

A Complete Solution for iBGP Stability

Ravi Musunuri Jorge A. Cobb
 Department of Computer Science
 The University of Texas at Dallas
 Richardson, TX-75083-0688
 Email: {musunuri,cobb}@utdallas.edu

Abstract—BGP routers within an Autonomous System (AS) exchange their inter-AS routing information via the internal Border Gateway Protocol (iBGP). Within an AS, every BGP router needs to maintain an iBGP peering session with every border BGP router. This peering scheme fails to scale due to the large number of iBGP peering sessions required. Current solutions to this scalability limitation divide the AS into clusters, with a route reflector acting as a representative of the cluster. Clustering, however, introduces routing anomalies. Furthermore, BGP also uses a Multi-Exit Discriminator (MED) to differentiate multiple links connecting the same pair of AS'ns. However, clustering in combination with the use of MED values further aggravates routing anomalies. In this paper, we propose a complete solution that solves both clustering induced anomalies and MED induced anomalies in iBGP. Our solution requires multiple path disseminations between route reflectors and selective single-path dissemination from each route reflector to its client.

I. INTRODUCTION

The Internet, at its highest level, is divided into administrative domains, commonly known as Autonomous Systems (AS'ns). The Border Gateway Protocol (BGP) [1] is the de-facto protocol for sharing inter-AS routing information between neighboring BGP routers. Neighboring BGP routers in different AS'ns share their inter-AS routing information via the external Border Gateway Protocol (eBGP). On the other hand, any two BGP routers in the same AS, even if they are not physically neighbors, share their inter-AS routing information via the internal Border Gateway Protocol (iBGP).

BGP routers reliably exchange the routing information with each other via peering sessions. A peering session between two routers in different AS'ns is known as an eBGP peering session, and a peering session between two routers within the same AS is known as an iBGP peering session.

Both eBGP and iBGP have been plagued with forwarding and divergence anomalies. Forwarding anomalies consist of permanent loops in the routing tables, while divergence anomalies prevent the routers from converging to a stable selection of paths. eBGP suffers mainly from divergence anomalies. eBGP divergence anomaly has been studied extensively, and, given that it is outside the scope of this paper, the reader is referred to [3], [4], [5], [6] for a discussion of the problem and proposed solutions.

iBGP, on the other hand, suffers from both divergence and forwarding anomalies. Two features of iBGP are the cause for these anomalies. First, iBGP employs Route-Reflection Clustering [2] to improve the scalability of number of iBGP peering sessions required. Although scalability is improved,

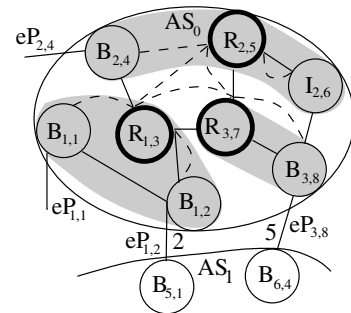


Fig. 1. Autonomous system

this technique has caused forwarding and divergence anomalies [7], [8], [9], [10]. Second, a Multi Exit Discriminator (MED) is a path value used to establish a preference among multiple links connecting the same pair of AS'ns. The MED value, in combination with clustering, may cause divergence anomalies [11], [12], [14]. In Section IV, we will explain each iBGP anomaly in detail.

In this paper we propose a complete Stable iBGP (S-iBGP) protocol that solves both forwarding and divergence anomalies. Our solution is more scalable than previous solutions and is also proven correct. Due to space limitations, we are not providing the proof of our solution in this paper. Our solution requires two minor changes to the iBGP messages. However, these required changes are internal to the AS. Therefore, they can be implemented locally inside an AS without disrupting the behavior of other AS'ns. Anomalies are resolved via dissemination of multiple paths between reflectors, and via selective dissemination of single path between reflectors and their clients.

II. NETWORK MODEL AND NOTATION

An AS is a graph consisting of a set of nodes V , that denotes BGP routers, and two sets of edges, E and E_p , that denote physical links between routers and peering sessions between routers, respectively. Edges in E are depicted as solid lines, while edges in E_p are depicted as dotted lines.

