Assessing Survivability of Inter-domain Routing System under Cascading Failures

Yujing Liu[1], Wei Peng[1], Jinshu Su[1], and Zhilin Wang[2]

[1] College of Computer, National University of Defense Technology
Changsha, China
{liuyujing,wpeng,sjs}@nudt.edu.cn
[2] Education Department, National University of Defense Technology
Changsha, China
wangzhilin@nudt.edu.cn

Abstract. The Internet is designed to bypass failures by rerouting around connectivity outages. Consequently, dynamical redistribution of loads may result in congestion in other networks. Due to the co-location of data plane and control plane traffic of Border Gateway Protocol (BGP), the survivability of inter-domain routing system is sensitive to severe congestion. Therefore, an initial outage may lead to a cascade of failures in the Internet. In this paper, we characterize the survivability of inter-domain routing system by reachability and number of rerouting messages, and propose a model for studying the relationship between the survivability and the capacity of AS links under intentional attacks and random breakdowns. Through simulations on an empirical topology of the Internet, we find that the cascading failures bring a great deal of added burden to almost all the core ASes. When the tolerance parameter of AS links is less than 0.1, the cascading effect tends to be amplified globally. Moreover, the effect triggered by intentional attack is greater than that triggered by random breakdown. But the difference between them is not as prominent as previous research due to the unique automatic-restoration process in inter-domain routing system.

Keywords: the Internet, inter-domain routing, survivability, cascading failure.

1 Introduction

The Internet is composed of tens of thousands of Autonomous Systems (ASes), which exchange routing messages with each other by the de-facto inter-domain routing protocol - BGP. The reliability of BGP is very important to achieve stable communications in the Internet. Currently, the routing control packets of BGP share resources such as bandwidth and buffer space with normal data traffic in Internet packet forwarding. This co-location of control plane and data plane makes BGP sensitive to severe network congestion.

In the Internet, traffic is rerouted to bypass malfunctioning segments, probably leading to overloads on some of other healthy networks, resulting in congestion there. Loss of routing messages due to the congestion can cause BGP

J. Su et al. (Eds.): ICoC 2013, CCIS 401, pp. 97–108, 2013.

session failures between ASes, leading to another round of rerouting. Similarly, the dynamical redistribution of traffic loads may disconnect other pairs of BGP sessions. Meanwhile, the previous 'failed' sessions may re-establish since the links are no longer congested after all the traffic were rerouted around them. A single fault in routers or links can trigger a sequence of route changes on a global scale. This process is what we call cascading failures in inter-domain routing system.

Previous works about robustness of BGP in congested networks study the relationship between traffic overload factors (i.e. queueing delays, packet sizes, TCP retransmission parameters and so on) and lifetime of a BGP session [1], [2]. This is a micro-view of survivability of BGP, focusing on single component in the system. However, inter-domain routing system is a complex network, whose behaviour is better characterized by the dynamics induced by interactions of BGP routers in the whole Internet. On the other hand, studies on cascading failures in complex networks have shown that networks with highly heterogeneous distribution of loads such as the Internet and electrical power grids are particularly vulnerable to attacks in that a large-scale cascade may be triggered by disabling a single key node [3], [4]. But in these models, overloaded nodes are either removed or avoided. They are not suitable to describe the unique 'virtual cut' and 'automatic restoration' characteristics of BGP links under dynamical congested state. Recently, a CXPST attack is presented utilizing this property of BGP to create control plane instability by using only data plane traffic [5]. Besides this specific attacking technique, it's also important to further analyse factors that affect the instability scope and the difference between random breakdowns and intentional attacks.

In this paper, we try to answer questions that: Considering the distinct property of inter-domain routing system, under what conditions can a global cascade take place? And who will be affected worst? Our contributions with respect to previous works are summarized as follows.

(1) We propose a model for studying the inter-domain routing process under a cascade of congested and resumed link states. In this model, the overloaded links are not removed from the network. They have chances to be restored in the future. Moreover, we apply *customer-prefer* and *valley-free* policy to routing process instead of simply using shortest path algorithm, in order to better comply with the actual situation.

(2) We characterize the survivability of inter-domain routing system by *reachability* and *number of rerouting messages*. The most critical ability of routing system is making routing decisions. Reachability can evaluate how the incomplete topology will affect the capability; and number of rerouting messages can evaluate the effect of instability of the routing system. Because a surge of BGP updates generated by large-scale reroutings may exceed the computational capacity of affected routers, causing a degradation of such capability.

