route if it could find a ROA where the holder of 203.0.113.0/24 authorized AS 64500 to originate a route for this prefix. In turn, the ROA could be validated if there was a certificate chain from a trust point to the certificate that signed the ROA.

The ROA conveys a simple authority, and does not convey any further routing policy information, nor whether or not the AS holder has consented to actually undertake the routing action.

Positives and Negatives

The SIDR work on origination has exposed some interesting design considerations when looking as how to approach this problem. One of these is the use of *positive security credentials*. In a heterogeneous environment where valid and invalid assertions coexist, one of the initial starting premises is that the invalid assumptions will not cooperate by identifying themselves.

This implies that the security credentials are intended to inform relying parties as to whether a route object is constructed using a valid origination (prefix and originating AS) as a positive confirmation of the integrity of the routing information. The implication is that attempts to create false origination information could not have valid security credentials. In a world of comprehensive deployment these positive credentials represent a sufficient capability, because within such an environment what is not provably "white" is assuredly "black". In other words, if all valid route objects are protected by a valid ROA, then if a route object does not have valid security credentials then it should be considered to be invalid.

However, in a world of partial deployment of this technology there would be a class of objects that have no associated security credentials. This implies that what is not able to be validated, either because validation fails or because there are no security credentials in the first place, cannot be considered as invalid, but at best "undecidable." In other words the route object itself may be valid, but there are no security credentials available to the relying party to make the call.

Now if the objective of this entire exercise of adding security credentials to origination was to identify lies in the routing system, then the ability to categorize route objects as either "valid" or "unknown" is not exactly the desired outcome.

> This is a clear example of the need for the "evil bit". In an April 1 RFC in 2003 Steve Bellovin defined the "evil bit" in the IP header, with the associated semantic that if this bit is set to 1 then the packet has evil intent, and secure systems would actively defend themselves against such packets. Despite obvious application in many security environments, getting the bad folk to set the evil bit on their IP packets has remained a highly elusive objective so far!
>
> If you are going to allow folk to identify the lies from the truths, and the liars won't cooperate by setting the evil bit, then you have to work through the approach of validating the truth-sayers and making the supposition that all of what's left are lies. Partial deployment makes this a little tricker, but nevertheless that's all we have.

This leads to the consideration of how a relying party can consider a route object can be considered to be invalid in a heterogeneous environment where not every truth is well supported with credentials that can be validated.

One approach is to make an object that is a specific "disavow" attestation. Earlier work in preparation for the SIDR standardization activity explored the concept of a specific disavowal, termed a "Bogon Route Origination Authority", or BOA, that attested to the intention of the holder of the address, or AS number, not to announce the address prefix or any more specific prefix of that prefix. The inference of this attestation was that any form of use of the prefix, or AS number, could be construed as invalid by virtue of the issued BOA. However this is a less than satisfactory measure in that there is still the problem of partial deployment,

and the appropriate interpretation "grey" space that is neither the subject of a positive authority or a negative disavowal, and there is the added issue of resolution of mixed cases where a route object is the subject of a valid authority and a valid disavowal.

The other approach here is to add an additional clause to the interpretation of a positive authority of the form "and no other". For example, if a positive authority is generated that refers to the origination of 203.0.113.0/24 by AS64500, then the addition of the rider clause would imply that an origination of this prefix by any other AS would be considered "invalid" rather than "unknown", and also that the origination of any prefix that is more specific than this prefix, such as 203.0.113.0/25, would also be considered to be "invalid".

The second approach can be transformed into something similar to the first by creating a positive assertion that refers to origination using a reserved AS number. i.e. a positive authority for 203.0.113.0/24 with AS 0, implies that any other use of this prefix would be invalid.

The implication of the second approach is that is a party elects to generate a positive attestation about a route object, the party is assumed to have generated a "locally complete" set of all such attestations, so that relying parties can make the assumption that within the domain constraint of a particular address prefix, there is complete use of secure credentials. The implication is that within this limited domain of consideration a relying party can make the assumption that everything that lies within the scope of this prefix (i.e., all more specific prefixes and all ASes) that is not validated through the available security credentials is indeed invalid. Perhaps unsurprisingly this is precisely the same assumption that is made by a BOA, in that a BOA has the same property of being "locally complete" with respect to the prefixes it describes.

At this stage there appears to be some sense in using a smaller set of signed objects to form the set of credentials to be used to secure route objects, so there is a preference to use a route origination attestation that includes the concept of an implicit disavowal in its interpretation.

Securing AS Paths

The work to date in SIDR has concentrated on the framework that will support validation of the origination of a route object, allowing relying parties to validate the origination of information into the inter-domain routing system. Origination protection in routing in and of itself does not do an awful lot if you can still lie in the AS path of a BGP route object. A prefix can still be hijacked in the routing system by generating a route object that preserves the originating AS number and pushing this route out into the routing system.