Set V is divided into m disjoint sub-sets, known as *clusters*, which are depicted as shaded regions. Each cluster i has a special router, known as *route-reflector* R_i . Route reflectors are depicted as bold nodes. Other routers in cluster i are known as *clients* of R_i . The client routers set in cluster i is denoted by C_i . Each node has a label of the form $X_{i,j}$, where X is one of R, B, I , that indicates if the node is a reflector, border, or interior router, respectively. The subscripts i and j denote the

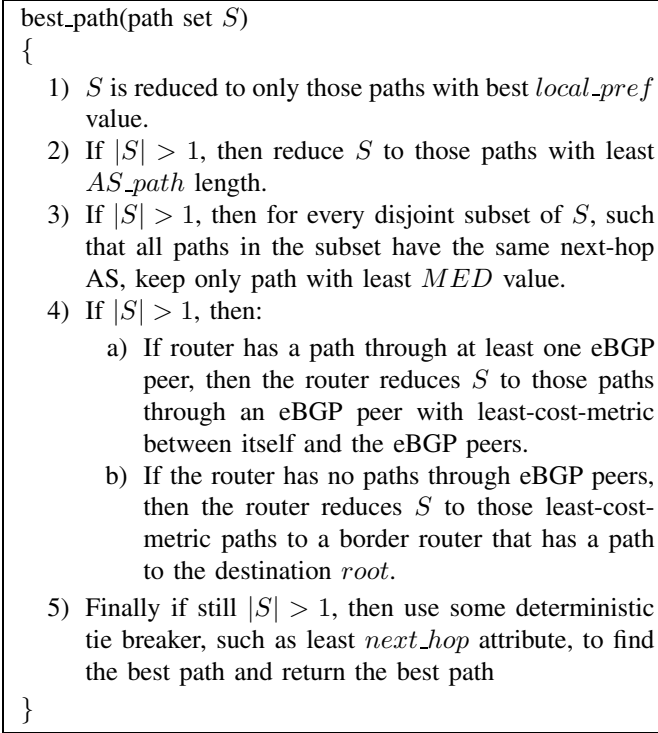


Fig. 2. Best Path Selection Algorithm

cluster number (*not* the AS number) and the router number within the AS, respectively.

Each reflector $R_{i,j}$ maintains a peering session with each reflector, $R_{m,n}$, $m \neq i$, of every other cluster m within its AS. Also, $R_{i,j}$ maintains a peering session with each of its clients, i.e., with each router in C_i . If there exists a physical link between two routers $X_{i,j}$ and $X_{m,n}$, then $(X_{i,j}, X_{m,n}) \in E$. Similarly, if there exists a peering session between $X_{i,j}$ and $X_{m,n}$, then $(X_{i,j}, X_{m,n}) \in E_p$.

Figure 1 shows an autonomous system, AS_0 , that has been divided into three clusters: cluster one contains routers $\{B_{1,1}, B_{1,2}, R_{1,3}\}$, cluster two contains routers $\{B_{2,4}, R_{2,5}, I_{2,6}\}$, and cluster three contains routers $\{R_{3,7}, B_{3,8}\}$. Each physical link is assigned a cost (e.g., the typical cost in the intra-domain routing protocol) and each inter-AS link is assigned an integer MED value. The cost of the minimum-cost path between two routers $X_{i,j}$ and $X_{m,n}$ is denoted by $sp(X_{i,j}, X_{m,n})$.

Without loss of generality, we assume a single destination AS, that we call *root*. Thus, every router, $X_{i,j}$, attempts to find a best path, $best_{i,j}$, to reach *root*. The end-to-end path between $X_{i,j}$ and *root* consists of two sub-paths. The first sub-path corresponds to the *internal Path*, $iP_{i,j}$, which is the shortest path from interior router $X_{i,j}$ to some border router $X_{m,n}$ within the same AS. The second sub-path corresponds to the *external Path*, $eP_{i,j}$, from the above border router $X_{m,n}$ to *root*. In particular, note that $eP_{i,j} = eP_{m,n}$. Path $eP_{m,n}$ has the following attributes.