(3) In our model, the survivability depends on congested state of AS links, which are decided by the comparison of their loads and capacities. Therefore, we study the relationship between survivability of inter-domain routing system and capacity of AS links. This examination reveals that when the tolerance

parameter of links is less than 0.1, a global cascading failure of the Internet will emerge.

(4) An initial failure is the trigger that causes consequent cascades. Here we focus on failure of a single AS link. We divide the initial failure as intentional attack and random breakdown, and further analyse the survivability under these two kinds of initial failures. We find that intentional attack causes greater effect than random breakdown. But the effect is weakened by the automatic-restoration process in inter-domain routing system.

2 Model for Cascading Failures in Inter-domain Routing System

We demonstrate the cascading process in inter-domain routing system by a simple example as shown in Fig. 1. In this simple topology of ASes, the initial path from S to D is $S \to A \to B \to D$. After an initial failure happens on the link between A and B, routers compute new routes to bypass the faulty link. Path 1, i.e., $S \to A \to C \to B \to D$ is chosen to carry the rerouted traffic from the initial path. However, the redistribution of traffic load surpasses the capacity of link $A-C$. Unlike the electrical power grids, the overload in the Internet will not lead to breakdown of the link, but cause a congestion between A and C. Furthermore, large amount of packets are lost. Unfortunately, the routing message in control plane and the data traffic in data plane share limited resources in BGP routers. So sever congestion will drop the *KEEPALIVE* messages spoken by BGP routers at two ends of link, and make them 'think' that the session between them is disconnected. Hence routers start to compute other new but less preferred route from S to D. It turns out to be path 2, i.e., $S \to A \to E \to F \to B \to D$. At this time, traffic is rerouted away from link $A - C$. The link is no longer congested. Routers in A and C resume their BGP session automatically. And the more preferred route - path 1 is available again. Traffic is rerouted to path 1, then another round of congestion happens.

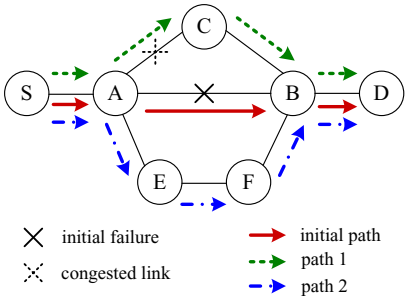

Fig. 1. An example of cascading failures in inter-domain routing system

If the topology of the network is as complex as the Internet, the dynamics of rerouting will be more complicated. A single initial failure may lead to lots of 'virtually cut' and 'automatically restored' links, causing a cascade of instabilities. To better understand the process and the effect of this type of failure, we propose a model to study it. As shown in Fig. 2, our model for cascading failures in inter-domain routing system consists of three interconnected components, and iterates along with time. One iteration indicates a step in the cascades.

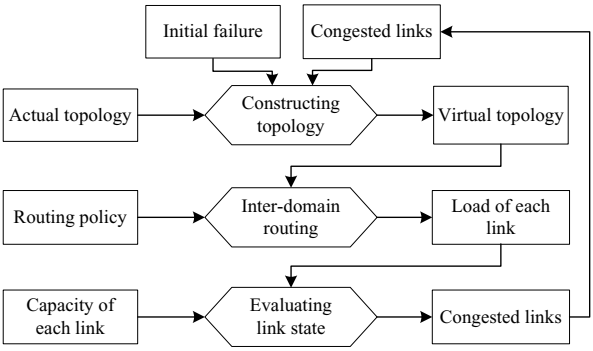

Fig. 2. Model for cascading failures in inter-domain routing system

Constructing Topology. The first component is to construct a virtual topology from the actual topology, initial failure and congested links. We present the actual topology as an annotated undirected graph $G = (V, E)$, where V is the set of all ASes in the inter-domain routing system, and E is the set of AS links annotated with their relationships, which include provider-customer, customer-provider and peer-peer. Since BGP is a policy-based routing protocol and AS relationship is the essential factor to set routing policy, it is important to take this information into account. The initial failure is the disconnection of an AS link $e^{ini} \in E$ at the first place that triggers follow-up instabilities. Congested links, denoted as $E^{con}(t) \subseteq E$, are AS links disconnected by the overload of traffic on them at a certain time t. Therefore, the virtual topology $G'(t) = (V, E'(t))$ is the set of ASes and links that are available to exchange routing messages under those failures at time t, i.e., $E'(t) = E \setminus \{e^{ini}\} \setminus E^{con}(t)$. It is worth noting that the removal of e^{ini} is perpetual while the removal of $E^{con}(t)$ is temporal. $E^{con}(t)$ changes according to different states of links at different time.

