BGPv4 Security Essentials[1]

Version 0.5

(Please send comments and correction to bgreene@cisco.com.)

BGP RISK ASSESSMENT IN TODAY'S INTERNET

Border Gateway Protocol version 4 (BGPv4) was created in early days of the Internet when the security risks were not as intense. Yet, as the protocol that glues together the largest, most stable, and most complex network ever created, BGPv4 has been refined and meet the increased operations and security risk. BGP has always been a flexible protocol – built specifically to allow new features and functionality to enhance its operational use, capabilities, stability and security. Understanding the process through which BGP has been refined is critical gain an accurate portrait of BGP's operations and security risk in today's Internet.

BGP'S SECURITY EVOLUTION

In the beginning, BGP did not have security explicit features. It did have provisions for security features to be added in the future. These features have been included over time. Each added as a fix to a perceived vulnerability or through operational experience. Most BGP *security* features are really *operational enhancements* added to provide a solution to real problem. So items like Route Flap Dampening, Community Filtering, and other *operational features* may not seem at first to be related to security. Yet, these features, which were devised to verify the correctness of routing information and the damage caused by the propagation of false routing information, directly relate to security, especially when you consider the results possible through malicious insertion of routing data in the Internet's global routing table.

What follows are some highlights of operational capabilities which have been added to BGP since its original deployment which have a direct impact on the security of the routing system.

[1] Security analysis outlined in this paper can only be certified with Cisco IOS 12.0S. Other BGP implementations may not use the approaches described in this paper. Hence, this can only be classified as a *Cisco BGP Security Risk Assessment* and not a general assessment of all the BGP implementations deployed on the Internet today.

Spoofing Risk

BGP Spoofing attacks are those in which the BGP peer is imitated. This can be TCP based spoofing targeting the BGP port of the router or spoofed BGP packets. There is a common perception that BGP is easy to spoof. Yet, some simple analysis demonstrates that spoofed attacks targeting BGP are not as simple as people believe.

To successfully spoof a TCP session supporting the BGP peers, the following must be achieved[2]:

- **Source IP address must be spoofed.** The source address of the BGP Neighbor must be spoofed. In most cases this address can be determined through ICMP traceroutes from various places on the Internet through the BGP peer.[3] This form of mapping is a more advanced technique, requiring an understanding of how routing and ISP peering works on the Internet.

- **Source Port must be spoofed**. One of the two BGP peers initiates the BGP session. The peer that initiates the BGP session will have a randomly or sequentially selected port number with in a port range (port range is depended on the TCP implementation)[4]. The BGP initiator will connect to its peer's BGP port 179. Since it cannot be predicted which side is using the random port and which side is using port 179 in the TCP session, packet capturing is required.

- **TCP Sequence Number must match.** The TCP sequence number allows for the reassembly of packets that arrive out of sequence. It provides a security role by insuring that the TCP packet which arrives matches the expect sequence number. If it doesn't the TCP packet will be dropped. While there are known techniques for TCP sequence number predictions, improvements in random sequence number generation over the years have added increase resistance to these forms of attacks. With these new techniques, capturing packets or breaking into the router are key steps needed to determine the TCP sequence number range.

- **IP's TTL must match during the initial TCP session start**. The IP TTL is a safety mechanism that ensures lost packets on the Internet will eventually expire and get dropped. Most ISP peering connections use eBGP. Since eBGP session assumes the peers are directly connected through a Layer 2 medium, the TTL of the IP packet is required to be 1. The BGP packet is dropped if the TTL is greater than 1. Since the BGP

[2] These spoofing requirements have only been validated with the TCP and BGP implementations in IOS 12.0S. Further testing on GateD, Zebra, and other vendor implementations are required.
[3] There are two techniques to hide the IP addresses used for eBGP sessions. One technique uses loopback address to do the eBGP peering. The other technique uses secondary addresses for the eBGP peering. Both techniques hide the peering addresses from traceroutes.
[4] The *randomness* of the port number depends on the TCP implementation.

speaker the attacker is trying to connect to will (most likely) be transmitting its packets with a TTL of 1, the attacker will need to be attached to the same layer 2 segment (local segment) as the router it is attempting to attack to receive the BGP packets. Alternatively, the attacker must measure the number of hops to the targeted eBGP router and determine the exact value needed to count down the TTL to equal 1.[5] Asymmetry on the Internet will add to the difficulty of TTL count down determination, but not eliminate the risk. It should be noted that TTL in of itself is not a security mechanism. However, it does add another layer of difficulty, when combined with other TCP/BGP session management/validation techniques. ***RFC 3682 The Generalized TTL Security Mechanism (GTSM)***[6] is a new proposal where the eBGP TTL is set to 255 – then only accept packets with TTLs with a TTL between 255-254. This prevents TTL crafting as a means of spoofing TCP/BGP packets.