- $eP_{m,n}(local_pref)$: Local preference value.
- $eP_{m,n}(AS_path)$: Sequence of AS'ms along the path from $X_{m,n}$ to *root*.

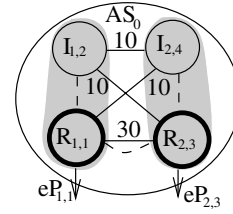


Fig. 3. CIF Anomaly.

- $eP_{m,n}(next_hop)$: IP address of next-hop BGP router of $X_{m,n}$.
- $eP_{m,n}(MED)$: Integer that ranks multiple links between the same pair of AS'ms. A smaller value implies a greater preference for $eP_{m,n}$.

Each router, $X_{i,j}$, including reflectors, maintains a set of available external paths, $AP_{i,j}$, that are advertised by all its peer routers. Additionally, each reflector $R_{m,n}$ also maintains a set of cluster's available paths, $CAP_{m,n}$, that are advertised by its peers in its own cluster. In Figure 1, $CAP_{1,3} = \{eP_{1,1}, eP_{1,2}\}$, where as, $AP_{1,3}$ might be $\{eP_{1,1}, eP_{1,2}, eP_{2,4}, eP_{3,8}\}$.

Each router $X_{i,j}$ finds its best external path, $best_{i,j}$, by calling the $best_path(AP_{i,j})$ routine [15], which is shown in Fig. 2.

Next, for a given reflector router, we introduce the notions of the *Feasible Paths Set* and *Cluster's Feasible Paths Set*, both of which are necessary in our protocol.

Definition 1: The *feasible paths set*, $FP_{i,j}$, at reflector $R_{i,j}$ is obtained by applying the first three steps in the $best_path$ routine with input set $AP_{i,j}$. A border router traversed by a feasible path is called a *feasible border router*. Similarly, the *cluster's feasible paths set*, $CFP_{i,j}$, at reflector $R_{i,j}$ is obtained by applying the first three steps in the $best_path$ routine with input set $CAP_{i,j}$.

III. ROUTE-REFLECTION CLUSTERING

Each router, $X_{i,j}$, advertises its best path, $best_{i,j}$, by using the *update* message [1] and withdraws the previously advertised infeasible paths. In the original route-reflection clustering protocol, best paths are advertised as explained below:

- Each border router, $B_{i,k}$, advertises its $best_{i,k}$ to its reflector $R_{i,j}$.
- At reflector $R_{i,j}$, if $best_{i,j} \in CAP_{i,j}$ then $R_{i,j}$ received its $best_{i,j}$ from some client router $X_{i,m} \in C_i$. In this case, $R_{i,j}$ advertises $best_{i,j}$ to others as follows:
 - $R_{i,j}$ advertises to all other reflectors, $R_{m,n}$, $m \neq i$.
 - $R_{i,j}$ advertises to all routers in C_i , except to the router $X_{i,m}$.
- if $best_{i,j} \notin CAP_{i,j}$ then $R_{i,j}$ advertises $best_{i,j}$ only to its client routers, i.e., to routers in C_i .

IV. IBGP ANOMALIES

iBGP has been plagued with numerous routing anomalies. We can divide iBGP anomalies into the *Clustering Induced (CI) Anomalies* and *Clustering and MED Induced (CMI) Anomalies*. Next, we explain these anomalies via examples. In these examples, we assume all external paths at border

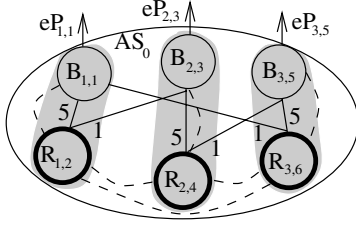


Fig. 4. CID Anomaly.

routers have the same *local_pref* value and *AS_path* length. Hence, each border router $B_{i,j}$ always chooses $eP_{i,j}$, if there exists one, as $best_{i,j}$, according to step 4(a) in the algorithm of Fig. 2.