Inter-domain Routing. The second component is to simulate inter-domain routing process. The propagation of routing messages is constrained by virtual topology and controlled by routing polices. According to economic considerations of ASes, there are some common points of routing polices summarized by previous research [6]. For import polices, if a BGP router receives routes

to the same destination from different neighbours, it prefers route from customer over those from peer then from provider. Metrics such as path length and other BGP attributes are used in route selection if the preference is the same for different routes. This policy is known as *customer-prefer*. For export polices, an AS does not transmit traffic between any of its providers or peers, which is called *valley-free* property. Under these circumstances, connectivity does not mean reachability in the inter-domain routing system. In our model, we assume that all ASes follow customer-prefer and valley-free polices, and simulate route selections from any source to any destination. Since the inter-domain traffic from source to destination follows the AS path in BGP route, the more AS paths an AS link participates in, the more traffic load the link will transmit. Moreover, the size of source and destination AS should be taken into account. Because large AS usually generates more traffic load. In this paper, we use the number of IPs that an AS announces to assess the size of the AS. Therefore, the load on an AS link is formulated as

$$L_e = \frac{\sum_{u,w \in V} \sigma_{uw}(e) \cdot \varphi(u) \cdot \varphi(w)}{\sum_{u,w \in V} \sigma_{uw} \cdot \varphi(u) \cdot \varphi(w)} \tag{1}$$

The summation is over all ASes in V. $\sigma_{uw}(e)$ denotes the total number of AS paths between u and w that pass through AS link e. σ_{uw} denotes the total number of AS paths between u and w. $\varphi(u)$ and $\varphi(w)$ denote the number of IPs that u and w have. The value of load is normalized into [0, 1]. Accordingly, we could calculate the load on any AS link at any time, denoted as $L_{e_i}(t)$.

Evaluating Link State. The next is to evaluate whether an AS link is congested. The capacity of a link is the maximum load it can handle. We assume that the capacity C_e of link e is proportional to its initial load $L_e(0)$, i.e.,

$$C_e = (1 + \alpha) \cdot L_e(0) \tag{2}$$

where $\alpha \geq 0$ is the tolerance parameter of the network [1]. If the load on an AS link increases and becomes larger than its capacity, the link is congested. The *KEEPALIVE* messages spoken by BGP routers at two ends of link will be dropped along with data packets. As a result, those BGP routers assume that the link is disconnected and end the BGP session between them. Hence we treat congested links as 'virtual cuts' of the network. Meanwhile, if the load on previous congested link decreases below its capacity, the link will 'automatically restore' and be capable of exchanging routing messages again. The current comparison of load and capacity determines congested state of AS link at the next time step in cascades, i.e., $\forall e_i \in E^{con}(t+1), L_{e_i}(t) > C_{e_i}$.

3 Assessing Survivability under the Model

To assess survivability of inter-domain routing system under this cascading model, the first step is to characterize the survivability quantitatively. In this paper, we propose two different metrics to measure the survivability from different perspectives.

Reachability. The virtually cut links entailed by overload could be plentiful at a certain time. On this incomplete topology regarding to the actual one, we wonder if the routing system is still able to find paths for any pair of source and destination. Hence, as shown in Eqn. (3), we define a metric *reachability*, denoted as R to measure this capability of the inter-domain routing system.

$$R(t) = \frac{\sum_{u,w \in V} \chi_{uw}(t)}{N \cdot (N-1)} \tag{3}$$

where $\chi_{uw}(t)$ is equal to 1 if there exist a path from u to w at time t. Otherwise, it is equal to 0. N is the total number of ASes in the Internet, i.e., $N = |V|$. Higher reachability indicates higher survivability of the routing system.