A TCP Reset (RST) attack is an attack profile frequently referenced for an attacker who has no direct access to the link. The TCP RST is a packet that will *reset* the TCP session supporting BGP. Tearing down the TCP session also tears down the BGP session, flushing the routes for that peer.

BGP's will recover and restart the peering session. The restart could happen from the targeted BGP Peer or the target's BGP Neighbor. This depends on the state of the BGP Finite State Machine. The results would be unpredictable port numbers. Whoever initiates the BGP session will have a random port number. The other peer will be port 179. The TCP RST would change who is the random port number and who is port 179. The only way to keep track of this port number variation is to sniff the wire between the peers.

In summary, if the specific TCP and BGP follows this implementation pattern, spoofed attacks targeted at BGP would be difficult but not impossible – usually requiring packet sniffing of the BGP session or in depth knowledge of the topology and config. Added resistance is gained from the two types of inter-ISP peering connections: private peering and IXP peering. Most peering connections are private peering with some sore of dedicated layer 2 medium (i.e. like POS) connecting the routers of the two providers. The second type is Internet exchange Point peering – where ISPs connected on a layer 2 shared medium. Both peering techniques provide a level of physical security and accountability that increment the BGP Spoofing difficulties. Together, they explain why BGP Spoof attacks are not a common attack vector on today's Internet.

Yet, the Internet community did not stop here. Spoofed attacks, while extremely difficult to execute, were feasible. So BGP evolved with the addition of RFC 2385 - *Protection of BGP*

[5] If you are 5 hops away from the router, you must set the packet's TTL to be 6. This would decrement the packet by one so that it will be a TTL of 1 when it hits the targeted eBGP session.

[6] **ftp://ftp.rfc-editor.org/in-notes/rfc3682.txt**

Sessions via the TCP MD5 Signature Option. TCP MD5 added another layer of difficulty to BGP Spoof attack vector. Now in addition to everything else, the MD5 key needs to be known or broken. So even if someone could *sniff the wire*, they would still need the MD5 key to execute an effective BGP Spoof attack.

Counter Measures: MD5 on BGP peering session mitigates most wire sniffing threats to BGP Spoofing. Use of diverse keys on eBGP session with fellow ISPs would mitigate the risk of the MD5 key from leaking. If operationally feasible, treating MD5 keys with changes policies similar to password change policies also mitigate the risk. Difficulties with MD5 key maintenance within an operational ISP environment is one of the core reasons given for ISPs choosing to not implement MD5 in their network.

Hijacking Risk

BGP Hijacking requires a success BGP Spoof. These attacks masquerade BGP status packets as coming from the neighbor. The packets would look legitimate, but would carry malicious BGP status updates. The updates could be tearing down the BGP session, inserting routing information, or withdrawing valid routing information. While sounding dangerous, effective BGP Hijacking requires additional knowledge of the current BGP interaction between the peers. For example, if a BGP Update message is sent attempting to inject a new prefix into the BGP Table, specific knowledge of the peering connection is required. Next-hop, BGP communities, prefix filters, and other details on how the peering is configured add to the difficulty of a successful BGP Hijack

Counter Measures: MD5 on BGP peering session mitigates most wire sniffing threats to BGP Hijacking. Work is in progress on a BGP over IPSEC option that would greatly increase the difficulty of hijacking. Although, the industry does not known if BGP over IPSEC will be *operationally deployable*.[7] Lessons learned from MD5's key maintenance limitations demonstrate that hashing or encrypting routing protocols are not trivial task in an operational network.

Route Flapping Risk

In the mid-1990s the Internet experienced a significant problem with excessive route flaps. A route flap is a two state change to a route in the Internet's Global Route table. This could be an existing route that is withdrawn or a new route that gets added. Each time one of these route

[7] *Operationally Deployable* means that an ISP can cost effectively deploy, troubleshoot, and maintain the protocol/configuration. If it is not cost effective, then the added cost may not off set the perceived benefits.

flaps occur, BGP must work through and update its tables. This impacts the load on routers – as they have to constantly recalculate BGP – forcing the router's CPU to saturate. The result is a network with convergence and stability problems.