> From this perspective it could be said that a little security, in the guise of origination protection, is perhaps worse than none at all, given that it proposes the superficial appearance of improving the integrity of the system while in fact achieving nothing tangible whatsoever.
>
> However, origination protection alone can provide some protection for certain classes of routing attack, in spite of a continued ability to misrepresent the AS path. The class of routing attack that can be prevented is that of the origination of more specifics.
>
> The routing attack on YouTube originated from Pakistan Telecom in February 2008 was a case of a more specific attack where the original aggregate route originated by YouTube of 208.65.152.0/22 was effectively usurped by the false origination of the more specific route 208.65.153.0/24 from AS 17557 (Pakistan Telecom).Were the YouTube route to have been protected by a ROA that specified that nothing more specific than the /22 prefix was to be considered valid, then this Pakistan Telecom attack would've been addressed by SIDR's origination protection constructs.
>
> Because of the ability to misrepresent AS Path, origination protection does not protect against all forms of route prefix hijacking, but it can limit the effective scope of such a routing attack from a global scope (of a more specific route

taking precedence globally) to a localized scope. The misrepresented AS path must be shorter than the genuine AS Path in order to take precedence in routing, or the genuine route must be locally withheld, and such techniques tend to implicitly scope the attack to a network subdomain that is "close" to origination of the attack.

The question is: how can we "protect" the AS path in BGP such that attempts to misrepresent the AS path in a BGP update can be detected as invalid by any BGP speaker?

In attempting to validate an AS path there are a number of potential validation questions. The first, and weakest, question is: "Are all AS's in the AS Path valid AS's?" A slightly stronger validation question is: "Do all the AS pairs in the AS Path represent valid AS adjacencies? (where both AS's in the pair-wise association are willing to attest to the mutual adjacency)". A yet stronger question is: "Do the sequence of AS's in the AS Path represent the actual propagation path of the BGP route object?"

In looking at the academic literature on the topic it appears that folk have examined three generic approaches to securing the AS Path:

a) **Incremental crypto-wrapping** - where every agent that handles the BGP update signs across the update components that are transitive. The best example of this approach in literature is sBGP. If the AS path purports to be a "snail trail" of the provenance of the update information through the network, then this approach certainly allows a receiver to validate this implicit assertion. However, this approach implies that the security credentials for AS Path protection are tightly bundled with the BGP update. It is noted that the parallel approach to protecting route origination that SIDR has been developing is an unbundled approach, where the security credentials for route origination are distributed independently of the propagation of BGP updates.

b) Plausibility protection - where there is a separate mechanism for flooding the equivalent of a set of inter-AS link state assertions that are verifiable in some fashion (so-called AS adjacency attestations) and when the inter-domain distance vector delivers a BGP update the implicit topology fragment described in the AS path is checked against the AS adjacency map that was assembled from the collection of the adjacency attestations. If the path is not described on the AS adjacency map then there are grounds to consider the AS Path to be implausible. The best example of this approach in literature is soBGP. This is incrementally deployable, and very lightweight in terms of BGP load. It is nowhere near as impervious to attack as incremental signing of the update, but, like protection of origination, if further limits the scope of the attack. In order to hijack a route, the attacker would need to reproduce the valid origination information and then construct an AS Path that is plausible in terms of the verifiable inter-AS adjacencies.

c) Reverse check - where a BGP speaker would , in response to a received update perform a direct query for each AS on the path, and ask the AS's query server: "Did you propagate this update?". This approach has been described in the IRV approach. The possibilities for this approach are intriguing as, in theory, not only can you confirm the propagation of a BGP update, but you can interrogate each AS on the path as to the state of its forwarding, and confirm that its current forwarding state matches the information in the routing update. The problem with this kind of approach is that this places a potential for denial of service attacks into the network's infrastructure. Fascinating as this approach may be, there has been little in the manner of followup studies of this approach in the literature on the subject so far.

Another way to look at this is that approach is that approach b) uses pre-provisioned security , approach a) is "in-line" security and c) is "post-facto" security.

These differences of degree expose differences of approach in AS path validation, and also expose to some extent the current uncertainty of the costs of path validation in operational environments. They also raise the question of what degree of validation outcomes can be achieved on a per-BGP Update processing level in BGP speakers in the routing framework, and what may have to be pre-validated outside of BGP.

Given that the SIDR effort is now nearing completion in the areas of defining the PKI for resource certificates, and mechanisms to protect route origination, it seems likely that in the coming months SIDR will turn its attention to mechanisms to secure the AS Path. To make progress in this space and arrive at a position of rough consensus over which of these approaches to Path validation is the most effective, some of these

questions about how much BGP processing load is "enough" and the overall scope of routing security will need to be answered.

Disclaimer

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

About the Author

GEOFF HUSTON is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region. he graduated from the Australian National University with a B.Sc, and M.Sc. in Computer Science. He has been closely involved with the development of the Internet for many years, particularly within Australia, where he was responsible for the initial build of the Internet within the Australian academic and research sector. He is author of a number of Internet-related books, and was a member of the Internet Architecture Board from 1999 until 2005, and served on the Board of Trustees of the Internet Society from 1992 until 2001.

http://www.potaroo.net