Rerouting Messages. The most essential task for a routing system of networks is to make routing decisions. However, if large amounts of paths need to be rerouted around faulty links, large amounts of BGP messages will be generated, sent and processed. In this case, the computational load on a router's CPU increases dramatically, possibly exceeding the capacity of processors, then weakening the system's ability of routing. So we propose the number of *rerouting messages* that received by *core ASes* at a given time to measure how the cascading failures affect every critical ASes in the Internet. We identify the core ASes by sorting the amount of traffic load that every transit AS transmits for other ASes, then selecting the top one percent ASes to form the core AS set V_c. In the Internet, most ASes are stub ASes rather than transit ASes. They are not our targets for this study. The number of rerouting messages received by u ($u \in V_c$) is defined as

$$RM_u(t) = \sum_{w \in V} \delta_{uw}(t-1, t) \cdot \rho(w) \tag{4}$$

where $\delta_{uw}(t-1, t)$ is the number of paths from u to w that are different at $t-1$ and t. In our model, paths are rerouted only due to changes of virtual topology. $\rho(w)$ is the number of IP prefixes in w. Since in BGP, the routing messages are generated regarding to every IP prefix. The more prefixes an AS has, the more routing messages it will generate when paths targeted to it need to be rerouted. The distribution of RM wrt. every AS in core AS set reveals different effects of the failure on different AS. Generally speaking, more rerouting messages indicates lower survivability of the inter-domain routing system.

In our model, many factors could affect the survivability of the inter-domain routing system. In this paper, we focus on analysing the tolerance parameter α and the initial failure of link e^{ini}. First of all, we examine the relationship between survivability and capacity of AS links to evaluate under what condition a global instability of the Internet will emerge. Secondly, we divide the initial failure as intentional attack and random breakdown, and further analyse the difference of survivability of inter-domain routing system under these two kinds of initial failures.

4 Simulation Results

We build a simulator to simulate the routing dynamics under our model. The topology of the Internet and the AS relationships are inferred from CAIDA's data set [7]. Although the inferred topology doesn't completely agree with the actual Internet, this data set is the most complete and accurate one that is used by present research. The number of IPs and prefixes in every AS are calculated from the BGP routing tables collected by Route Views [8] and RIPE RIS [9]. Many ASes monitored by these two projects distribute in the core of the Internet, so their routing tables cover almost all the routed IP prefixes in the world. More precisely, we construct a connected network with 41204 ASes and 121310 AS links. The tolerance parameter α is set to be 0.1, 0.3, 0.5, 0.7 and 0.9, representing five levels of capacity of links. However, in fact, the capacities of links are various. We simplify the situation in our simulator at first. Then we plan to differentiate each link according to its position in the routing hierarchy in our future work.

The distributions of transmitted load of transit ASes and load of AS links are shown in Fig. 3 and Fig. 4. As we can see, the Internet controlled by inter-domain routing system exhibits a highly heterogeneous distribution of load. We choose 79 candidates from the top 1% of transit ASes to construct core AS set, and examine the rerouting dynamics in 20 time steps, denoted as T. We cut the AS link with highest load, denoted as e_A, to simulate the intentional attack, whereas cut the link e_B at random to simulate the random breakdown. Then we run policy-based BGP among ASes and measure its survivability under different conditions. Next are some results.

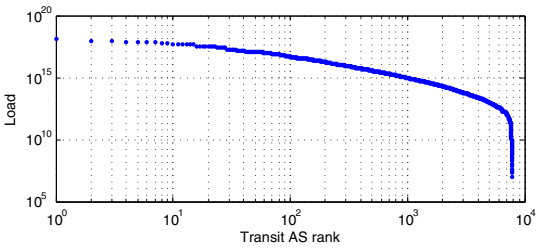

Fig. 3. Load of transit ASes

Fig. 5 and Fig. 6 are reachability on links with different capacities under different initial failures. The first time step of every curve is the reachability under the initial failure. Then it is followed by 20 time steps of consequent reroutings. From the results we can see that no matter what kind of initial failure is, the cascading effect on reachability of inter-domain routing system is amplified when the tolerance parameter α is equivalent to 0.1. It's rational to infer that the effect is also amplified when α is less than 0.1. Meanwhile it is constrained into a very limited scope when α is equal to and greater than

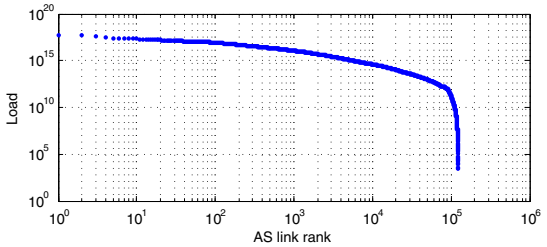

Fig. 4. Load of AS links

0.3. Moreover, as anticipated, with the same tolerance parameter, the cascading effect under intentional attack is greater than that under random breakdown. However, due to the automatically restoration of some faulty links, the difference between these two cases is deflated comparing with previous research [3].