In the mid-1990's, excessive route flaps caused a significant stability problem on the Internet. The industry responded with a *route flap dampening* algorithm proposed by Curtis Villamizar. Equipment vendors quickly implemented the new dampening technique, allowing for ISPs to deploy updated code and mitigating the effects of the excessive route flaps. The results and algorithm are articulated in RFC 2439 - *BGP Route Flap Damping*.

Several years later, the RIPE-NCC Routing Working Group reviewed the effectiveness and lessons learned from the excessive route flapping on the Internet. These lessons pointed out how the route flap dampening algorithm could be used as a denial of service attack. Someone could "flap" the route of their target. The flapping route would trigger the route flap dampening feature on routers, removing that route from the Internet's Global Route table until the flapping stops. Given this security risk, the RIPE-NCC Routing WG published a recommended route flap dampening configuration for all ISPs. This configuration would protect specific routes from never getting dampened (like the DNS root servers) and weighing the flapping penalties for other range of routes. This work is articulated in RIPE-229 - *RIPE Routing-WG Recommendations for Coordinated Route-flap Damping Parameters.*

Counter Measures: Implementation of RIPE-229 minimizes the dampening risk to critical Internet resources. Implementation of aggressive route filtering on customers and peers also minimize the security risk posed from route flap dampening attack vectors.

De-aggregation Risk

On Apr 25, 1997 at 11:30 a.m. AS 7007 announced more specific routes (/24s) for practically the entire Internet. This *de-aggregation* of the Global Internet Route table had immediate effects on the entire Internet. Routing was globally disrupted as the more specific prefixes took precedence over the aggregations routes. Routers with 32M of memory, which was fine for the Global Internet Route table of that time, now had tens of thousands of more routes – in some cases causing router crashes. More specifics being advertised from AS7007 into AS1239 sucked traffic from all over the Internet into AS1239 – in some cases saturating links and causing more outages.

The Internet community immediately reacted, adding filters, applying dampening features, unplugging AS7007, and working together to mitigate the effects. Within hours, the Internet was stable again. Yet, this *de-aggregation* even had immediate ramifications. A new push for ISPs to implement route filtering on their customers was initiated. At the same time, customers wanted a

new BGP feature that would be a failsafe tool incase a de-aggregation event/attack happened in the future.

The risk of a de-aggregation event/attack is real. Multihomed customer with BGP speaking routers could be broken into and used to launch a de-aggregation attack. If the upstream ISP does not implement ingress route filtering on their customer, the effective of this attack would impact the ISP. Propagate across the entire Internet is less likely. Most ISPs are implementing stricter route filters on their peer connections – limiting what they send and receive. This filtering compartmentizes the risk to a few ISPs – leaving the others unaffected.

The lesions learned from the AS7007 incident demonstrated the need for Murphy's Law protection feature that would limit the maximum number of prefixes that would be accepted from another peer. A Max Prefix-Limit feature was added to many BGP implementations. Usually this feature would shutdown a BGP peer when the maximum number was reached (the max number being configurable). Alternatively, the max prefix-limit feature would just notify via syslog when the max threshold was reached. Max Prefix-Limit features is another example how BGP is refined through experience.

Counter Measures: Max Prefix Limits on peer connections combined with aggressive route filtering of the ISP's customers effectively mitigates the de-aggregation risk. ISPs should only permit customer prefixes for those IP address blocks that have been allocated to them by the IANA system. These IP allocation records can be validated through the RIR databases, their customers, and their peers (if the customer is a multi-homed customer).

DUSA Route Injection Risk

Some IP addresses should never appear in the Global Internet Route Table. Documenting Special Use (DUSA) IPv4 addresses have been allocated by the Internet Assigned Numbers Authority (IANA) for very specific functions on the Internet. These are:

- 0.0.0.0/8 and 0.0.0.0/32 - Default and broadcast
- 127.0.0.0/8 - Host loopback
- 192.0.2.0/24 - TEST-NET for documentation
- 10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16 - RFC 1918 private addresses
- 169.254.0.0/16 - End node auto-config for DHCP
- 192.88.99.0/24 – RFC 3068 Anycast Prefix for 6to4 Relay Routers

Malicious route injection of any of these addresses might cause disruptions in networks that use these addresses within their designated functions. BCP for ISPs are to filter these routes coming into and out of their network.

