Outbound Load Balancing in BGP Using Online Simulation

Tao Ye, Hema Tahilramani Kaur, Shivkumar Kalyanaraman
Electrical Computer and Systems Engineering Department,
Rensselaer Polytechnic Institute, Troy, NY-12180.
{yet3,hema,shivkuma}@networks.ecse.rpi.edu

***Abstract*—BGP is the standard inter-AS routing protocol in current Internet. However, since it does not taken into account network conditions when making routing decision, its routing may not be optimal in terms of QoS constraints. This paper investigates the load balancing problem when an AS has multiple outbound links connected with other ASes and presents a dynamical tuning approach to achieve load balancing without modifying BGP protocol in any way. In this approach, the load balancing is generalized with a black-box optimization problem, thus, the on-line simulation system, which is specifically designed to tackle this class of problems, can be applied to adaptively adjust BGP configuration for load balancing. The simulation results shows the proposed scheme greatly improve the load balancing of BGP routers.**

I. Introduction

In current Internet, IP traffic is mapped onto the network by standard routing protocols, such as, OSPF for intra-AS traffic and BGP for inter-AS traffic. These routing protocols are so-called topology-driven protocol. That is, when routing the traffic, the routing algorithms used in these protocols normally do not take into account current network conditions and QoS constraints. Therefore, the routing generated by these algorithms tends to result in a highly unevenly mapping of traffic. That is, some links may get very congested since most of traffic go through it and the other may be consistently underutilized. This phenomenon has been confirmed in many traffic measurements[1], [2] which observed large variation in link utilization across the network. Traffic engineering tries to eliminate this situation by taking into considreation network conditions, such as traffic demands when making routing decisions.

The goal of traffic engineering is to optimize the distribution of traffic load accross the network such that network resources are optimally utilized. One approach is to deploy MPLS and contraint-based routing to achieve traffic engineering[3]. In this approach, exisiting IGP protocols are extended to exchange additional network information, such as, link bandwidth and state, and optimal routing is computed with these factors besides network topology. In addition to this approach, there are increasing insterests in the methods which achieve similiar traffic engineering functionality without requiring modificiations on existing routing protocols. Basically, these methods are established upon the assumption that there are adjustable parameters in routing algorithms and the performances of routing algorithms are sensitive to these parameter. In these methods, the parameters of routing algorithms are dynamically adjusted based on current network conditions and QoS constraints. The tuning of the prameters is usually formulated as a black-box optimization problem, i.e., the network is considered as a black-box with routing parameters as its input and a performance metric as its output. With such formulation, a black-box optimization algorithm, such as, genetic algorithm and simulated annealing can be used to find near-optimal solutions. For example, in [4], such an approach is presented, where the link weights of OSPF algorithm are optimized for distributing the given traffic demand evenly across the network. Since this approach does not change existing routing protocls in any way, it can be quickly employed on the network.

Based on the above black-box optimization approach, an online simulation scheme has been proposed in [5] for the adaptive configuration of network protocols. As shown in Figure 1, in this scheme, network conditons are monitored and the parameters of the network are adaptively optimized based current network conditions. With such a scheme, any network protocol can be tuned for better operation as long as it is sensitive to its parameter setting. Whenever network conditions change, the optimization will be redone to adapt the protocol setting to current conditions. In this way, this scheme provides an adaptive tuning of network protocols.

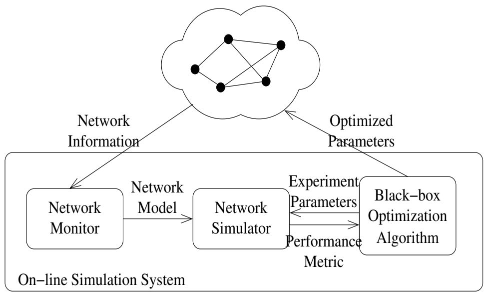

Fig. 1. On-line simulation framework for network protocol optimization

This paper will consider the load balancing problem in BGP enviroment. With the knowledge of traffic demand, routing decision can be better made for traffic engineering purpose. Ideally, the routing should be made with a networkwide consideration of traffic demand. This is feasible for intra-AS routing since all the routers are under the same adminstrative orginazation and it is possible to get traffic demand statistics and obtain the control on these routers. However, this is not true for inter-AS routing, i.e., for different administrative domain, the traffic demand statistics are usually kept private and the unified control on all routers are impossible. Therefore, the traffic engineering can only be performed locally. In this paper, we will consider a specific load balancing problem, i.e., the outbound load balancing of a single AS and attempt to apply the on-line simulation scheme to this problem.