A. Clustering Induced (CI) Anomalies

CI anomalies occur due to clustering of routers, even if we ignore MED values. CI anomalies [7], [8] cause *Forwarding Anomalies* (routing loops) and *Divergence Anomalies* (failure to achieve a stable configuration).

1) *CI Forwarding (CIF) Anomaly*: Figure 3 [9] shows an example of a CIF anomaly. $I_{1,2}$ and $I_{2,4}$ learn their external path, $eP_{1,1}$ and $eP_{2,3}$, respectively, from their reflector. Thus, $I_{1,2}$ routes its data messages through $R_{1,1}$, and $I_{2,4}$ routes its data messages through $R_{2,3}$. However, due to the costs assigned to each link, the internal paths are as follows: $iP_{1,2}$ is $I_{1,2} \rightarrow I_{2,4} \rightarrow R_{1,1}$, and $iP_{2,4}$ is $I_{2,4} \rightarrow I_{1,2} \rightarrow R_{2,3}$. Hence, there is a forwarding loop between $I_{1,2}$ and $I_{2,4}$.

2) *CI Divergence (CID) Anomaly*: An example of a divergence anomaly [7] is shown in Fig.4. Each $R_{i,j}$ always prefers $eP_{i+1,j+1}$ ¹ over $eP_{i,j-1}$ due to following:

$$sp(R_{i,j}, B_{i,j-1}) > sp(R_{i,j}, B_{i+1,j+1}). \quad (1)$$

The following explains the divergence steps.

- 1) Lets assume $AP_{1,2} = \{eP_{1,1}, eP_{2,3}\}$, $AP_{2,4} = \{eP_{2,3}\}$ and $AP_{3,6} = \{eP_{3,5}\}$. Hence, $best_{1,2} = eP_{2,3}$ due to condition 1, $best_{2,4} = eP_{2,3}$ and $best_{3,6} = eP_{3,5}$.
- 2) Next, if $R_{2,4}$ receives an update message from $R_{3,6}$ with $eP_{3,5}$, then $AP_{2,4} = \{eP_{2,3}, eP_{3,5}\}$ and $best_{2,4} = eP_{3,5}$ due to condition 1. Also, $R_{2,4}$ withdraws $eP_{2,3}$, and thus $R_{1,2}$ updates its $best_{1,2}$ to $eP_{1,1}$.
- 3) Next, if $R_{3,6}$ receives an update message from $R_{1,2}$ with $eP_{1,1}$, then $AP_{3,6} = \{eP_{1,1}, eP_{3,5}\}$ and $best_{3,6} = eP_{1,1}$ due to condition 1. Also, $R_{3,6}$ withdraws $eP_{3,5}$, and thus $R_{2,4}$ updates its $best_{2,4}$ to $eP_{2,3}$.
- 4) Next, if $R_{1,2}$ receives an update message from $R_{2,4}$ with $eP_{2,3}$, then $AP_{1,2} = \{eP_{1,1}, eP_{2,3}\}$ and $best_{1,2} = eP_{2,3}$ due to condition 1. Also $R_{1,2}$ withdraws $eP_{1,1}$, and thus $R_{3,6}$ updates its $best_{3,6}$ to $eP_{3,5}$.

The above cyclic exchange of path update messages may continue indefinitely, and thus, a stable set of best paths may never be achieved.

B. Clustering and MED Induced (CMI) Anomalies

CMI anomalies occur due to combination of clustering and MED values [11], [12], [14]. Similar to CI anomalies, CMI anomalies also cause route divergence.

¹mod 3 is applied to subscript i and mod 6 is applied to subscript j

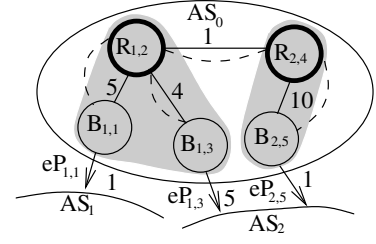


Fig. 5. CMID Anomaly.

1) *CMI Divergence (CMID) Anomaly*: The divergence example [11] is shown in Fig. 5. The steps in the divergence are as follows.

