The Impacts of Link Failure Location on Routing Dynamics: A Formal Analysis

Xiaoliang Zhao, Beichuan Zhang, Daniel Massey, Andreas Terzis, Lixia Zhang [*]

ABSTRACT

One approach to understanding the complex global routing dynamics is to identify the impact of various factors in the routing system. In this paper we focus on one of these factors, the location of link failures. We build a formal analysis framework to examine whether link failures occurring at the network core and network edges have different impact on routing dynamics, as measured by the number of affected nodes, number of affected routes, and total number of updates. We validate our analytical results by simulations. Our results show that, on average, edge link failures tend to affect more nodes than core link failures, and core links failures tend to affect more routes. As the network grows in size, most of the routing updates will be caused by edge link failures.

1. INTRODUCTION

There exist frequent link failures in today's Internet [1]. Once a failure is detected, individual routers adjust their forwarding path entries affected by the failure and exchange routing updates to inform neighbor nodes about their routing changes. Although these route adjustments and message exchanges follow well defined protocol steps, one cannot infer from protocol specifications about the resulting routing dynamics, especially in a large-scale network where tens of thousands of routers interact through routing message exchanges. It is generally agreed that routing dynamics in the global Internet is rather complex and still beyond comprehension at this time.

In this paper we take an initial step toward understanding routing dynamics in large scale networks by using formal analysis to examine the impact of failure locations, more specifically we are interested in whether link failures at the edge and core of a network lead to different degrees of routing dynamics, as measured by the number of nodes affected, the number of changed routes, and the number of routing updates exchanged between nodes. When a link fails, it may cause routing update messages propagate to the rest of the network. There is a routing "cost" associated with such update propagation. Previous measurements [2, 3] suggest that not all link failures are equal in terms of their routing "cost." Failures occurring at different locations in the network seem to result in different amount of update exchanges and different convergence time. However these previous studies mainly reported the measurement results collected from specific vantage points and during specific time periods, thus it is unclear how general the results may be.

In this paper we take a formal analysis approach which is expected to lead to results that are generally applicable. Formal analysis can provide a solid base of understanding, however it often requires simplifying assumptions to make the problem tractable. These simplifying assumptions may render the results less applicable to real systems. On the other hand, simulations can provide a good approximation to realistic systems, however today's simulation tools cannot handle very large systems such as the Internet. In this work we use simulations to verify the analytical results for small networks, and use the analytical results to draw conclusions about very large networks.

Modeling the operational Internet is a difficult task. In order to keep the problem tractable, we use an incremental approach, starting with a simple model which captures only the most basic properties of the network, then incrementally taking additional factors into account. Our basic model includes a two-tier hierarchical network topology, a routing protocol which is a simplified version of BGP, and a simple routing policy. Based on this model, we analyze the impacts of core link failures and edge link failures. We then extend the basic model by taking into account the number of interconnections between core nodes and the number prefixes announced by core nodes to represent the real system more closely.

Our results show that different locations of link failures do lead to different impact on routing dynamics. The main contributions of this paper can be summarized as follows:

- Generally speaking, an edge link failure affects more nodes than a core link failure.

[*] Xiaoliang Zhao (xleonzhao@gmail.com) is an independent consultant; Beichuan Zhang (bzhang@cs.ucla.edu) and Lixia Zhang (lixia@cs.ucla.edu) are with UCLA; Daniel Massey (massey@cs.colostate.edu) is with Colorado State University; Andreas Terzis (terzis@cs.jhu.edu) is with Johns Hopkins University. This work was done when Xiaoliang, Beichuan and Dan were at USC/ISI.

SIGCOMM ASIA WORKSHOP '05, Apr. 12-14, 2005, Beijing, China.
Copyright 2005 ACM 1-59593-0302/05/0004 ...$5.00.

