Authors: Elizabeth Kaufman elkaufman@pollere.com
Dylan Greene dylan@juniper.net
Pete Moyer pete@juniper.net
Dave Stine dstine@pollere.com

# Mission-Critical BGP (MC-BGP)
# Defensive Inter-Domain Routing for Mission Critical Networks
# Draft Version 7

## 1. Document Status:

This document is a draft overview and specification intended to solicit comments and recommendations. The MC-BGP protocol remains under development.

## 2. Introduction

### 2.1. Industry Context and Problem Statement:

Inter-domain routing for IP internetworks is subject to a range of attacks. These attacks have been documented in many places, both as generic threats and with the specific target of BGP-4. Many have been implemented in software that is freely available worldwide for download. These routing-oriented attacks divide roughly into three classes:

**2.1.1.** Those that exploit inherent design vulnerabilities in TCP/IP;

**2.1.2.** Those that exploit common implementation problems in TCP/IP stacks or BGP-4 routing processes, and

**2.1.3.** Those that exercise innate weakness of BGP-4, in particular that it will accept and install any update that claims to offer a more optimal route for any network, without reference to the source of the advertisement. One can add to this the additional point that BGP has no mechanism to detect or terminate a Byzantine peer (a peer that behaves pathologically but does not fail completely).

Most serious proposals to enhance or replace BGP focus on the third class of problem, with varying assumptions about the security and stability of the available infrastructure. MC-BGP assumes an environment where local (intra-AS) device insertion is exceptionally difficult, and the AS topology (including prefix origination and AS_PATH) may be highly volatile in ways that cannot be predicted or characterized, even provisionally, in advance of a topology change.

.

## 2.2. What is MC-BGP?

MC-BGP is an extension of BGP-4 designed to provide non-cryptographic, policy stabilized inter-domain routing for Mission Assurance Category I (MAC I) and other high availability networks that must support significant unpredictable mobility. It uses an optional, transitive (Inter-AS) Path Attribute to attach policy criteria to high-value routes, making them difficult to over-write dynamically, provided the route remains valid (i.e., next hop is available and it is not withdrawn). MC-BGP can utilize partial or complete data validation repositories, but does not require their presence, and will not fail if it loses contact with a data validation overlay. It allows limited path preference selection (to provide, among other things, options for future Quality of Protection (QoP) support), and offers a dynamic, configurable mechanism to respond to badly-behaved peers. It is agnostic with regard to authentication method and infrastructure confidentiality and integrity services (encryption). It is fully interoperable with RFC standard BGP-4, and can function usefully in incremental deployments. MC-BGP does not perform AS_PATH validation, and does not require (but will not prevent) external AS-to-prefix or AS-to-Origin validation.

### 2.2.1. Mission Assurance Category I and Inter-domain Routing:

Department of Defense Instruction 8500.2 defines MAC I as: "Systems handling information that is determined to be vital to the operation readiness or mission effectiveness of deployed and contingency forms in terms of both content and timeliness. The consequences of loss of integrity or availability of a MAC I system are unacceptable and could include the immediate and sustained loss of mission effectiveness. Mission Assurance Category I systems require the most stringent protection measures." How this definition maps specifically to the accreditation of routing protocols and routing design is a discussion outside the scope of this document. The authors concluded, nevertheless, that MAC I networks require, above all, inter-domain routing that is resistant to external interference with MAC I mission support, that prioritizes convergence under all conditions, and maintains availability of user data transit services (packet forwarding) for high priority networks even under conditions of severe disruption.

## 2.3. Comparison with other BGP Alternatives:

Multiple proposals exist to enhance or replace BGP. These include S-BGP, soBGP, PSBGP and PGBGP. The first three options focus primarily on ensuring that routing information is valid and cryptographically assured, with varying models and degrees of flexibility. PGBGP focuses on a problem space similar to that of MC-BGP, which is the issue of how to handle a potentially suspicious route, but it does not place the same priority as MC-BGP on rapid convergence, and does not specifically address the needs of MAC I mission support.

