Path-vector Routing Stability Analysis

Dimitri Papadimitriou
Alcatel-Lucent Bell
Antwerp, Belgium
dimitri.papadimitriou@alcatel-lucent.com

Florin Coras
Technical University of Catalonia
Barcelona, Spain
fcoras@ac.upc.edu

Albert Cabellos
Technical University of Catalonia
Barcelona, Spain
acabello@ac.upc.edu

Abstract

In this paper, we define a set of metrics that characterize the local stability properties of path-vector routing protocols such as BGP (Border Gateway Protocol). By means of these stability metrics, we propose a method to analyze the effects of BGP policy- and protocol-induced instability on local routers.

1. Introduction

Research efforts to understand BGP's instability led to categorize them into policy- and protocol-induced instabilities.

Policy-induced instabilities: addressing routing stability consistently with planned BGP routing policy implies eliminating non-deterministic routing states resulting from policy interactions and in particular, non-deterministic and unintended but unstable states. Griffin et al.'s seminal work [1] modeled BGP as a distributed algorithm for solving the Stable Paths Problem, and derived a general sufficient condition for BGP stability, known as "No Dispute Wheel". This sufficient condition guarantees the existence of a stable solution to which BGP always converges. Informally, this sufficient condition allows nodes to have more expressive and realistic preferences than always preferring shorter routes to longer ones. The game theoretic approach introduced in [2] relies on the best-reply BGP dynamics: a convergence game model in which each Autonomous System (AS) is instructed to continuously execute the following actions: i) receive update messages from BGP peering nodes announcing their routes to the destination, ii) choose a single peering node whose route is most preferred to send traffic to, iii) announce the new route to peering nodes. However, as proved in [2], best-reply BGP dynamics is not incentive-compatible even if No Dispute Wheel condition holds: even if all but one AS are following the BGP rules, the remaining AS may not have the incentive to follow them. Interestingly, as demonstrated in [2], incentive compatibility of best-reply BGP dynamics requires combining an additional global condition (Route Verification) together with the "No Dispute Wheels" to guarantee stability. Consequently, all known conditions for global stability are sufficient but not necessary conditions (checking them is an NP-hard problem and enforcing them requires a global deployment of an additional mechanism); on the other hand, local instability effects have yet to be characterized.

Protocol-induced instabilities: BGP is an inter-AS path-vector routing protocol subject to Path Exploration phenomenon like any other path-vector algorithm: BGP routers may announce as valid, routes that are affected by a topological change and that will be withdrawn shortly after subsequent routing updates. This phenomenon is the main reasons for the large number of routing updates received by BGP routers which exacerbate inter-domain routing system instability and processing overhead [3]. Both result in delaying BGP convergence time upon topology change/failure [4]. Several mitigation mechanisms exist to partially limit the effects of path exploration; however, none actually eliminate its effects. Hence, BGP is intrinsically subject to instability.

The goals of this paper are to 1) Develop a methodology to process and interpret the data part of BGP routing information bases in order to identify and document occurrences of Internet BGP routing stability phenomena; 2) Determine a set of stability metrics and develop methods for using them in order to provide a better understanding of the BGP routing system's stability; and 3) Investigate how path-vector routing protocol behavior and network dynamics mutually influence each other. The proposed approach aims to bring rigor and consistency to the study of routing stability. For example, it would allow for a unified approach to the cross-validation of techniques for looking at improving path exploration effects on the routing system.

2. Routing Table Stability and Metrics

2.1 Preliminaries

The AS topology of the routing system is described as a graph G = (V,E), where the vertices (nodes) set V, |V| = n, represents the AS, and the edges set E, |E| = m, represents the links between AS. At each node $u \in V$, a route r per destination d ($d \in D$) is selected and stored as an entry in the local routing table (RT) whose total number of entries is denoted by N, i.e., |RT| = N. At node u, a route r_i to destination d at time t is defined by $r_i(t) = \{d, (v_k=u, v_{k-1},\ldots,v_0=v), A\}$ with $k > 0 \mid \forall\ j,\ k \geq j > 0, \{v_j, v_{j-1}\} \in E$ and $i \in [1,N]$, where $(v_k=u, v_{k-1},\ldots,v_0=v)$ represents the AS-path, v_{k-1} the next hop of v along the AS-path from node u to v, and A its attribute set. Let $P_{(u,v),d}$ denote the set of paths from node u to v towards destination d where each path p(u,v) is of the form $\{(v_k=u, v_{k-1},\ldots,v_0=v), A\}$. A routing update leads to a change of the AS-path $(v_k, v_{k-1},\ldots,v_0)$ or an element of its attribute set A. A withdrawal is denoted by an empty AS-path (ε) and A = ∅: {d,ε,∅}. According to the above definition, if there is more than one AS-path per destination d, they will be considered as multiple distinct routes.