- On average, a core link failure affects more routes (i.e., the number of routing table entries) than an edge link failure, especially when the number of network core nodes is relatively small compared to the number of edge nodes. However when core nodes are connected by multiple links, a core link failure has on average lower routing impact compared to an edge link failure.
- Based on the current trend that the number of edge nodes grows much faster than the number of core nodes, in a very large network the number of total routing updates will grow in proportion to the numbers of edge nodes and edge links.

The rest of the paper is organized as follows. Section 2 describes the basic network model used in our analysis. Section 3 analyzes the routing impact of core link failures and edge link failures, validates our analytical results by simulations, and examines results in larger networks. In Section 4, we relax some assumptions on core nodes, and re-examine the results. We present related work in Section ref:related and conclude in Section 6.

2. NETWORK MODEL

In this section, we describe our basic network model and the metrics we used to measure the impact of link failures on routing changes. This basic network model makes analysis tractable, while capturing the fundamental characteristics of Internet routing.

2.1 Network Topology

While considerable progress has been made in modeling the properties of the Internet AS graph starting with the seminal work presented in [4], we chose to use a simpler model to keep our analysis tractable. Previous studies [5, 6] have shown that the Internet topology exhibits a hierarchical structure, where a small number of networks, mainly Tier-1 ISPs, form the network core and provide transit services, and the majority of networks form the network edge. Core networks have rich connectivity among themselves, while edge networks have only limited number of connections and rely on the core networks for global reachability. Our topology model captures these essential characteristics by constructing a 2-tier hierarchical structure; we plan to extend this two-tier model to more realistic topologies in our future work.

Because our focus is on the impact of link failures at the inter-domain level, we model each network as a single node which belongs to either the network core or the edge. The set of core nodes, V_c, forms a clique. Each of the edge nodes, $u \in V_e$, is directly attached to one or more core nodes. All the links in the topology can be classified into two classes: *core links*, which connect core nodes, and *edge links*, which connect edge nodes to core nodes. Formally, the topology can be defined as an undirected graph $G(V, E)$ satisfying the following conditions:

- $V = V_c \bigcup V_e, V_c \cap V_e = \emptyset$
- $E = E_c \bigcup E_e, E_c = \{(u, v)|u, v \in V_c\}, E_e = \{(u, v)|u \in V_e, v \in V_c\}$
- $(u, v) \in E_c, \forall u, v \in V_c$

The notations we used to describe the parameters of $G(V, E)$ are listed in Table 1.

Meaning	Notation	
Total number of nodes	N	
Degree of node u	$d(u)$	
	edge	core
number of nodes	N_e	N_c
number of links	L_e	L_c

Table 1: Network Graph Parameters

2.2 Modeling Network Growth

Since the Internet continues to grow, we would like to know how to evolve our topology model to reflect this network growth. Such an accurate model of Internet growth is currently a topic of active research topic. In this paper we adopted a simplified network growth model to capture the essential Internet growth pattern. A number of studies [7, 8] have shown that the growth in the number of edge networks is much faster than that of core networks, and is the major contributor to the Internet growth. Such difference in growth rates is essential to our model. A simple way to differentiate these two rates is by growing the number of edge nodes exponentially while growing the number of core nodes linearly. More precisely, we first pick a small number s as seed and generate a sequence of topologies with $N_c = i * s$ and $N_e = 2^{i-1} * s$. Since we cannot predict the exact growth rate of the Internet, we use $(1 \leq i \leq n)$ (where n is a small constant) to investigate different growth rates.

2.3 Routing Protocol and Policy

To model the routing protocol and policies that control routing information exchanges, we chose to use the Simple Path Vector Protocol (SPVP) [9], which has been used in previous BGP routing studies. Informally, SPVP can be summarized as follows:

- When a node u has multiple paths to node v, u chooses the shortest length path as its best route to v. If there is a tie, the path with the smallest next hop node ID is preferred.
- When the best route changes, u will send a routing update to all its neighbors.

In our model, every node originates its own set of prefixes and advertises these prefixes to all of its neighbors. Every node also constructs a routing table containing the best routes to all other prefixes. Since core nodes form a clique, the best route between any two core nodes is one hop, and the best route between any two edge nodes, in the absence of any failures, is at most three hops (i.e., edge-core-core-edge).