**Table A: Partial Comparison of BGP Alternatives**

OCW is Outside the Convergence Window; WCW is Within the Convergence WIndow

| | **S-BGP** | **soBGP** | **psBGP** | **PGBGP** | **MC-BGP** |
|---|---|---|---|---|---|
| **BGP4 incremental deployment** | Possibly contiguous Inter-AS; | Designed to do so but some analysis that suggests issues. | Possibly inter-AS | Yes | Yes |

| | | | | | |
|---|---|---|---|---|---|
| **Router Authentication** | IPsec | Signed EntityCert | Integrated; AS cert | N/A | Uses Infrastructure; PKI or shared secret or other. |
| **Peering Integrity** | IPsec/null | N/A | IPsec | N/A | Uses Infrastructure; IPsec,<br>L2 crypto or other |
| **Peering Confidentiality** | Not required | N/A | IPsec | N/A | Uses Infrastructure |
| **UPDATE validation** | Central PKI<br>WCW (or not) | PKI<br>WCW/OCW | PKI<br>WCW | Anomaly-based<br>Shunt to Operator<br>WCW | Via local route master<br>OCW |
| **AS-prefix binding** | Central PKI/Attestation<br>WCW (or not) | PKI<br>WCW/OCW | AS PKI+ inter-AS cross-assertion | Statistical or Registry<br>WCW | Via local route master<br>OCW |
| **AS-origin binding** | Central PKI/Attestation<br>WCW | PKI/Attestations<br>WCW/OCW | AS PKI + inter-AS cross-assertion | Registry<br>WCW | Via local route master<br>OCW |
| **AS_PATH validation** | Central PKI/sig. chain<br>WCW (or not) | PKI, Attestations<br>WCW/OCW | PKI path check<br>WCW | ? | Via local route master<br>OCW |
| **Trust model** | Hierarchical centralized | Web of trust or Hierarchy | Hierarchical/ centralized | AS model | AS model |
| **Policy-based Route Persistence** | N/A | Cryptographically hardened (originator) | Cryptographically hardened | Operator Intervention | Yes |
| **AS trust adjustments** | No | No | No | Operator Intervention | Yes |
| **Byzantine peer mitigation** | Operator/Authorizer Revocation? | Operator Revocation? | Operator Revocation? | Operator Intervention | Dynamic |

## 2.4. Comparison with BGP4+Best Common Practices (BCP):

A partial interpretation of MC-BGP is that it offers a dynamic mechanism to circulate policy that is currently handled through a mix of intransitive attributes (local_pref), partially transitive attributes (MED), proprietary knobs (weight), and lots and lots of filters. This description isn't completely inaccurate, but MC-BGP does offer some capabilities that extend beyond BGP4+BCP. These include:

**2.4.1.** Ability to apply sensible policy to a large population of unpredictably mobile networks (full ASes and wandering prefixes), especially where full

AS and/or prefix information cannot be known in advance.

**2.4.2.** Ability to converge dynamically and correctly in conditions of core segmentation, with potential mis-configuration of edge peers (for example, to correctly select previously unfavorable exit routes to external networks).

**2.4.3.** Ability dynamically to communicate/ negotiate routing policy at inter-organizational boundaries via a simple, standard set of policy semantics.

**2.4.4.** Ability to simplify policy configuration by configuring it once and authoritatively either at the point of origin or via a data validation route injection mechanism (Route Master).

**2.4.5.** Ability dynamically to respond to Byzantine peers.

Note that there is no mechanism yet proposed that eliminates the need for significant and careful configuration of large routing infrastructures; MC-BGP does not claim to offer any magic in this area.

## 3. MC-BGP Technical Overview:

### 3.1. Design Summary:

MC-BGP introduces a new, optional transitive Route Resilience Path Attribute which carries policy values on a per-route basis; the Route Resilience attribute is used in the BGP decision tree to control whether an installed route will be replaced when a new route is received via an UPDATE without a prior timeout or withdraw. It can be used to make a known good route highly “sticky,” or, in some cases, to reduce hold-down, permit MOAS, or manage deaggregation of installed prefixes. MC-BGP also introduces a mechanism to respond dynamically to Byzantine peers, and to enable policy-based path preference (for example, in support of QoP).

### 3.2. Design Principles:

**3.2.1. Deployability:** MC-BGP is not an academic project; it is intended for actual use to provide policy-stabilized inter-domain routing on mission critical IP networks. In its design, we solicited input from network operations experts as well as those with experience in protocol design. We continue to solicit relevant review.