Counter Measures: DUSA Filters on the ISPs peering and customer connection effectively prevent these attacks. These DUSA filters should be on route coming into and out of the ISP.

Un-Authorized Route Injection Risk

Advertisement of routes in which the network does not have allocation authority pulls traffic away from the authorized network. This causes a DOS on the network who allocated the block of addresses and may cause a DOS on the network in which it re-advertised. The opportunity of malicious abuse presents itself when you combine the industry trends of multiple links to the Internet (referred to as multihomed customers) combined with inadequate security practices in these networks. The address spaces of these multihomed customers are frequently scanned[8] and the routers potentially violated.[9] These violated multihomed routers speaking BGP with their upstream ISP are now potential platforms for BGP attacks. The easiest attack vector being advertisement of someone else's IP address block.

IP Addresses are allocated under the guidelines of RFC2050 – *Internet Registry IP Allocation Guidelines*. These guidelines set up a hierarchy of allocations with the IANA on the top, then the Regional Internet Registries (RIRs), then the ISPs who act as Local Internet Registries (LIRs), and finally the customers of the ISPs (LIRs). No one on the Internet *owns* IP addresses. These addresses are allocated to RIRs, ISPs, and customer as a temporary resource. Customers who move from one ISP to another must return the allocation to their old ISP.

This *provider based allocation* system requires an up to date databases of who is allocated which IP address block.[10] While the function of these databases are to keep track of the allocations, it also provides intelligence information on which block of addresses are allocated to end users of the Internet. This allows an attack vector where someone breaks into a BGP speaking router, advertises the specific address of the target, and has the Global Internet Route table forward traffic to the violated router vs the target.

Counter Measures: Aggressive egress routing filtering prefixes set to other ISPs on the peering points (and customers) mitigate the risk of malicious advertisement of un-authorized routes into the Global Internet Route table. With significant limitations, these IP allocation records can be validated through the RIR databases, the ISP's customer databases, and their peer contacts (if the customer is a multi-homed customer). This egress filtering contains malicious advertisement from a violated router within an ISP's Autonomous System – keeping the advertisement from spreading to other ISPs.

[8] Scan rates are based on internal data from CERT and Cisco's PSIRT.
[9] Customers who do not implement BCP principles for securing a router are the most common victims.
[10] Each RIR maintains a database of their allocations to the ISPs. Some ISPs maintain their own databases of their allocations to their customers. ISPs are required update their RIR's databases before they are granted additional allocations. This requirement keeps the databases up to date.

Un-Allocated Route Injection Risk

Advertisement of IP addresses that have yet to be allocated by IANA can pose several problems on the Internet. Two feasible ones which can be inferred from the AS 7007 incident and the experience with Code Red and Nimda are BGP table explosions and the use of latent backscatter as a DOS tool. Most ISPs do not filter Bogons – the term used to describe the IANA reserved address space. A malicious attack might use a violated BGP speaking router to start advertising large ranges of Bogon space – with the objective of overloading BGP and forwarding tables in routers.

Bogon advertisement could feasible turn the advertising router into an *Internet Sink Hole*. Many spoof DOS/DDOS attacks use the unallocated addresses as their source addresses. When these DOS/DDOS attack hit a target, the target normally responds with ICMP Unreachable messages back to the source address. These ICMP Unreachable messages echoing from a target are called the *backscatter* of an attack.

Several common DOS/DDOS attacks use source addresses from the un-allocated blocks. Normally, the DOS/DDOS target's gateway router or the upstream ISP drops the majority ICMP Backscatter. Since at any give time there are tens of DOS/DDOS happening on the Internet, the opportunity exist to exploit the backscatter as an attack. This *backscatter* traffic can be pulled to a target by advertising the un-allocated IP address blocks – creating a DOS against the advertising router.

Counter Measures: Bogon Route Filtering – filtering all address blocks that have yet to be allocated – is an effective counter to this attack vector. IANA maintains public list of Ipv4 allocations (http://www.iana.org/assignments/ipv4-address-space). The *IANA Reserved* blocks are the Bogon blocks. This list is used to generation Bogon filters for networks who wish to use them.

