Evaluating Internal BGP Networks from the Data Plane

Feng Zhao, Xicheng Lu, Baosheng Wang, Peidong Zhu

School of Computer, National University of Defense Technology,
Changsha 410073, Hunan, China
fengzhao1980@tom.com

Abstract. This paper focuses on the design of IBGP networks, which are very important to the reliability and stability of Internet. Although several metrics have been presented to measure the robustness of IBGP networks, they only considered the impact of route reflection networks on the control plane. A robust network should have low sensitivity to traffic load variations. So we propose a new metric to characterize the impact of route reflection networks on the data plane, which is called TDR (Traffic Diversion Rate). Simulation results show that adopting the optimal route reflection topology that minimizes TDR will make the network lose or shift much less traffic, compared with adopting the optimal route reflection topologies found according to other metrics.

1 Introduction

Full-mesh IBGP does not scale well. As an alternative to full mesh IBGP, BGP route reflection is often used in IBGP topology design for large ASes. If there is no IBGP session that will fail, it does not matter adopting which IBGP topology. However, link or router failures occur as part of everyday operation in backbone networks of a large AS. They can cause IBGP session failures. When an IBGP session is lost, all related routes in the BGP routing tables have to be withdrawn and thus IP networks may become unreachable. Also when routes change, some traffic flows may be forwarded along different intradomain paths. Traffic shifts will modify the distribution of the traffic inside the network and change the load of some links. As a consequence, some links can even become congested. Robust IBGP networks are very important to the stability and reliability of Internet.

To measure the robustness of route reflection networks, reference [1-4] propose several metrics: *Hop Count, Reliability Product, IBGP Expected Lifetime, Expected Session Los, Resilience, IBGP Failure Probability, Expected Connectivity Loss.* However, they only considered the impact of route reflection networks on the control plane. Even small routing shifts of popular routes can impact the data plane by causing a large swing in traffic, perhaps leading to congestion, loss, delay, and jitter. A robust network should have low sensitivity to traffic load variations [5]. So we set out to evaluating the control plane quality by evaluating the data plane. We propose a new metric to characterize the impact of route reflection networks on the data plane, which is called TDR (Traffic Diversion Rate).

2 Traffic Diversion Rate

A typical network in an AS is represented as graph $G(V, E)$. Node set V represents all the routers. E is the set of physical links. We use S to denote the set of all network states, which includes all failure states and the state without any failure. F_s is the set of physical components that fail in state $s \in S$, and $F_s \in V \cup E$. Other components, which are not in F_s, work properly. The probability that state s occurs is r_s.

If router i can not exchange BGP routing information with router j directly or indirectly due to IBGP session failures, i and j are separated logically from each other and we denote this relation as $[i \leftarrow\mapsto j]$.The probability that i and j are separated logically from each other in failure state s is denoted by $Pr_s[i \leftarrow\mapsto j]$.

We denote the ingress-to-egress traffic matrix by M and denote the traffic that flows from router i to router j by M_{ij}.We define the Traffic Diversion Rate in state s as T_s.

$$T_s = \frac{\sum_{i,j \in V_r} M_{ij} Pr_s[i \leftarrow \mapsto j]}{\sum_{i,j \in V_r} M_{ij}}$$

Therefore, the Traffic Diversion Rate over the entire state space is $T = \sum_{s \in S} r_s T_s$

After having defined the reliability metric for IBGP networks, we describe the optimization problem based on this metric as follows:

Problem 1 (Robust Reflection – TDR (RR-TDR)):

Given: a network $G(V, E)$; the upper bound of the node degree: $\{c_i\}$, where $i \in V$; the probabilities of failure scenarios : $\{r_s \mid s \in S\}$,where $S = V \cup E$; IBGP session failure probability $\{q_s \mid s \in S\}$; the ingress-to-egress traffic matrix $\{M_{ij}\}$,where $i, j \in V$; IGP weights, we look for a reflection topology $G_r^*(V, E_r)$ such that (1) $h_i \le c_i$, for $\forall i \in G_r^*.V$; (2) $TS(G_r^*) \le TS(G_r)$, for any reflection topology $G_r(V, E_r)$ which satisfies $h_i \le c_i$, for $\forall i \in G_r.V$.

The goal of the IBGP topology design problem is to find a route reflector topology with minimum Traffic Diversion Rate.

3 Experiments

The optimal IBGP topologies based on different metrics may be different. We think between two IBGP topologies, if the network loses or shifts less traffic under an IBGP topology than under the other, then the IBGP topology is better than the other.

The source data of our experiments comes from a real network, the US research network (Abilene). By using TOTEM, we got the realistic traffic matrix M of Abilene on Jan 1, 2005. Also by using TOTEM, we calculated the number of external routes that are obtained by a router from its EBGP peers and are further injected into the IBGP network.