**3.2.2. Verifiability:** MC-BGP was designed with an eye to providing meaningful, measurable operational metrics that could be analyzed by an external statistical or policy based routing analysis package. While it does not require a heavy management infrastructure, it is designed to integrate well in environments that must certify baseline network functionality rigorously and in real-time.

**3.2.3. Re-use:** In the design of MC-BGP, we attempted to re-use, as much as possible, mechanisms within BGP-4 that have been proven in large and diverse operational networks. Specific principles and/or mechanisms include:

**3.2.3.1.Autonomous System (AS) Sovereignty:** MC-BGP utilizes transitive (inter-AS) mechanisms to provide policy-stabilization for routing updates, but these mechanisms are subject to administrative over-ride or augmentation in routing policy. Inter-domain routing is based on the fundamental concept of ASes that are administratively distinct, sovereign and internally self-sufficient with respect to routing policy. Although there are many blurred edges in deployment (and MC-BGP permits this blurring), MC-BGP policy attributes are designed to scale along the AS model, and to be exchanged as appropriate on an inter-domain basis. The installation and re-distribution of those attributes are subject to peering agreements, and may be instantiated in routing policy (configuration), as is true with all standard, transitive BGP-4 attributes.

**3.2.3.2.Path Attributes:** Path Attributes are a widely-used feature of BGP4; both commercial and private implementers have significant experience optimizing BGP decision trees to process them within the convergence window. This mechanism is thus preferable to other implementation options (such as idiosyncratic coding of communities) that lack a track record for deployment and optimization.

**3.2.3.3. Peer Authentication:** BGP-4 is agile with respect to peer authentication, and does not require support for any specific mechanism. MC-BGP is similarly able to use any (or no) authentication method supported by the relevant platforms. While authentication methods are important and require significant operational review, they change much more rapidly than does a routing protocol; it seems inappropriate to specify specific authentication technologies within a routing specification. That choice is left to the implementer and to the network operator.

**3.2.3.4.Capabilities Exchange for MC-BGP Peer Identification:** MC-BGP peers identify one another during the initial peering set-up via the standard BGP-4 capabilities exchange. This approach is to ensure seamless interoperability with BGP speakers.

**3.2.3.5. BGP-4 Framework and Mechanisms:** MC-BGP extends rather than replaces RFC 4271. Where an MC-BGP Infrastructure Index is not set, MC-BGP behaves like BGP-4.

## 4. MC-BGP Technical Specification:

### 4.1. Peer Authentication:

**4.1. Peer Authentication:** MC-BGP is agnostic with regard to peer authentication technology. It can use any available mechanism to authenticate, including (but not limited to) pre-shared secrets, digital certificates, or simple passwords. Like other proposals, MC-BGP presumes the fallback availability of IPsec, but does not require it, and can easily deploy over L2 encrypted infrastructures, or other mechanisms. MC-BGP employs a functional legitimacy model that enables it to trigger alarms or tear down peering with a neighbor that, at any point in the peering session, becomes Byzantine or appears to behave badly. MC-BGP routers may be configured to peer promiscuously.

**4.2. Set-up and Identification:** MC-BGP routers identify one another using the Capabilities Optional parameter during initial peering setup as specified in RFC 3392. A router that identifies itself as an MC-BGP router must support the MC-BGP Path Attribute as described below. MC-BGP routers may be configured to peer only with routers that identify themselves as MC-BGP capable. We anticipate, however, that most MC-BGP deployments will be incremental, and MC-BGP routers must be configurable to peer with BGP-4 speakers. MC_BGP routers support 3 options in the peering setup:

### 4.2.1. MC-BGP router (mandatory for MC-BGP):

Identifies the router as an MC-BGP speaker.

### 4.2.2. MC-BGP Route Master (optional):

Identifies a peer as presenting externally validated (hence highly resilient) routes; it is optional for a non-Route Master to accept this identification from a peer, and should be subject to configuration and, if the operator chooses, additional authentication.

### 4.2.3. MC-BGP mapping (optional):

MC-BGP routers may participate in a logical graphing activity, where they share logical adjacency data with each other in support of map aggregation for external network management/monitoring functions. This is optional, and can be unilateral or bi-lateral in a peering session. As described below, it occurs on an as-available basis, and must not interfere with packet forwarding or routing table convergence. (Note that his functionality still requires lab benchmarking to make recommendations on deployment and potential associated overhead.)