- Let us assume $AP_{1,2} = \{eP_{1,1}, eP_{1,3}\}$ and $AP_{2,4} = \{eP_{2,5}\}$. Thus, $best_{2,4} = eP_{2,5}$, and $best_{1,2} = eP_{1,3}$ [$sp(R_{1,2}, B_{1,1}) > sp(R_{1,2}, B_{1,3})$].
- Next, if $R_{1,2}$ receives an update message from $R_{2,4}$ with $eP_{2,5}$, then $AP_{1,2} = \{eP_{1,1}, eP_{1,3}, eP_{2,5}\}$. $R_{1,2}$ prefers $eP_{2,5}$ over $eP_{1,3}$ [$MED(eP_{2,5}) < MED(eP_{1,3})$], and $eP_{1,1}$ over $eP_{2,5}$ [$sp(R_{1,2}, B_{2,5}) > sp(R_{1,2}, B_{1,1})$]. Hence, $R_{1,2}$ changes its $best_{1,2}$ to $eP_{1,1}$.
- Next, if $R_{2,4}$ receives an update message from $R_{1,2}$ with $eP_{1,1}$, then $AP_{2,4} = \{eP_{1,1}, eP_{2,5}\}$. $R_{2,4}$ prefers $eP_{1,1}$ over $eP_{2,5}$ [$sp(R_{2,4}, B_{2,5}) > sp(R_{2,4}, B_{1,1})$]. Hence, $R_{2,4}$ changes its $best_{2,4}$ to $eP_{1,1}$, and also withdraws $eP_{2,5}$.
- Next, because $eP_{2,5}$ has been withdrawn, $AP_{1,2} = \{eP_{1,1}, eP_{1,3}\}$. $R_{1,2}$ prefers $eP_{1,3}$ over $eP_{1,1}$ [$sp(R_{2,5}, B_{1,1}) > sp(R_{2,5}, B_{1,3})$]. Hence, $R_{1,2}$ updates its $best_{1,2}$ to $eP_{1,3}$, and also withdraws $eP_{1,1}$.
- Next, because $eP_{1,1}$ has been withdrawn, $AP_{2,4} = \{eP_{2,5}\}$, and $R_{2,4}$ updates its $best_{2,4}$ to $eP_{2,5}$.

The above cyclic exchange of path update messages may continue indefinitely, and thus, a stable set of best paths may never be achieved.

V. STABLE IBGP

In this section, we present our protocol. Before doing so, we give a brief overview of our notation².

A. Pseudo-code Notation

A network consists of a set of BGP aware routers interconnected via peering sessions. We use two types of messages (*update*, *brpref*), and the peering session from $X_{i,j}$ to a peer router $X_{m,n}$ is denoted by $ps(X_{i,j}, X_{m,n})$.

The code of a router consists of a set of actions. An action is of the form: $\langle \text{guard} \rangle \rightarrow \langle \text{command} \rangle$. A receiving guard at router $X_{i,j}$ is of the form: **rcv** m **from** $X_{m,n}$, where $X_{m,n}$ is a peer of $X_{i,j}$. An action with this guard is enabled if and only if there is a message of type m in peering session $ps(X_{i,j}, X_{m,n})$. Furthermore, if this action is chosen for execution, then this message is removed from peering session $ps(X_{i,j}, X_{m,n})$. The $\langle \text{command} \rangle$ in an action is a sequence of assignment, iteration, and send statements.

²our notation is loosely based on guarded command notation in [16].

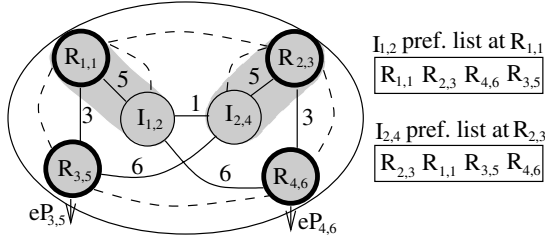


Fig. 6. Selective Path.

The parameter declared in a process is used to write a set of actions as a single action, with one action for each possible value of the parameter. For example, if we have the following parameter definition **par** $g:\{r,s\}$ then the action: **rcv update from** $g \rightarrow$ **send update to** $X_{i,j}$ is a shorthand notation for the two actions: **rcv update from** $r \rightarrow$ **send update to** $X_{i,j}$ and **rcv update from** $s \rightarrow$ **send update to** $X_{i,j}$.