2.2 Routing Table Stability

The stability of a routing system is characterized by its response (in terms of processing of routing information) to inputs of finite amplitude. Routing system inputs may be classified as i) internal system events such as routing protocol configuration change or ii) external events such as those resulting from topological changes. Both types of events lead to the exchange of routing updates that may result in routing states changes. Indeed, BGP does not differentiate routing updates with respect to their root cause, their identification (origin), etc. during its selection process.

Definition 1: Let RT(t) represent the routing table at some time t. At time t+1, $RT(t+1) = RT_0(t) \oplus \Delta RT(t+1)$ where, $RT_0(t)$ is the set of routes that experience no change between time t and t+1, and ΔRT(t+1) accounts for all route changes (additions, deletions, and changes to previously existing routes) between time t and t+1.

The magnitude of the output of a stable routing system is small whenever the input is small. That is, a single routing information update shall not result in output amplification. Equivalently, a

stable system's output will always decrease to zero whenever the input events stop. A routing system, which remains in an unending condition of transition from one state to another when disturbed by an external or internal event, is considered to be unstable. More precisely, let $|\Delta RT(t+1)|$ be the magnitude of the change to the routing table (RT) at some time t+1, we distinguish three different equilibrium states for the routing table:

Definition 2: when disturbed by an external and/or internal event, a RT is considered to be *stable* if the following condition is met: $|\Delta RT(t+1)| \leq \alpha$, $t \rightarrow \infty$, where $\alpha > 0$ is small. In these conditions, if the routing system returns locally to its initial equilibrium state, it is considered to be (asymptotically) stable.

Definition 3: when disturbed by an external and/or internal event, a RT is considered to be *marginally stable* if the following condition is met: $\alpha < |\Delta RT(t+1)| \leq \beta$, $t \rightarrow \infty$, where $\beta > 0$ is small, $\alpha < \beta$.In these conditions, if the routing system transitions locally to a new equilibrium state, it is considered to be marginally stable.

Definition 4: when disturbed by an external and/or internal event, a RT is considered to be *unstable* if the following condition is met: $|\Delta RT(t+1)| > \beta$, $t \rightarrow \infty$. In these conditions, the routing system remains locally in an unending condition of transition from one state to another and it is considered to be unstable

The values α and β shall be set based on operational criteria. Among other factors, α and β depend on the observation sampling period that must be set to the Minimum Routing Advertisement Interval (MRAI) in order to ensure one routing update per sampling period. A similar reasoning to the one applied for the Loc_RIB stability (that corresponds to the BGP routing table) can be applied to the Adj_RIB_In (which stores incoming routes from neighbors). It is also interesting to measure the instability induced by the BGP selection process.

2.3 Stability Metrics

To measure the degree of stability of the Loc_RIB, Adj_RIB_In, and determine how close the routing system is to being unstable the following stability metrics are defined:

- Stability $\varphi_i(t)$ of selected routes $r_i(t)$: characterizes the stability of the selected routes r_i ($i \in [1,|D|]$) stored at time t in the Loc_RIB ($|Loc\_RIB| = N$) by quantifying the magnitude of change for these routes from time t to t+1.

```
When route r_i is created: φ_i(t) ← 0
if r_i experiences a path or an attribute change
(r_i(t+1) ≠ r_i(t)) then φ_i(t+1) ← φ_i(t) + 1
else /* r_i experiences no changes */
     if φ_i(t) = 0 then φ_i(t+1) ← 0
     else if φ_i(t) > 0 then φ_i(t+1) ← φ_i(t) - 1
          end if
     end if
end if
```