According to [6], in the current Internet, there are two dominant relationships among different networks, namely, customer-provider relationship and peering relationship. These relationships guide routing policy configuration in most cases. In our topology model, if a core node u connects to an edge node v, u will be the provider of v, and v is the customer of u. Any two core nodes are peers. Accordingly, we have the following routing update policies:

- Core nodes provide transit service only to their customers, i.e., a core node only announces its own and its customers prefixes to its peers.

- Edge nodes do not provide transit service, i.e., edge nodes do not send updates for prefixes learned from their providers.

These policies imply that an edge node will only appear as route origin, and link failures do not trigger any edge to generate routing updates.

2.4 Metrics

When a link fails, nodes that were using the disabled link will re-calculate their best routes and send routing updates to neighbors with the new route information. To understand the impact of the location of a link failure on routing, we define three metrics. First, we are interested in *how many nodes need to make routing changes due to the link failure*, which is measured by the *average number of affected nodes*. This metric reflects the scope of the impact. The reason we use averages is because comparison between individual instances of link failures will not lead to a general conclusion. For example, if a particular edge link failure affects more nodes than a particular core link failure, it does not imply other edge link failures will necessarily affect more nodes than core link failures. Therefore, we take average to obtain statistically meaningful results.

Second, we are interested in *how many routes have to be changed*, which is measured by the *average number of affected routes*. Specifically, we count how many best routes need to be replaced due to the link failure. This metric reflects the potential impact on data forwarding since data flows traversing the affected routes may potentially experience packet delays or losses.

Third, we are also interested in *the percentage of triggered updates contributed by edge link failures*, to see which type of link failures triggers more routing traffic. At a high level, if each edge link may fail with probability p_e, each edge link failure triggers t_e updates on average, and there are L_e edge links in total, then the total number of routing updates due to edge link failures is $p_e t_e L_e$. Similarly, the total number of updates due to core link failures is $p_c t_c L_c$. Let $U^e = t_e L_e$, $U^c = t_c L_c$, $\eta_u = \frac{U^c}{U^e}$ and $\eta_p = \frac{p_c}{p_e}$. The percentage of updates due to edge link failures, γ_e, is

$$\gamma_e = \frac{p_e U^e}{p_e U^e + p_c U^c} = \frac{1}{1 + \eta_p \eta_u}$$

In our analysis, we will derive the *total number of triggered updates*, U^e and U^c, and then calculate γ_e.

2.5 Simplifying Assumptions

In the basic model, we have three simplifying assumptions to keep the problem tractable. We subsequently relax these assumptions to make our model more realistic.

First, we assume that all links have the same probability of failure. In practice, it is generally believed that Internet core links tend to be more reliable than edge links. Internet core links correspond to the connection between large ISPs who have resources dedicated to monitoring and management. We relax this assumption in Section 3.6. Second, we assume each node announces a single address prefix. In practice, both core and edge nodes (i.e., an Autonomous System) can originate routes to multiple prefixes, and a large ISP (i.e., core node) tends to originate a large numbers of prefixes. We relax this assumption in Section 4.1. Third, we assume any two nodes are connected by a single link. In reality, two core nodes (i.e., two large ISPs) are often interconnected at multiple locations, which results in multiple physical links between them. We relax this assumption in Section 4.2.

In this work, we do not consider the impact of slow routing convergence when counting the number of updates. After a link failure, BGP may explore multiple transient paths before converging on the new set of stable paths [10, 11, 12]. In general topologies, studies [2] have shown that edge link failures tend to create more transient path exploration thus more updates than core link failures. However, we believe that the topology model and routing policy used in this work have the effect of limiting the occurrence of slow convergence [13]. Examining the impact of slow convergence is part of our future work.