B. Basic Idea

In this section, we explain the basic idea of our Stable iBGP (S-iBGP) protocol. Our protocol requires two changes from the original route-reflection protocol, as follows.

- Each $B_{i,k}$ advertises its $best_{i,k}$ to its reflector $R_{i,j}$. This is similar to the original protocol.
- **Multiple Path Dissemination (MPD) between Reflectors:** In our modified protocol, each $R_{i,j}$ is required to advertise multiple feasible paths, $CFP_{i,j}$, to other reflectors $R_{m,n}$, $m \neq i$. In the original protocol, reflectors used to advertise the single path $best_{i,j}$. In Figure 5, reflector $R_{1,2}$ advertises $CFP_{i,j}$ set, which is equal to $\{eP_{1,1}, eP_{1,3}\}$. MPD is the same idea used in [14], which requires a change in the format of iBGP *update* message between reflectors.
- **Selective Path Dissemination (SPD) between Reflectors and Clients:** In our modified protocol, we employ Selective Path Dissemination (SPD) from reflector $R_{i,j}$ to its client routers in C_i . Router $R_{i,j}$ maintains a *preference list* for each client in C_i . The preference list is simply a list of all border router id's sorted in order of their distance (or cost) from the client.

A path P is advertised by $R_{i,j}$ to its client c , $c \in C_i$, if P is a feasible path, and furthermore, out of all the feasible paths, the distance (or cost) from c to the border router of P is the least. In the original protocol, $R_{i,j}$ used to advertise the same path, $best_{i,j}$, to all its clients.

As an example, consider Figure 6. $R_{1,1}$ and $R_{2,3}$ maintain the preference lists of $I_{1,2}$ and $I_{2,4}$. Even though $R_{1,1}$ prefers $eP_{3,5}$ over $eP_{4,6}$ (because $sp(R_{1,1}, R_{3,5}) < sp(R_{1,1}, R_{4,6})$), $R_{1,1}$ selectively advertises $eP_{4,6}$ to $I_{1,2}$.

Each router in C_i is responsible for sending its preference list to $R_{i,j}$. This requires iBGP to add a new *brpref* message.

The two changes above solve the known iBGP anomalies. Our solution requires $R_{i,j}$ to advertise only one path to client routers in C_i . This is important for the scalability of the solution.

```

reflector( $R_{i,j}$ )
constant
 $C_i$  : integer set /*  $R_i$ 's client router ids */
 $E_i$  : integer set /*  $R_i$ 's eBGP peer ids */
 $B_i$  : integer set /* cluster  $i$  border router ids */
 $R$  : integer set /* ids of  $R_{m,n}$  ( $m \neq i$ ) in  $AS_0$  */
 $B$  : integer set /* border router ids in  $AS_0$  */

variable
 $best_{i,j}$  : path /* best path */
 $AP_{i,j}, AP_{i,j}^*$  : path set /* avail., withdrawn paths */
 $NP, NP^*$  : path set /* peer's available,
withdrawn paths */
 $CAP_{i,j}$  : path set /* cluster  $i$  avail. paths */
 $cpref$  :  $|C_i| \times |B|$  integer array
/* client's preference list */

parameter
 $c$  : element of  $C_i$ 
 $g$  : element of  $R \cup B_i \cup E_i$ 

begin
rcv update( $NP, NP^*$ ) from  $g \rightarrow$ 
 $AP_{i,j}, AP_{i,j}^* := (AP_{i,j} \cup NP) - NP^*, AP_{i,j}^* \cup NP^*$ ;
 $best_{i,j} := best\_path(AP_{i,j})$ ;
 $CAP_{i,j} := (CAP_{i,j} \cup NP) - NP^*$  if ( $g \in C_i$ )
for every  $x$  in  $C_i$ 
send update(select_path( $FP_{i,j}, x$ ),  $AP_{i,j}^*$ ) to  $x$ ;
for every  $x$  in  $R$ 
send update( $CFP_{i,j}, AP_{i,j}^*$ ) to  $x$ ;

[]
rcv brpref( $br$ ) from  $c \rightarrow$ 
 $cpref[c] := br$ 
end

function select_path(paths set  $RP$ , client  $c$ )
 $i = 0$ ;
while  $cpref[c][i] \notin RP$ 
 $i := i + 1$ ;
return  $cpref[c][i]$ ;