**4.3. MC-BGP Route Resilience Path Attribute:** MC-BGP introduces a new transitive Route Resilience Path Attribute, mandatory for MC-BGP speakers, marked as optional for purposes of interoperability with BGP-4 speakers, who should ignore and re-advertise it without modification. MC-BGP routers must evaluate the MC-BGP Path Attribute first of policy configuration, prior even to AS path length (but after verifying next hop, etc.). For purposes of rapid convergence, the Infrastructure Index should be evaluated first; if that value is zero, the router should exit analysis of the MC-BGP Path Attribute and continue to its normal BGP-4 decision tree even if other values of the Route Resilience Attribute are set.

**4.3.1.** MC-BGP routers must attach the MC-BGP Path Attribute to routes that they originate, and must retain that attribute when received from another MC-BGP speaker. Based on routing policy, MC-BGP routers may augment, decrement, or zero the Infrastructure Index received inter-AS, based on existing peering agreements and/or individual AS policy. (See the Administrative Trust Policy described below.) MC-BGP routers may be configured to attach the MC-BGP Path Attribute to a route received from a BGP-4 speaker, where administratively reasonable to do so.

**4.3.2.** The Route Resilience Path Attribute is an aggregate of the following values:

**4.3.2.1. Infrastructure Index:** (4 bits) a numeric value used to characterize critical routes that are highly stable and highly available (such as backbone routes); routes that should be selected by a normal BGP-4 decision tree must be set to an Infrastructure Index value of 0. The number of bits is primarily for future use, or potentially to enable granular statistical analysis of route behavior by external engines; it seems unlikely that any single AS will need to exercise the full range of the Index (though important routes will need the highest possible rating).

**4.3.2.2. Mobility Indicator:** (1 bit) identifies a mobile network (a network that may experience rapid changes of origin or AS Path, or both). This bit could be used to reduce hold down on highly mobile routes.

**4.3.2.3. MOAS Permitted:** (1 bit) this bit indicates whether a network is administratively permitted to appear as Multi-Origin. MC-BGP only permits MOAS for MC-BGP routes if this bit is set.

**4.3.2.4. Path Quality**: (4 bits) a rating of path desirability based on administrative considerations such as cost, security, etc.. This metric can be used to configure MC-BGP to select certain routes based on known attributes of the path. For example, if it is administratively preferable to prefer a terrestrial route to a space route, or an encrypted link to an unencrypted one, MC-BGP can be configured to assess these values as part of its path selection process. This metric can be used in support of QoP, where applicable.

## 4.4. MC-BGP Data Validation Model:

MC-BGP does not require any eternal mechanism for full or partial route validation, but it can use and prioritize validated routes if/as they are available. Where data validation mechanisms are present, MC-BGP allows injection of highly resilient routes via a “Route Master” router.

(Note: **this is a non-cryptographic model** that utilizes route resilience to inject highly authoritative routes. Discussions continue as to whether there are lightweight mechanisms (such as Whisper) that may provide higher assurance at the point of route injection.)

### 4.4.1. MC-BGP Route Master:

The MC-BGP Route Master is a router that may send UPDATES for any route where the high-order Infrastructure Index bits are set. Route Masters may be, but probably ought not to be, a next hop router for some or all of those routes. From a design perspective, at least for MAC I networks, it seems likely that a Route Master should not support packet forwarding of user transit data, but the protocol is agnostic with regard to this issue.

#### 4.4.1.1. Route Master Validation Repository:

The MC-BGP Route Master learns its routes from a repository of validated routes. That repository may be a routing registry, part of a CA hierarchy, etc.. The Route Master may re-advertise dynamically learned routes back to the repository for operational review.

#### 4.4.1.2. Route Master Peering Design:

To function, the Route Master must peer with one or more MC-BGP peers on the network, but it need not peer with all of them. Its routes are sufficiently resilient that they will propagate normally across the routed infrastructure. Route Masters are statically configured, if used, and can be deployed many ways, provided they don't inject conflicting routes within a single AS.

#### 4.4.1.3. Route Master Failure:

Route Master routes will persist on the network until/unless they are withdrawn, or the next hop router becomes unavailable. If the Route Master fails, its advertised routes will be unaffected. Provided the Route Master is not a next hop router, packet forwarding will continue as usual on the available paths.