3. ANALYSIS ON BASIC MODEL

3.1 Overview

The failure of link (u, v) affects nodes that use this link to reach some destinations. These nodes will change their routing tables and send out routing updates. Let $r(u, v)$ be the set of nodes who use u to reach v, including node u. When link (u, v) fails, nodes at both sides of the link are affected, thus the total number of affected nodes is $r(u, v) + r(v, u)$. Each affected node at u's side has $r(v, u)$ routing table entries affected, and each node at v's side has $r(u, v)$ routing table entries affected, therefore the number of affected routes is $2 * r(u, v) * r(v, u)$. Since edge nodes do not provide transit service for other nodes, they do not send routing updates after a link failure. Affected core nodes will send one update for each affected routing table entry to all their neighbors except the failed link. For instance, assuming v is the core node, it will send $r(u, v) * (d(v) - 1)$ updates, where $d(v)$ is v's degree,

In our analysis, we will follow the above logic to obtain the metrics for each individual link failure, sum over all target links (i.e., core or edge), and compute the average by dividing the sum by the number of links. Our conclusions are mainly based on the final analytical results, not the proof themselves. Due to page limit, we only show a few proofs to demonstrate main techniques. Complete proofs of all theorems can be found in [13].

3.2 Edge Link Failure

For the failure of an edge link $(u, v | u \in V_e, v \in V_c)$, we have the following analytical results.

Theorem 3.1. *The average number of affected nodes after an edge link failure, $E(|V^e|)$, is*

$$E(|V^e|) = 1 + \frac{N_e}{L_e}(N - 1)$$

Proof. Use $a(u, v)$ to denote the number of nodes affected by an edge link failure of (u, v), then $a(u, v) = 1 + r(v, u)$. If we take the sum $a(u, v)$ over all u's links and all edge nodes, and divide it by the total number of edge links, we get

$$\begin{aligned} E(|V^e|) &= \frac{1}{L_e} \sum_{u \in V_e} \sum_{(u,v) \in E} a(u, v) \\ &= \frac{1}{L_e} \sum_{u \in V_e} \sum_{(u,v) \in E} (1 + r(v, u)) \end{aligned}$$

Since every node except u must use exactly one of u's link to reach u,

$$\sum_{(u,v)\in E} r(v,u) = N-1, \sum_{(u,v)\in E} 1 = d(u)$$

Therefore,

$$\begin{aligned} E(|V^e|) &= \frac{1}{L_e}\sum_{u\in V_e}(d(u)+N-1) \\ &= \frac{1}{L_e}(\sum_{u\in V_e} d(u) + \sum_{u\in V_e}(N-1)) \\ &= \frac{1}{L_e}(L_e+(N-1)N_e) = 1+\frac{N_e}{L_e}(N-1) \end{aligned}$$

□

When all edge nodes are single-homed, which means $N_e = L_e$, any edge link failure will partition an edge node from the rest of the network and affect all nodes, $E(|V^e|) = N$, as expected. As edge nodes increase their connectivity, L_e increases, which reduces the average number of nodes affected by an edge link failure. This theorem shows that N_e/L_e is the quantitative measure of how effective multi-homing is in reducing the scope of an edge link failure's impact.

Theorem 3.2. *The average number of affected routes after an edge link failure, $E(|P^e|)$, is*

$$E(|P^e|) = \frac{2N_e}{L_e}(N-1)$$

Similarly, multi-homing reduces the number of affected routes by increasing L_e.

Theorem 3.3. *The total number of updates summed over all edge link failures, U^e, is*

$$U^e = L_e(N-2)$$

The average number of updates is $N-2$, which implies that for each edge link failure, on average all nodes except u and v will receive *one* update. When all edge nodes are single-homed, this is obvious since all other nodes rely on link (u,v) to reach u and will be notified of its failure by one update. In the case of multi-homing, after (u,v) fails, some nodes may be notified *multiple* times if they are multi-homed, and some nodes may not be notified if u is multi-homed. However, this theorem shows that, on average, each node still receives *one* update. This result demonstrates that formal analysis can help reveal insights that are otherwise difficult to see.