The rest of the paper is organized as follows: in Section 2, the outbound load balancing problems in BGP enviroment is described and the optimization-based approach is presented. In Section 3, various simulation results are presented, which show that load balancing can be greatly improved with the proposed scheme. Section 4 concluded this paper and discuss further directions.

II. Outbound Load Balancing in BGP

It is now very common that an AS may peer with multiple neighboring ASes. In this case, there are multiple outbound links connected to other ASes. The outbound traffic of this AS may be routed on any of these links depending on the decision made by the inter-AS routing algorithm, usually BGP. The load balancing problem we will consider here is to distribute the outbound traffic over the output links such that each outbound link is evenly utilized. It should avoid the situation where some links get heavily congested while the others are underutilized.

BGP is the standard routing protocol for inter-AS traffic, which uses the path vector routing algorithm to choose a route traversing the least number of ASes. BGP routing algorithm itself cannot achieve load balancing since it makes the routing decision based on the length of AS path without considering the outbound traffic demand. However, BGP offers the flexibility of overiding the default routing with `local_pref` attributes. `local_pref` attribute of a path is first considered when BGP make the routing decision and the path with highest `local_pref` is selected when there are multiple alternative links for one destination. Therefore, if we know the desired routing which is able to realize our traffic engineering objective, we can use `local_pref` to overide the default BGP routing and deploy the ideal routing instead. Based on this idea, this paper proposes a load balancing scheme, which caculates the ideal routing based on current network condtions and QoS constraints deploys it with `local_pref` attributes. In the following, we will describe the details of this scheme.

A. Traffic Demands

Given a certain outbound traffic demand, load balancing is to split this traffic demand and distribute them onto different links. Usually, the traffic demand is composed of a number of traffic flows. In the smallest granualrity, a traffic flow is determinted by IP 5-tuples. In the coarse granualtiry, a traffic flow can also be identified by the source and destination AS pair. In [1], the measurement has shown that the traffic aggregate based on destionation prefixes are suitable for load balancing since they are relatively stable through the day on per-hour time scales. This granularity is taken in our load balancing scheme, i.e., the traffic demand is splited based on the per-prefix level of flows. Currently, many methods are available to obatin the traffic demand statistics[6], such as SNMP management information base and traffic measurement.

For a typical BGP router, there are normally thousands of destination prefixes in the routing table. It will be very complex to work with such a large number of traffic flows. Many measurments[1], [2] have demonstrated Internet traffic exibits the so-called elephant and mice phenomenon, i.e., the small number of traffic streams, i.e., *elephants*, generate a large portion of total traffics, while a large number of steams, i.e., *mice*, generate a small portion of total traffic. And the elephant traffic flows are usually very stable over time and therefore suitable to be rerouted for load-balancing purpose. For example, in [2], it has been found that the top 9% of flows between ASes accounts for 86.7% of packets and 90.7% of bytes transmitted. Therefore, it is not neccessary to consider all the traffic flows in the load balancing scheme. In this paper, we only attempt to adjust the routing of the top 10% destination prefixes in the routing table based on their traffic demands, i.e., a few hundred of flows.

B. Optimal Routing Calculation for Load Balancing

In addition to traffic demands, network topology should also be known to calcuate optimal routing for traffic engineering objectives. In the case of outbound load balancing, only the number of outbound links and their bandwidthes need to be known. Given the knowledge of traffic demand and outbound link information, the optimal routing for load balancing can be calculated.