Theoretically, there could be multiple levels of reflection. However, in practice, the two-level reflection is most often used. In the experiments, we only focus on the two-level reflection. We further assume that the reflection graph has not redundancy.

In the experiments, we assume that all routers won't fail and the failure probabilities of all links are uniform. Also the conditional failure probabilities of the IBGP sessions that are affected by the link failures are uniform, which are not zero. And we assume the number limit of IBGP sessions is 6. Then with over 1000 lines of Matlab code, we find the optimal IBGP topologies for different metrics. The optimal IBGP topologies for some different metrics may be the same. We get the 5 IBGP topologies for these 8 metrics: topology 1 for the metrics of *Hop Count*, *Reliability Product* and *IBGP Failure Probability*; topology 2 for the metric of *Expected Lifetime;* topology 3 for the metric of *Expected Session Loss;* topology 4 for the metric of *Resilience* and *Expected Connectivity Loss;* topology 5 for the metric of *Traffic Diversion Rate.*

We use SSFNet to study the impact of these different route reflection networks on the data plane. In our simulations, there is a network of 11 stub ASes and a transit AS. The topology of the transit AS is the core Abilene network topology. Each stub AS has one router (which connects a router in the transit AS) running BGP and one host. In this transit AS, BGP is running at all routers. All of the Internal BGP peering sessions use loopback addresses for peer destination addresses. And each router in the AS runs OSPFv2.

To simulate that there will occur IBGP session failures, we set the experiment parameters as follows: BGP hold time: 60 seconds; OSPF router dead interval: 80 seconds; OSPF hello interval: 10 seconds; Link failure duration: 90 seconds. With these parameter settings, the OSPF routing recovery time will be greater than 70 seconds. Thus a link failure will cause some IBGP sessions to fail. It may take 165 seconds for failed IBGP sessions to be reestablished after the OSPF routing recovers. So we set the simulation time 500 seconds to allow failed IBGP sessions to be reestablished before a simulation ends. At the time of 200th second, the application session of each host starts to send packets to other hosts. The host in AS i will send packets to the host in AS j with the rate $M(i, j)/100$. At the time of 210th second, we inject a network failure by bring down a link in the transit AS. And the link is recovered at the time of 300th second. Because there are 5 optimal IBGP topologies for the metrics and 14 links in the core, we perform 70 simulations and record the number of packets lost in the transmission.

The average number of lost packets for each link failure is shown in figure 1. From this figure we can see that the number of lost packets is the least under the optimal IBGP topology based on the TDR metric.

Although we do the experiment only with a traffic pattern and a failure model, we believe under other traffic patterns and other failure models, the optimal topology based on TDR metric will make the network lose or shift much less traffic, because TDR metric is derived from the data plane directly.

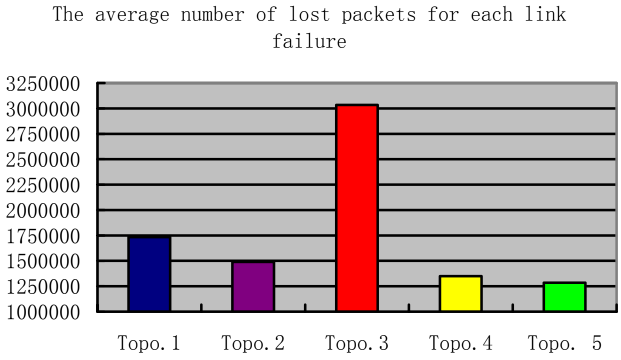

Fig. 1. Number of IBGP sessions: Rocketfuel ISP topologies

4 Conclusion

This paper proposes TDR, a new metric to characterize the impact of route reflection networks on the data plane. Our experiment shows that under the same traffic pattern and the same failure model, different IBGP topologies will make the network exhibit different behaviors. And the optimal topology based on TDR metric will make the network achieve better performance than other optimal IBGP topologies based on the metrics proposed before.

References

1. L. Xiao, and K. Nahrstedt, Reliability models and evaluation of internal BGP networks, in Proceedings of IEEE INFOCOM, 2004.
2. L. Xiao, J. Wang, and K. Nahrstedt, Optimizing ibgp route reflection network, in Proceedings of IEEE ICC, 2003.
3. L. Xiao, J. Wang, and K. Nahrstedt, Reliability-aware ibgp route reflection topology design, in Proceedings of IEEE ICNP, 2003.
4. L. Xiao, and K. Nahrstedt, Reliability of Internal BGP Networks: Models and Optimizations, Technical Report UIUCDCS-R-2005-2608/UILU-ENG-2005-1800, 2005.
5. R. Teixeira, N. Duffield, J. Rexford, and M. Roughan, Traffic matrix reloaded: impact of routing changes, In Proc. of PAM 2005, 2005.