3.3 Core Link Failure

Theorem 3.4. *The average number of affected nodes after a core link failure, $E(|V^c|)$, is*

$$E(|V^c|) = \frac{1}{2L_c}(2N(N_c-1) - \sum_{w\in V_e} d(w)^2 + L_e)$$

When every edge node w is single-homed, $d(w) = 1$ and $N_e = L_e$, which gives $E(|V^c|) = 2N/N_c$. When multi-homing is used, edge nodes depend less on individual core links, therefore the number of affected nodes decreases. Unlike the case of edge link failure, the quantitative effect of multi-homing cannot be decided by L_e alone; it also relies on how the edge links are distributed among edge nodes, as measured by $\sum_{w\in V_e} d(w)^2$.

Theorem 3.5. *The average number of affected routes after a core link failure, $E(|P^c|)$, is*

$$E(|P^c|) = \frac{1}{L_c}(N(N-1) - 2L_e - \beta N_e(N_e-1))$$

PROOF. When a core link (u,v) fails, the number of affected routes is $2*r(u,v)*r(v,u)$. Summing this term over all core links will give a complex formula. Therefore we take another approach to get a more intuitive result. Since there are in total $N(N-1)$ routing entries in the network, we just need to find how many routes are not affected by core link failures and subtract it from $N(N-1)$. Routes that are not affected by core link failures should not contain any core link, which means they belong to one of the following categories: core-edge, edge-core, and edge-core-edge. Routes in the first two categories connect an edge node to a core node, therefore their total number is $2L_e$. Routes in the third category connect two edge nodes via a core node. In total, we have $N_e(N_e-1)$ routes connecting two edge nodes. And two edge nodes are connected either by routes in the form of edge-core-edge or by routes in the form of edge-core-core-edge. Let β be the percentage of the edge-core-edge routes. Therefore, we have totally $N(N-1) - 2L_e - \beta N_e(N_e-1))$ affected paths when failing every core link. By taking the average, we have the final result. □

β is a topology-dependent factor, which measures how much the end-to-end connectivity between edge nodes is independent from core links. For example, if all edge nodes connect to the same core node, none of the routes between two edge nodes need to traverse a core link, thus $\beta = 1$, and we have the least number of affected routes per core link failure.

Theorem 3.6. *The total number of updates summed over all core link failures, U^c, is*

$$U^c = L_e(N-1) - \sum_{u\in V_c} d_e(u)^2$$

where $d_e(u)$ is the number of edge nodes that a core node u connects. When N is fixed, multi-homing will increase both L_e and $\sum_{u\in V_c} d_e(u)^2$, but the latter increases faster since $L_e = \sum_{u\in V_c} d_e(u)$. Therefore, multi-homing will decrease U^c. This is because multi-homing decreases the dependency of edge nodes on individual core links, and in turn, decreases the number of triggered updates after a core link failure.

3.4 Simulation

We use SSFNET [14] in simulation to verify our analytical results in relatively small networks. SSFNET implements the full version of BGP, therefore it simulates various protocol operations and different timers, providing a more realistic scenario to verify the analytical results.

3.4.1 Simulation Settings

In all simulations, BGP parameters are set to default values, e.g., 30 seconds with random jitter for the $MRAI$ timer. After the initialization phase, a link is brought down, which means the BGP session on this link is terminated. This link failure will triggers some routing updates in the network. After the network converges again, we count the number of affected nodes, affected routes, and updates. After repeating this for every link in the network, the average numbers are calculated.

In each topology, core nodes form a clique, and edge nodes are attached to randomly selected core nodes. To simulate the limited connectivity of edge nodes and the increasing practice of multi-homing in the Internet, we set the degree of each edge node to be from 1 to 3. When we increase the network size, we set the network growth seed $s = 6$ and $i \in [1, 5]$.

3.4.2 Simulation Results

We compare the results obtained from simulations and those from calculations based on the theorems. Figure 1 shows that for all three metrics and both edge link failure and core link failure, simulation results and theoretical calculations match each other very well, which shows that our analysis is valid even with various BGP protocol operations and timers.