Suppose for the concerned AS there are m outbound links l_i, $i = 1 \dots m$, and each link has a bandwidth of c_i, $i = 1 \dots m$. The outbound traffic of this AS will be routed onto these links. Assume each link gets a share of output traffic s_i, $i = 1 \dots m$, the utilization of link l_i can be written as s_i/c_i. The objective of load banlancing is to minimize the maximum link utilization among all the outbound links, i.e.,

$$\text{minimize} \max_{i=1\dots m} \frac{s_i}{c_i} \tag{1}$$

Assume there are n top 10% traffic flows and let d_j, $j = 1 \dots n$, denote the loads of these flows, what our load balancing scheme attempts to do is to adjust the routing of these n flows to minize the objective in Equation(1. Let $\mathcal{D}_i$ the subset of the n traffic flows which are output on link l_i under adjusted routing and f_i the load on link l_i generated by the other 90% traffic flows, which distributed onto this link by BGP default routing, then Equation (1) becomes:

$$\text{minimize} \max_{i=1\dots m} \sum_{j\in\mathcal{D}_i} p_{ij} + h_i \tag{2}$$

where $p_{ij} = d_j/c_i$, which reprents the percentage load if flow d_j is output on link l_i, and $h_i = f_i/c_i$, which represents the percertage load generated by the other 90% flows on link l_i. This problem can also be written as the following integral programing problem:

$$
\begin{aligned}
\text{minimize} \quad & t \qquad\qquad\qquad\qquad\qquad\qquad\qquad (3)\\
\text{subject to} \quad & \sum_{j=1}^{n} x_{ij} p_{ij} + h_i \le t,\ i = 1 \ldots m \\
& \sum_{i=1}^{m} x_{ij} = 1,\ j = 1 \ldots n \\
& x_{ij} \in \{0, 1\},\ i = 1 \ldots m,\ j = 1 \ldots n
\end{aligned}
$$

where x_{ij} is a binary number and $x_{ij} = 1$ means flow d_j is output on link l_i, otherwise $x_{ij} = 0$. Note that traffic flow d_j may not have all outbound links as its alternative paths and we can arbitrarily assume a very large p_{ij} for those links. The problem represented by Equation(3) is actually a classical task scheduling problem with unrelated parallel machines[], where a number of tasks with different sizes are assigned to a set of parallel machines. The processing time of one task is different on different machines and the objective there is to minimize the completion time of all tasks by carefully distributing these tasks onto the parallel machines. This problem is NP-hard and approximation algorithms can be used to obtain approximate solutions. For example, in [7], linear programming technique is first used to obtain a basic solution where there are at most $m - 1$ non-integral x_{ij}. Then for these non-integral x_{ij}, an exhaustive enumeration is performed to find the optimal scheduling. Combining the solutions of these two steps can produce an approximate solution with a upper bound of $2t^*$, where t^* denotes the value of t produced by the optimal solution. The time complexity of this method is exponential in the value of m.

Another way to tackle the problem represented by Equation(2) is to formulate it as a black-box optimization problem and use general optimization algorithms, such as genetic algorithm and simulated annealing, to search for near-optimal solutions. It has been demonstrated that optimization algorithms are very effective and can deliever near-optimal solutions for many practical problems[4], [8]. The advantage of black-box optimization approach is that flexible optimization objectives can be used without modifying the underlying optimization scheme For example, we can also optimize for minimum average pac loss instead of maximum link utilization in Equation (1). I also possible to optimize for multiple objectives by comb these objectives with multi-objective optimization techniqu For example, in addition to traffic engineering, it is also sired that shortest pathes are used as much as possible. W multi-objective optimization, we can obtain the solution wh achieve the objectives of both traffic engineering and short path.

In this paper, the black-box optimization approach has be used for BGP load balancing scheme. Many effective timization algorithms are available. The Recursive Rand Search(RRS) algorithm[9] has been used in our scheme. T algorithm is specifically designed for the requirments of adaptive networking optimization problems, where the efficiency is greatly emphasized, i.e., finding a good result within a short time is more important than full optimization. This is different from most of existing optimization algoirthms, such as genetic algorithm and simulated annealing, which normally put full optimization first. The RRS algorithm has been demonstrated to be able to perform very efficiently for engineering problems where high-dimensinal and noisy objective functions usually present. Another unique advanateg of RRS is to be able to automatically exclude negligible parameters from the optimization process. Readers can refer to [9] for the details of RRS.