```

Fig. 7. Reflector Router Pseudo-code

The reason anomalies are resolved is because our protocol creates shortest path trees rooted at the feasible border routers. Every interior router and border router, which are not feasible, joins a tree rooted at the nearest feasible border router.

Next, we present our protocol pseudo-code at the reflector and the client routers, respectively. In this code, we are ignoring issues related to eBGP.

C. Reflector Router Pseudo-code

The specification of reflector $R_{i,j}$ is shown in Fig. 7. The constants, variables and parameters are self-explanatory and consistent with the notation used in this paper. Variable, *cpref*, is used to store the sorted preference list of border router ids for each router in C_i .

In the first action, $R_{i,j}$ receives an *update* message from a peer. The peer may be another reflector $R_{m,n}$, $m \neq i$, a border router $B_{i,k}$ in C_i , or a router in a neighboring AS. The first assignment statement updates the available paths

($AP_{i,j}$) and withdrawn paths ($AP_{i,j}^*$) variables. The second assignment statement finds the new best path ($best_{i,j}$) by using $best_path(AP_{i,j})$. The third assignment statement updates $CAP_{i,j}$ variable, if the update message received is from a client router in C_i . The first group of send statements send to each client c an update message containing a selective path and the set of withdrawn paths. The selective path of c is obtained from the $select_path()$ function. This function takes $FP_{i,j}$ and c as input parameters and returns the selective path for client c . The second group of send statements send an update message to all other reflectors containing the cluster's feasible paths and withdrawn paths.

In the second action, the reflector receives a $brpref$ message from a client c . This receive action has one assignment statement, which updates the client's preference list array $cpref$.

D. Client Router Pseudo-code

The *client* specification is simple, as shown in Fig. 8. The constants and variables are self explanatory and consistent with the notation used in this paper. Variable $pref$ stores the list of border router ids sorted from the nearest (or least cost) to the farthest (or greatest cost) border router. In case of equal cost paths, every client uses the same deterministic tie-breaking rules, such as lowest border router identifier.

In the first action, $X_{i,j}$ receives an *update* message from reflector $R_{i,k}$. $R_{i,k}$ advertises a selective path through the nearest feasible border router. Upon receiving this message, $X_{i,j}$ updates $best_{i,j}$, $AP_{i,j}$, and $AP_{i,j}^*$ variables. These variables might be used to advertise paths to eBGP peers, if $X_{i,j}$ has any eBGP peers. In the second action, $X_{i,j}$ sends a sorted list of border routers ids to $R_{i,k}$ by using a $brpref$ message upon timeout. Timeout is enabled iff there are no $brpref$ messages in the channel $ps(client, reflector)$. We can easily implement timeout by using timers as explained in [16].

```

client( $X_{i,j}$ )
constant
 $R_{i,k}$  : integer      /* reflector id */
 $B$    : integer set  /* border router ids in  $AS_0$  */
variable
 $best_{i,j}$  : path      /* best path */
 $AP_{i,j}, AP_{i,j}^*$  : path set /* avail., withdrawn paths */
 $SP, WP^*$  : path set /* selective path and
                    withdrawn paths from  $R_{i,k}$  */
 $pref$  :  $|B|$  integer array
                    /* sorted list of border router ids */
begin
rcv update( $SP, WP^*$ ) from  $R_{i,k}$   $\rightarrow$ 
 $AP_{i,j}, AP_{i,j}^* := (SP \cup AP_{i,j}) - AP_{i,j}^*, AP_{i,j}^* \cup WP^*$ ;
 $best_{i,j} := best\_path(AP_{i,j})$ ;
[]
timeout  $brpref \# ps(X_{i,j}, R_{i,k}) = 0 \rightarrow$ 
send  $brpref(pref)$  to  $R_{i,k}$ 
end