In calculating the average affected routes for core link failures, β is obtained in the following way. For an edge node u, it connects to multiple core nodes $C = \{c_1, c_2, \ldots, c_m\}$, and each c_i connects a set of edge nodes, so u can reach those nodes via c_i and form an edge-core-edge path. Therefore we have

$$\beta = \frac{1}{N_e(N_e-1)} \sum_{e \in V_e} | \bigcup_{c \in nbr(e)} nbr_e(c)|$$

where $nbr(u)$ is the set of neighboring nodes of u, $nbr_e(u)$ is the set of neighboring edge nodes of u.

3.5 Trends in Large Networks

Since simulation is practical only on relatively small topologies, we need rely on analytical results to understand the routing impacts in large networks. We generate a sequence of network topologies with linear increase of core nodes and exponential increase of edge nodes, calculate their topological parameters, such as N_e, L_e, etc., and then use these parameters to calculate the values for all metrics. The results are shown in Figure 2.

Observation 1. *Figure 2(a) shows that an edge link failure affects more nodes than a core link failure.*

This implies that on average, an edge link failure affects larger scope than a core link failure. This is the result of edge nodes' limited connections compared with core nodes. For an edge node with a small number of links, all the other nodes in the network need to use this small number of links to reach the said node. Therefore, on average, there are a large number of nodes rely on an edge link. When the edge link fails, a large number of nodes will be affected. On the contrary, core nodes usually have much more links, thus there are less nodes rely on each individual core link, and the impact of its failure will have a smaller scope.

Observation 2. *Figure 2(b) shows that an edge link failure affects less routes than a core link failure when the network size is large.*

Since edge nodes grow much faster than core nodes, there will be a large number of edge nodes and a relatively small core in large networks, which means a very large number of routes relying on a relatively small set of core links. This increases the average number of routes affected by core link failures, and becomes much more pronounced when the network is very large.

The total number of updates triggered by link failures are obtained from the following two equations.

$$U^e = L_e(N-2) \qquad U^c = L_e(N-1) - \sum_{u \in V_c} d_e(u)^2$$

Since $\sum_{u \in V_c} d_e(u) = L_e$, we can derive the bounds of $\sum_{u \in V_c} d_e(u)^2$. When all edge links are evenly distributed among all core nodes, $d_e(u) = L_e/N_c$, and $\sum_{u \in V_c} d_e(u)^2$ reaches its minimum value of L_e^2/N_c. When all edge links concentrate on a single core node, it reaches its maximum value of L_e^2. Note $L_e = N_e$ in the latter case. Also, since usually $L_e \cong k \cdot N_e$, where k is a small integer, we can express the bounds of U^c as:

$$L_e(N-1) - L_e^2 \le U^c \le L_e(N-1) - \frac{L_e^2}{N_c}$$

$$L_e(N_c - 1) \le U^c \le L_e(N_c + N_e - 1 - N_e \cdot \frac{k}{N_c})$$

As the network size increases, $N_e \gg N_c$, $U^e \cong L_e N_e$. For U^c, its bounds are approximated by $L_e N_c$ and $L_e N_e$. Therefore, we have

Observation 3. *The number of updates triggered by core link failures depends on how edge links are distributed among the core nodes. If the core nodes have more or less the same number of edge links, core link failures and edge link failures result in similar amount of routing updates; if edge links are distributed unevenly among core nodes, a core link failure tends to trigger fewer updates than an edge failure.*

In simulations, since edge links are attached to randomly selected core nodes, the distribution is close to uniform, and Figure 2(c) confirms that both types of failures trigger similar amount of updates. In the real Internet, the power-law of AS degree suggests that the distribution of edge links on the core may be quite uneven, and as a result, core link failures may trigger much less routing updates than edge link failures.