The computation of the stability metric for an entire routing table (RT) can then be derived form the stability of its individual routes. Let $|\Delta r_i(t+1)|$ denote the change in stability metric for a single route r_i from time t to t+1. These values are used to compute $|\Delta RT(t+1)|$ defined as the change in stability metric for the entire routing table from time t to t+1. Moreover, $|\Delta RT(t+1)|$ is normalized so that $0 \leq |\Delta RT(t+1)| \leq 1$, where 0 implies perfect stability, and 1 indicates complete instability.

```
For i=1 to N /* total nbr of routes in RT(t+1) */
   if r_i(t+1) is a new route then |Δr_i(t+1)| ← 0
   else if φ_i(t)=0 & φ_i(t+1)=0 then |Δr_i(t+1)| ← 0
        else if φ_i(t+1) > φ_i(t)
             then |Δr_i(t+1)|←[φ_i(t)+1]/[φ_i(t+1)+1]
             else |Δr_i(t+1)|←[φ_i(t)]/[φ_i(t+1)]
             end if
        end if
   end if
end i loop
|ΔRT(t+1)| ← Σ_i Δr_i(t+1)/ N
```

- Most stable route in the Adj_RIB_In ($|Adj\_RIB\_In| = M$): quantifies the relative stability between the routes to the same destination d, learned from different upstream BGP peers. Let $W_u \subset V$ denote the set of node's u BGP peers, $|W_u| = W \leq M$, and w one of its elements such that $(u,w) \in E$. Let $\varphi_{i,j}(t)$ denote the stability of the route r_i to destination d as received by peering router j ($j \in [1,W]$). At node u, $r'_{i,stable}(t)=\min\{\varphi_{i,j}(t), \forall j \in [1,W] \mid \{(v_k=u,v_{k-1}=w,\ldots,v_0=v),A\} \in P_{(u,v);d}, \forall w \in W_u\}$ defines –independently of the BGP selection rules– the selectable route that is the most stable for destination d at time t. Next, we define $\Delta\varphi_i$ as the relative measure of route's r_i stability with respect to the most stable route for the same destination d, $\varphi_{i,stable}$.

```
For i=1 to N /* |dst in Adj_RIB_In| = |Loc_RIB| */
    for j=1 to |W_u| /* nbr of peers for i^th dst */
        △φ_i,j(t+1)←[φ_i,j(t+1)+1]/[φ_i,stable(t)+1]
    end j loop
    △Φ_i(t+1) ← Σ_j Δφ_i,j(t+1)/|W_u|
end i loop
△Φ ← Σ_i △Φ_i(t+1)/N
```

- Best selectable route in the Adj_RIB_In: quantifies the relative stability between routes to the same destination d as learned from all upstream peers and the one amongst them selected by BGP at time t as the best route (thus following the BGP route selection). The computational procedure is the same as the previous one if one replaces $\varphi_{i,stable}$ by $\varphi_{i,selected}$.
- Differential stability between the most stable route in the Adj_RIB_In and the selected route stored in the Loc_RIB for the same destination d: characterizes the stability of the currently selected routes for a given destination d against most stable routes as learned from upstream neighbors. This metric provides a measure of the stability of the learned routes compared to the stability of the currently selected route. A variant of this metric, denoted $\delta\varphi_i$ ($i \in [1,|D|]$), characterizes the stability of the newly selected path p*(u,v) at time t for destination d against the stability of the path p(u,v) that is used at time t (i.e., stored in the Loc_RIB) for destination d and that would be replaced at time t+1 by the path p*(u,v): $\delta\varphi_i(t) = \varphi_i(t) - \varphi_i^*(t)$. In turn, if $\delta\varphi_i(t) > 0$, then the replacement of $r_i(t)$ by $r_i^*(t)$ increases stability of the route to destination d; otherwise, the safest decision is to keep the currently selected route $r_i(t)$ stored in the Loc_RIB.