Direct DOS/DDOS Risk to the Router (Resource Saturation Attacks)

DOS/DDOS Attacks directly against the BGP protocol port (port 179) are perceived to be an easily executable attack vector. Yet, as seen in previous sections, BGP implementations should not accept any packets that are not exact matches (i.e. successful spoof attacks). What are common are various forms of *resource saturation attacks*. These attacks (like a TCP syn flood against port 179) attempt to flood the application port. In reality, they end up flooding a resource like the input queue, forcing the router's processors to work over time with queue maintenance. At times, queue and processor resources can reach the point where control plane packets are

dropped. When control plane traffic is dropped, the routing protocol sessions drop resulting in a router flap.[11]

ACLs to protect BGP mitigate some direct attacks, but not spoofed attacks. Spoofed attacks only need to match the source IP address of the BGP peer. Once, spoofed, the packet passes right through the ACL. IP Source validation on the edge of an ISP's network would also help mitigate the risk, but this would need to happen across the entire breath of the Internet to have any real effect. Other proposed BGP mitigate techniques are just as vulnerable to these sorts of *saturation attacks*. BGP with MD5, BGP over IPSEC, or other BGP security proposals are all exposed to resource saturation attacks.

Counter Measures: The most common counter measure is to increase the input queue depth to the point where router has enough room to drop the attack packets and still have room to keep the control plane traffic. Other techniques include TCP state management techniques that would not response or clear out SYN and SYN/ACK floods.[12] Smaller routers can still have their resources over loaded – flapping the route. In this case, multi-layered/multi-level redundancy designed used on today's ISP networks allow for the back-up path to maintain network integrity.

Risk Related to an ISP's Routing Architecture

The way the network is designed effect how it responds to attack. As seen with the 2001 Code Red and Nimda incidence, ISPs who advertise a default route on one or more of their routers turn those routers in to magnets for malicious traffic with no path in the forward table. Routers under direct attack which flap under an attack does not stop the attack. The traffic of the attack still has to go somewhere – which means another router can be effected by the attack. Security is an essential part of ISP network design. Those ISPs who do not know about these security architecture principles tend to have networks that experience more attack stress than is necessary.

Counter Measures: The common ISP routing principle of <u>not</u> running default router is a common and effective counter measure. Black hole routing un-allocated prefixes (Bogons) on all routers are an alternative some providers consider. Creating sink hole networks – which advertise Bogons, infrastructure aggregates, and default inside the ISP – is another technique used pull backscatter and lost attack traffic to a section of the network designed to drop packet.

[11] Router flaps happen when a direct DOS attack saturates resources and drops the routing protocol packets. The routing protocol session will drop, which drops forwarding over the link to that router, changing the destination of the attack flow. At that time, the routing protocol recovers, restoring traffic flow, and allowing for the attack flow to hit the targeted router again – flapping the router again. The result is an oscillating attack.

[12] SYN Flood is a resource saturation attack consuming the TCP stack resources.

Risk Related to BGP Bugs

BGP implementation bugs do happen, but are normally caught and corrected through the internal test or through ISP operations teams. As a norm, they do not pose a *security* risk to the Global Internet Table. They have do cause operational risk to the Global Internet Table. Some bugs – especially interoperability bugs – have posed a significant operational risk to the Internet. Some have cause outages. Others have caused inconveniences. Most are quickly contained and fixed by provider-vendor collaboration.

What does pose a risk to the Global Internet Table is the breakdown of inter-vendor compatibility testing. As seen with a couple of recent incompatibility issues, BGP vendor compliance testing are now done live on the Internet. In the past BGP vendor compliance testing was done in University router test beds and interoperability nets (like Interop). Over the years, the rapid growth of the Internet and compeditive pressures has pulled apart inter-vendor interoperability testing. Today, most interoperability bugs are discovered through operational experience on live networks. This is not an optimal situation for the industry.

BGP Community Attribute Risks

BGP Communities are attributes linked to route advertisements. They are used for a variety of policy, filtering, and path selection task. These communities are additive and transitive BGP attributes. Which means that as the BGP route advertisement passes a router, that router pass the community to the next ISP (transitive) and can add another community to the list (additive). This opens the door for malicious manipulations of path selection, path preference, and abuse of other functions where BGP communities are used.

The industry responded with the additional capabilities to filter BGP communities. This allows ISPs to pick groups of communities are and are not allowed through their network.

Counter Measures: BGP Community filtering on the ISP peering and customer edges matching an ISP's policy of which Communities are passes, which are added, and which are deleted. BGP Community filtering is currently evolving. Today's techniques are tedious and have operation limitations in the way ISPs maintain and enforce the permit/deny for community attributes that pass through their networks.