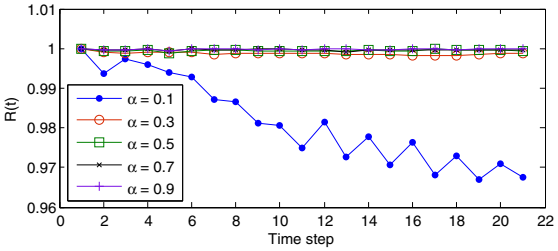

Fig. 5. Reachability under intentional attack

Fig. 7 and Fig. 8 are average numbers of rerouting messages in every time step under different conditions. $RM(t)$ is calculated as $(\sum_{v \in V_c} RM_v(t))/|V_c|$. From this metric we confirm our previous findings that the cascading failures emerge when the tolerance parameter of links is less than 0.1. And a failed link with higher traffic load can cause more added burden on the routing system.

Fig. 9 and Fig. 10 show distributions of rerouting messages associated with every core AS under different conditions. In these cases, we put emphasis on comparing the rerouting messages generated by initial failure and by cascading failures. So we just consider conditions that α is 0.1, 0.5 and 0.9, represented as low, median and high tolerant AS links. In the case of initial failure, RM_v is the number of rerouting messages in the initial time, i.e., the first time step. In the case of cascading failures, RM_v is calculated as $(\sum_{t \in T} RM_v(t))/|T|$, i.e., the average number of rerouting messages during the following 20 time steps. From the distributions we can see that most of core ASes encounter dramatically increased routing messages due to the cascading effect on the routing system. Especially when α is 0.1, the failures tend to affect all the core ASes, causing a global effect. Table 1 shows the number of rerouting messages that the 95th

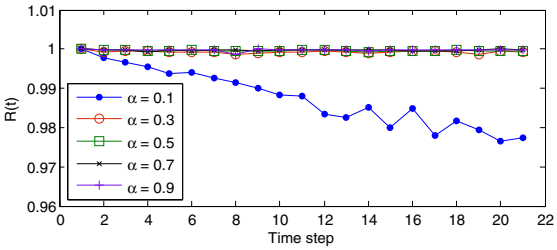

Fig. 6. Reachability under random breakdown

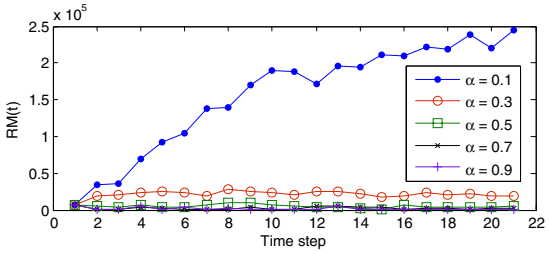

Fig. 7. Average number of rerouting messages under intentional attack

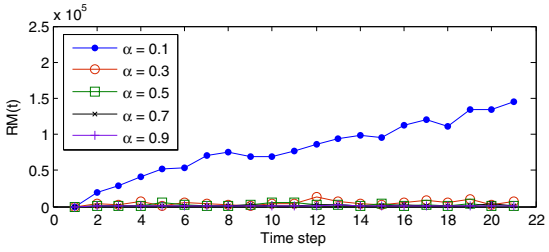

Fig. 8. Average number of rerouting messages under random breakdown

Table 1. Number of rerouting messages received by the 95th percentile core ASes in CDF

	Initial failure	Cascading failures		
		$\alpha = 0.1$	$\alpha = 0.5$	$\alpha = 0.9$
Intentional attack	0.25×10^3	3.47×10^5	6.01×10^3	2.36×10^3
Random breakdown	0	1.21×10^5	3.36×10^3	1.14×10^3

percentile core ASes receive in CDF. The amount of routing messages generated by consequent cascading failures increases by a factor of at least 10 than that only generated by the initial link failure. These additional overloads bring a great deal of added burden to almost all the core ASes, crippling their routers' ability to make routing decisions.