Application of the metric $\delta\varphi_i$ during the BGP selection process would prevent replacement of more stable routes by less stable ones but also enable selection of more stable routes than the currently selected routes. However, for this assumption to hold we must prove the consistency of the stability-based selection with the preferential-based selection model that relies on path ranking function. For each $u \in V$, there is a non-negative, integer-value ranking function λ_u, defined over $P_{(u,v),d}$, which represents how each node u ranks its paths: if $p_1(u,v)$ and $p_2(u,v) \in P_{(u,v);d}$ and $\lambda_u(p_1) < \lambda_u(p_2)$ then p_2 is said to be preferred over p_1.

Definition 5: The route selection problem is consistent with the stability function $\delta\varphi(t)$ if for each $u \in V$ and $p_1(u,v)$ and $p_2(u,v) \in P_{(u,v);d}$ (1) if $\lambda_u(p_1) < \lambda_u(p_2)$ then $\delta\varphi(t) = \varphi_1(t) - \varphi_2(t) \geq 0$ and (2) if $\lambda_u(p_1) = \lambda_u(p_2)$ then $\delta\varphi(t) = 0$.

Theorem 1: if $p_1(u,v)$ and $p_2(u,v) \in P_{(u,v);d} \wedge p_2(u,v)$ is embedded in $p_1(u,v)$ then the route selection problem is consistent with the stability function $\delta\varphi$ and the route selection is stretch decreasing.

Proof: Assume without loss of generality that $p_2(u,v) = (v_k,v_{k-1},\ldots,v_{i+1},v_{i-1},\ldots,v_0)$ is embedded in $p_1(u,v) = (v_k,v_{k-1},\ldots,v_{i+1},v_i,v_{i-1},\ldots,v_0)$. Then, applying formula $\delta\varphi(t) = \varphi_1(t) - \varphi_2(t)$, we find $\delta\varphi = \varphi_1(v_k,v_{k-1},..,v_{i+1},v_i,v_{i-1},..,v_0) - \varphi_2(v_k,v_{k-1},..,v_{i+1},v_{i-1},..,v_0)$.

Assuming that paths p_1 and p_2 result from the composition of sub-paths (without increasing their total length), p_1 and p_2 can be each written as the concatenation of three sub-paths $p(v_k,v_{k-1},..,v_{i+1})p(v_{i+1},v_i,v_{i-1})p(v_{i-1},..,v_0)$ and $p(v_k,v_{k-1},..,v_{i+1})p(v_{i+1},v_{i-1})p(v_{i-1},..,v_0)$, respectively. The resulting stability function: $\delta\varphi = [\varphi(v_k,v_{k-1},..,v_{i+1}) + \varphi(v_{i+1},v_i,v_{i-1}) + \varphi(v_{i-1},..,v_0)] - [\varphi(v_k,v_{k-1},..,v_{i+1}) + \varphi(v_{i+1},v_{i-1}) + \varphi(v_{i-1},..,v_0)] = [\varphi(v_{i+1},v_i,v_{i-1}) - \varphi(v_{i+1},v_{i-1})]$.

We can observe that the difference $\delta\varphi$ result exclusively from the sub-paths defined between nodes v_{i+1}, v_{i-1} and assuming that the only instabilities are policy and/or protocol-induced, we obtain $\delta\varphi = [\varphi(v_{i+1},v_i) + \varphi(v_i,v_{i-1}) + \varphi(v_{i+1},v_{i-1}) - \varphi(v_{i+1},v_{i-1})] = [\varphi(v_{i+1},v_i) + \varphi(v_i,v_{i-1})] \geq 0$, proving the first part of Theorem 1.

Moreover, from its decomposition, the length d_1 of path p_1 verifies $d_1(u,v) > d_2(u,v)$, where d_2 is the length of path p_2. Hence, the route selection is stretch decreasing.

3. Experimental Results

This section presents a set of experimental results obtained by applying the metrics defined in Section 2 to real-world BGP data. The dataset we used was obtained from the Route Views project [5] that comprises archives containing BGP feeds from a set of worldwide distributed Linux PCs running Zebra.