Cascade Failures

BGP's Prefix Updates have two types of attributes – transitive and non-transitive. The non-transitive attributes are not passed beyond a neighboring ISP. The transitive attributes are passed to the entire Internet. Hence, based on some BGP bugs and theoretical experiments, it is feasible to have a cascade failure of BGP triggered by some transitive attribute. While not proven, it does merit systematic investigation to determine the validity and risk. The consequences of one BGP update causing a cascade of BGP RIB failures warrants the investigation.

Counter Measures: Since no known vulnerability or exploit has been identified, it is not know if existing tools would be a counter measure to this type of attack vector.

GUARDED TRUST, MUTUAL SUSPICION, AND BGP SECURITY

BGP Peering assumes that something could go wrong with the policy filters between the neighboring routers. Filters are all created to mutually reinforce each other. If one policy filter fails, the policy filter on the neighboring router will take over – providing redundancy to the policy filters. This mutually reinforcement concept used BGP peering filters are created are also called guarded trust, mutual suspicion, or Murphy Filtering.[13]

For example,

- *ISP A* trust *ISP B* to send X prefixes from the Global Internet Route Table.
- *ISP B* creates an egress filter to insure only X prefixes are sent to *ISP A*.
- *ISP A* creates a mirror image ingress filter to insure *ISP B* only sends X prefixes.
- *ISP A's* ingress filter reinforces *ISP B's* egress filter.

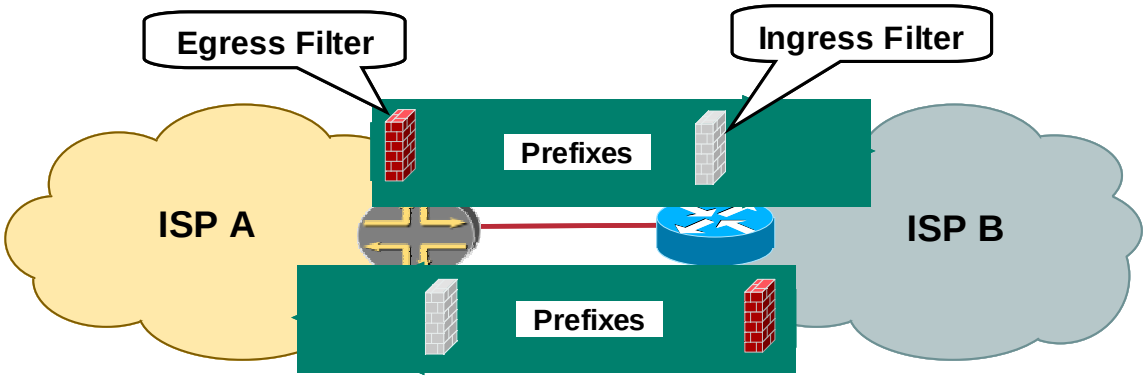

Figure 1 - BGP Peering and Reinforcing Prefix Filtering

[13] After Murphy's Law – where in Murphy's Law of Networking, anything that can go wrong, will go wrong at your most mission critical time. Hence back-ups, redundancy, and fail over mechanisms are a plus.

This approach allows ISPs to protect themselves from the failures in one of their peers. It is a widely accepted security strategy that is a Best Common Practice (BCP) concept in ISP Peering connections. ISPs who mindful adopt this approach mitigate the threat to themselves and help to compartmentizes problems to a few BGP Autonomous Systems.

TERMS AND DEFINITIONS

BGP Identifier - A 4-byte unsigned integer that indicates the sender's ID. In Cisco's implementation, this is usually the router ID (RID), which is calculated as the highest IP address on the router or the highest loopback address at BGP session startup. (Loopback address is Cisco's representation of the IP address of a virtual software interface that is considered to be up at all times, irrespective of the state of any physical interface.)

ACKNOWLEDGEMENTS

Thanks to all the people who have helped peer review, edit, and add to this work:

Rob Thomas [robt@cymru.com]
Daniel P (Dan) Koller [dpkoller@lucent.com]
Stephen Kent [kent@bbn.com]
Ross Callon [rcallon@juniper.net]
Russ White [ruwhite@cisco.com]
Alvaro Retana [aretana@cisco.com]
John G. Scudder [jgs@cisco.com]
Barry Friedman [friedman@cisco.com]
Anantha Ramaiah [ananth@cisco.com]
Satish Mynam [mynam@cisco.com]
Chris M. Lonvick [clonvick@cisco.com]
Paul Donner [pdonner@cisco.com]