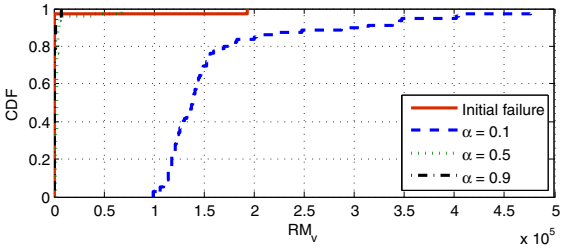

Fig. 9. Distribution of rerouting messages under intentional attack

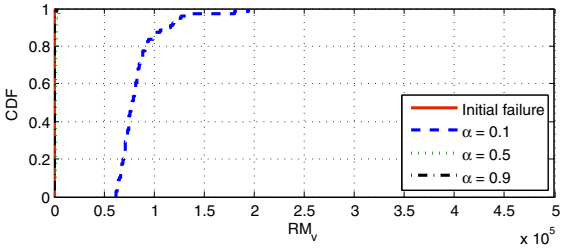

Fig. 10. Distribution of rerouting messages under random breakdown

Moreover, if we further classify the core ASes into three ranks according to the load they transmit for others, we find that the heaviest laden ASes are sorted into median or low ranks. It's rational to infer that the ranked ASes have matching capability to process routing messages. So the large amount of additional overloads on median- and low-rank ASes will have worse effect than that on high-rank ASes.

5 Conclusion and Future Work

In this paper, we present a model for cascading failures in inter-domain routing system, which is proposed for the first time to the best of our knowledge. Then we propose two metrics for measuring the survivability of the inter-domain routing system, and assess the survivability under different conditions based on empirical topology and property of the AS-level Internet.

From the simulation results, we get the following insights. First of all, due to the co-location of data plane and control plane in BGP, the inter-domain routing system is affected by the cascading effect triggered by link failures. This cascading effect brings a great deal of added burden to almost all the core ASes, crippling their ability to make routing decisions. Secondly, the cascading effect is amplified when the tolerance parameter of AS links is less than 0.1. Moreover, the effect triggered by intentional attack is greater than that triggered by random breakdown. But the difference between them is not as prominent as previous research due to the unique automatic-restoration process in inter-domain routing system.

In future work, we will examine the affecting scope of cascading failures topologically, to see whether it spreads over the global Internet or just causes impact to a local area. In addition, we are going to differentiate the capability of each link according to its position in the routing hierarchy. Moreover, it's also important to study the relationship between the initial failed portions of the Internet and its survivability, because the intentional attacks or the random breakdowns may take place to several links or ASes.

Acknowledgement. This research is supported by Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1012); Program for Science and Technology Innovative Research Team in Higher Educational Institutions of Hunan Province (Network Technology, NUDT); Hunan Province Natural Science Foundation of China (11JJ7003); the National Natural Science Foundation of China (Grant Nos. 61070199, 61003303); and the National High Technology Research and Development Program of China (Grant No. 2011AA01A103).

References

1. Shaikh, A., Varma, A., Kalampoukas, L., Dube, R.: Routing Stability in Congested Networks: Experimentation and Analysis. In: SIGCOMM 2000, pp. 163–174. ACM, New York (2000)
2. Xiao, L., He, G., Nahrstedt, K.: Understanding BGP Session Robustness in Bandwidth Saturation Regime. Technical Report, UIUCDCS-R-2004-2483, `http://hdl.handle.net/2142/10918`
3. Motter, A., Lai, Y.: Cascade-based Attacks on Complex Networks. Phys. Rev. E 66, 065102 (2002)

4. Crucitti, P., Latora, V., Marchiori, M.: Model for Cascading Failures in Complex Networks. Phys. Rev. E 69, 045104 (2004)
5. Schuchard, M., Mohaisen, A., Kune, D., Hopper, N., Kim, Y., Vasserman, E.: Losing Control of the Internet: Using the Data Plane to Attack the Control Plane. In: CCS 2010, pp. 726–728. ACM, New York (2010)
6. Gao, L.: On Inferring Autonomous System Relationships in the Internet. IEEE/ACM Transactions on Networking (ToN) 9(6), 733–745 (2001)
7. CAIDA - The Cooperative Association for Internet Data Analysis, `http://www.caida.org`
8. University of Oregon Route Views Project, `http://www.routeviews.org`
9. RIPE Routing Information Service (RIS), `http://www.ripe.net/ris`