### 4.5. MC-BGP Optional Logical Mapping:

MC-BGP routers may negotiate unilateral or bilateral logical map exchanges during their initial capabilities exchange. If mapping is enabled for a given peering session, the mapping peer(s) will periodically (depending on map changes and cycle availability) transmit a table of known adjacent neighbors with their local addresses. Mapping is an auxiliary function designed to aid in management and mapping of heavily encrypted infrastructure, where current logical mapping data may be otherwise difficult to maintain; mapping data does not function in the routing decision process.

**4.6. Multi-path and Tie-Breaking with Route Resilience:** MC-BGP permits multi-path configurations for MC-BGP routes with equal route resilience unless forbidden by policy (configuration). Where multi-path is disabled for a route, equal resilience routes are subject to the BGP-4 decision tree and, if necessary, standard BGP-4 tie-breaking mechanisms.

**4.7. De-aggregation:** MC-BGP permits deaggregation unless prohibited by policy (configuration). In order for an MC-BGP router to install a more specific route to a highly resilient route, the more specific route must have an equal or greater resilience than the original aggregate. Note that Route Resilience must be applied with care to external (outside the AS) aggregates, since effective deaggregation will need to be coordinated to converge routing correctly.

### 4.8. Byzantine Peers:

A Byzantine peer is a router that functions improperly (possibly due to hijacking or mis-configuration) but continues to function well enough to maintain a peering session and send false or malformed UPDATE messages. MC-BGP routers must support a range of options to respond to Byzantine peers. Specifically, where an MC-BGP router receives an update for an existing installed high Infrastructure Index route (for which it has not received a proper withdraw) it must be able to take one or more of the following actions:

**4.8.1.** Generate an alarm;

**4.8.2.** Maintain the peering session and ignore the bad route;

**4.8.3.** Tear down the peering session but maintain relevant routes learned from the now-Byzantine peer, where that peer is not the next hop. (Note that this response is useful only in relatively rare and odd circumstances.)

**4.8.4.** Tear down the peering session and purge routes learned from or that transit that peer.

### 4.9. Administrative Trust Policy:

Route Resilience is a transitive attribute, but no AS is required to install or redistribute another ASes Infrastructure Index values, unless they commit to do so via a peering agreement, SLA, or other agreement. MC-BGP routers must support an Administrative Trust Policy that allows a router to increase, decrement or zero the Infrastructure Index on a received route, on a per-peer, per-AS, or per-origin basis. The Administrative Trust Policy acts only on Infrastructure Index value, not on other aspects of the Route Resilience Attribute.

### 4.10. MC-BGP Deployment Considerations:

Note that MC-BGP was not designed to solve every known problem with inter-domain routing, nor will it provide an independent hardened routing capability over a completely unsecured network. In addition, as with any new mechanism, MC-BGP may force re-consideration of other routing configuration strategies in use on an existing BGP-4 network. The following is a partial list of known interactions.

### 4.10.1. Local_Pref:

Local_Pref is a non-transitive (intra-AS only) BGP Path Attribute that expresses administrative preference for one path over another. MC-BGP uses a transitive Path Attribute that subsumes and expands on the capabilities of Local_Pref; MC-BGP routers thus ignore Local_Pref.

### 4.10.2. Route Weighting and Other Vendor Proprietary “Knobs”:

Many of the features of MC-BGP have been partially implemented by various vendors using proprietary “knobs” and configuration-driven route weighting. MC-BGP is a protocol-based mechanism that over-rides these knobs, and must inter-operate predictably in a multi-vendor environment.

### 4.10.3. Path Selection:

A successful routing process is designed to select a good (i.e. usable) path to any reachable network and, where more than one route may exist, to apply some selection criteria towards identifying which path might be optimal, where optimal is a function of policy applied to known path characteristics. MC-BGP is designed for networks that must optimize for high-availability. BGP-4 typically optimizes for the shortest AS path (which may not be and often is not the shortest route in hops). Any adjustment to account for policy criteria (such as path preference) may result in a less optimal route, where optimal is defined purely as shortest. Potential impact on round-trip times may factor into the decision to apply path preference to any given route.

#### 4.10.4. Infrastructure Index Values and Loop Free Topologies:

In general, inter-domain routing protocols perform well at preventing and/or eliminating routing loops. Any mechanism that inserts static or virtually static routes (such as high Infrastructure Index values) needs to be analyzed to avoid creating the potential for routing loops, in particular along inconsistently-administered AS boundaries.