Stability of Selected Routes: Considering that the number of selected routes is around 318k, Fig.1 shows that on average the Loc_RIB contains a few, between 60 and 120, unstable routes with minor contribution to the metric which can be interpreted as a sign of routing table stability. During the third day (5k minutes), 2 spikes separated by around two hours indicate large changes in stability. The first spike seems to suggest that 21 times more routes than the average experienced instabilities. However, BGP quickly converges to a new state that is disturbed by the second (smaller) spike since part of the affected routes return to the state before the occurrence of the first event. Overall, the plot shows a constantly changing but bounded churn within the table.

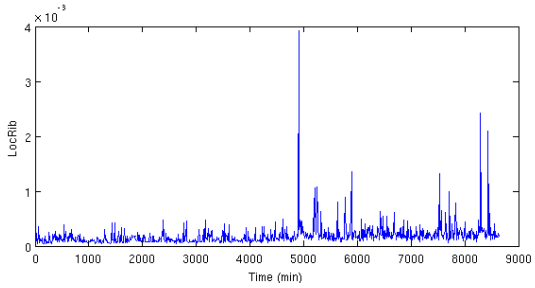

Fig.1. Stability of the selected routes

Most Stable Route: Fig.2 shows that on average the routes have slowly decreasing stability when compared to the most stable route. As a result, the plot has a small but positive slope. It does sometime present local minima which mark the points in time when the most stable route experiences changes. The average of the maximum metric value per destination d shows a positive but larger slope: the most unstable routes have a faster paced increasing instability. Further, during the entire observation duration (6 days), a subset of routes continuously presented instabilities leading to a monotonic increase of the metric.

Best Selectable Route: It can be seen from Fig.3 that the BGP selected route has, on average, a better stability than the other routes out of which it is selected. However, comparison between Fig.2 and Fig.3 reveals that the selected route exhibits slightly more changes than the most stable route (a lower metric value indicates a higher instability). Additionally, for the *avg* curve, the local maxima are correlated with the local minima of the previous metric and are likewise due to a diminishing stability of the most stable route. One can also observe the same monotonously increasing trend of the metric for both the average and the maximum, due to routes with sustained instability.

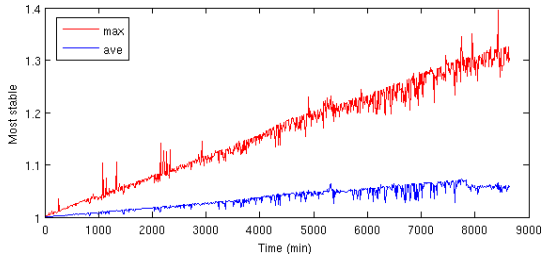

Fig.2. Most stable metric

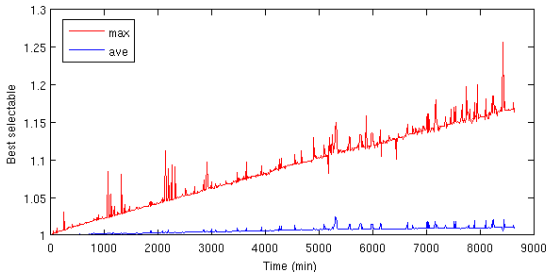

Fig.3. Best selectable metric

4. Future Work

In this paper, we define and provide a first validation of several stability metrics that characterize the effects of BGP policy- and protocol-induced instabilities on local routers. Our initial results show that the proposed method enables detecting instability events and their impact on local routing tables. Ongoing work includes verifying if the route selection problem is consistent with the stability function $\delta\varphi$, and determining the general trade-offs between stability-based route selection and the resulting stretch increase/decrease factor on the selected routing paths

5. References

[1] Griffin, T., Shepherd, F. B., Wilfong, G., The Stable Paths Problem and Interdomain Routing, *IEEE/ACM Trans. Networking*, 10(1):232–243, 2002.

[2] Levin, H., Schapira, M., Zohar, A., Interdomain routing and games, *Proc. ACM Symposium on Theory of Computing (STOC)*, 2008.

[3] Huston, G., Damping BGP, *RIPE 55, Routing Working Group*, 24 October 2007.

[4] Labovitz, C., Ahuja, A., Bose, A., Jahanian, F., Delayed Internet Routing Convergence, *IEEE/ACM Trans. Networking*, 9(3):293-306, 2001.

[5] Univ. of Oregon. RouteViews. *http://www.routeviews.org*