C. The Scheme

The load distribution generated by the above optimization method can be deployed by seting `local_pref` atrribute for each prefix. The whole procedure can be summarized as follows:

1) Extract destination prefixes $d_j,\ j = 1 \ldots n$ with top 10% traffic demands from the routing table;
2) Caculated $p_{ij} and h_i,\ i = 1 \ldots m,\ j = 1 \ldots n$ accourding to the traffic demand of each prefix and the capacity of each outbound link. Assign an arbitrarily large value, say 1, for p_{ij} where there is no routing of d_j on l_i;
3) Compute the approximate solution using the optimization algorithm;
4) Based on the obtained x_{ij}, set `local_pref` of link l_i for prefix d_j to a high value if $x_{ij} = 1$.

Network conditions keep changing and traffic demands also vary from time to time. Whenever it is found that previous load balancing is broken by these changes, the above procedure can be repeated with new traffic demands to achieve new load balancing. In this way, adaptive load balancing is accomplished. Network conditions can be directly monitored to dectect signficant changes and start the tuning process accordingly. The maximum link utilization can also be monitored and when it increases beyond a certain level, the tuning can be performed. This adaptive BGP load balancing procedure is implemented in on-line simulation framework.

III. Simulation Results

In this section, we present the simulation results for the load balancing scheme described as above. The BGP network is sim-

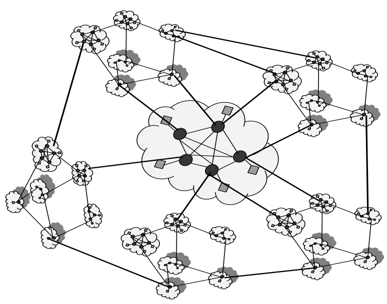

Fig. 2. Simulation network topology

First, we will show the effectiveness of the scheme.

Second, we will show the adaptiveness of the scheme.

Third, we will show the flexibility of the scheme, for example, use other optimization objectives.

IV. CONCLUSION

This paper presented an black-box optimization approach for the outbound load balancing in BGP. An general adaptive network configuraiton scheme, i.e., on-line simulation scheme, has been used for this purpose. Basically, this approach formulates the load balancing of BGP as a black-box optimization problem and uses an optimization algorithm search for near-optimal solutions. Simulation results have shown that the load balancing factor can be greatly improved by the described scheme.

Although the work presented here is limited to examine the objective of load blancing, this approach is highly flexible and can be used to achieve other optimization objectives, such as, minimization of the sum of link utilization. In addition to BGP, the on-line simulation scheme has also be succussesfully used for the adaptive configuration for other network protocols, such as OSPF[11] and RED[12].

REFERENCES

[1] S. Bhattacharyya, C. Diot, J. Jetcheva, and N. Taft. Pop-level and access-link-level traffic dynamics in a tier-1 pop. In *ACM SIGCOMM Internet Measurement Workshop*, November 2001.
[2] Wenjia Fang and Larry Peterson. Inter-as traffic patterns and their implications. In *Proceedings of Global Internet 99*, Rio, Brazil, 1999.
[3] D. Awduche and et. al. Requirements for traffic engineering over mpls. RFC 2702, Septemper 1999.
[4] Bernard Fortz and Mikkel Thorup. Internet traffic engineering by optimizing ospf weights. In *Proceedings of the INFOCOM 2000*, pages 519–528, 2000.
[5] Tao Ye and et al. Traffic management and network control using collaborative on-line simulation. In *Proc. of IEEE ICC'01*, Helsinki, Finland, 2001.
[6] M. Grossglauser and J. Rexford. Passive traffic measurement for ip operations. In *The Internet As A Large-scale Complex System*. Oxford University Press, 2002.
[7] C. N. Potts. Analyssis of a linear programming heuristic for scheduling unrelated parallel machines. *Discrete Appl. Math.*, 10:155–164, 1985.
[8] Zelda B. Zabinsky. Stochastic methods for practical global optimization. *Journal of Global Optimization*, 13:433–444, 1998.
[9] Tao Ye and Shivkumar Kalyanaraman. A recursive random search for optimizing network protocol parameters. Technical report, ECSE Department, Rensslaer Polytechnique Institute, Dec 2001.
[10] SSFNET. network simulator. http://www.ssfnet.org.
[11] T. Ye, Kaur H. T., and et. al. Dynamic optimization of ospf weights using online simulation. submitted.
[12] Tao Ye and Shivkumar Kalyanaraman. Adaptive tuning of red using on-line simulation. In *Proceedings of IEEE GLOBECOM'2002*, Taibei, Taiwan, Nov. 2002.