```

Fig. 8. Client Router Pseudo-code

VI. RELATED WORK

There are many studies [11], [12], [13], [14] related to CMID anomalies. In the Walton et.al. [13] solution, a reflector finds a best path through each of the neighboring AS and advertises this path to iBGP peers if the best path's $local_pref$ and AS_path length values are equal to the reflector's overall best path's corresponding attributes. Basu et.al [14] showed a counter-example to Walton's solution. They also presented a solution, in which, the reflector advertises the paths obtained from $CFP_{i,j}$. They also proved the correctness of their solution. These solutions, in which, multiple path advertisements are required between every pair of iBGP peers, may not be scalable. This defeats the whole purpose of route reflection. Griffin et.al. [7] presented the necessary conditions to avoid the CI anomalies, but in doing so the network topology is restricted. In [10], the authors proposed a solution to solve some CI anomalies. On the other hand, in this paper, we are proposing a complete scalable S-iBGP protocol to solve both CI and CMI iBGP anomalies.

VII. CONCLUDING REMARKS

BGP has both divergence and forwarding anomalies. Divergence anomaly in eBGP has been well studied area in research community. In iBGP, both divergence and forwarding anomalies can occur. We proposed a S-iBGP to solve all known iBGP anomalies.

REFERENCES

- [1] Y. Rekhter, and T. Li. IETF RFC-1771: A Border gateway protocol 4 (BGP-4). Mar. 1995.
- [2] T. Bates and R. Chandra. IETF RFC-1996: BGP Route Reflection - An Alternative to Full Mesh iBGP. June, 1996.
- [3] K. Varadhan, R. Govindan, and D. Estrin. Persistent route oscillations in inter-domain Routing. Computer networks, Vol. 32, Issue 1, Jan. 2000, Pages 1-16.
- [4] T. Griffin, F.B. Shepherd, and G. Wilfong. The stable paths problem and interdomain routing. IEEE/ACM TON. Vol. 10, Issue 2, Apr. 2002, Pages 232-243.
- [5] J.A. Cobb, M.G. Gouda, and R. Musunuri. A Stabilizing solution to the stable path problem, Lecture Notes in Computer Science 2704, 2003, Pages 169-183.
- [6] L. Gao, and J. Rexford. Stable internet routing without global coordination. IEEE/ACM TON, Vol. 9, Issue 6, Dec. 2001, Pages 681-692.
- [7] T.G. Griffin, and G. Wilfong. On the Correctness of iBGP configuration. Proc. of SIGCOMM Conference, Aug., 19-23, 2002.
- [8] J.G. Scudder, and R. Dube. BGP Scaling Techniques Revisited. ACM Computer Communication Review, vol. 29, no. 3, Oct. 1999.
- [9] Dube, R. and Scudder, J., G., IETF Internet Draft, Route Reflection Considered Harmful. May, 1999.
- [10] R. Musunuri, and J.A. Cobb. Stable iBGP through selective path dissemination. IASTED PDCS, 2003.
- [11] D. McPherson, V. Gill, D. Walton, and A. Retana, IETF RFC-3345: Border Gateway Protocol (BGP) persistent route oscillation condition, Aug. 2002.
- [12] T.G. Griffin, and G. Wilfong, Analysis of the MED oscillation problem in BGP. Proc. of ICNP Conference, Nov., 12-15, 2002.
- [13] D. Walton, D. Cook, A. Retana, and J. Scudder. BGP persistent route oscillation solution. IETF Internet draft, May 2002.
- [14] A. Basu, C. L. Ong, A. Rasala, F.B. Shepherd, and G. Wilfong. Route oscillations in I-BGP with route reflection. Proc. of SIGCOMM Conference, Aug. 19-23, 2002.
- [15] B. Halabi, and D. McPherson. Internet routing architectures. Cisco press, second edition, 2000.
- [16] M.G. Gouda. Elements of network protocol design. Jhon Wiley & Sons, Inc., 1998.