3.6 The Impact of Failure Probability

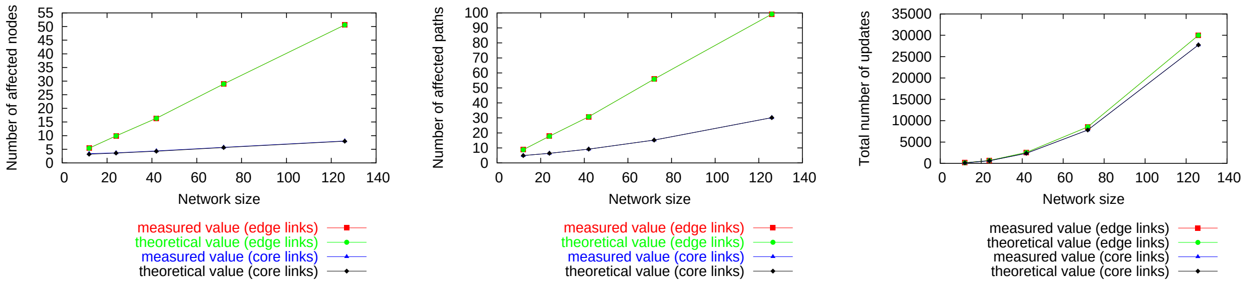
Now we relax the assumption that core links and edge links have the same probability to fail. Since

$$\gamma_e = \frac{1}{1 + \eta_p \eta_u} \quad \eta_p = \frac{p_c}{p_e} \quad \eta_u = \frac{U^c}{U^e}$$

In large networks, $0 \le \eta_u \le 1$. In reality, core links usually are much better maintained than edge links, therefore $\eta_p \ll 1$. In this case, γ_e will approach to 1 in large networks, which means that most routing dynamics will be attributed to edge link failures.

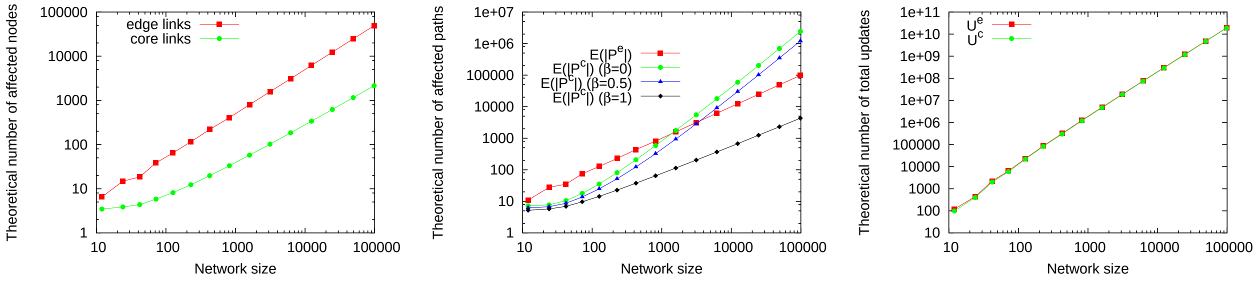
4. EXTENDED MODEL FOR NETWORK CORE

In this section, we relax two assumptions in the basic model to make the network core more realistic: a core node can originate multiple prefixes, and there can be multiple links between core nodes.



(a) Average Affected Nodes (b) Average Affected Paths (c) Total Updates

Figure 1: Simulation Results



(a) Average Affected Nodes (b) Average Affected Routes (c) Total Updates

Figure 2: Large Networks (calculation)

4.1 Multiple Prefixes Originated by Core Nodes

In reality, large ISPs (i.e., core nodes) usually originates much more prefixes than small ISPs or customer networks (i.e., edge nodes). To capture this in our model, we assume that every core node originates $M_c \geq 1$ prefixes. Since routing entries are associated with prefixes, we count routes at per prefix basis.

Using multiple prefixes does not change the results for affected nodes and total updates after edge link failures, because these results are independent to the number of prefixes originated by core nodes.

Theorem 4.1. *The average number of affected routes after an edge link failure is*

$$E(|P^e|) = \frac{1}{L_e}(N_e N_c M_c + N_e(N + N_e - 2))$$

Theorem 4.2. *The average number of affected routes after a core link failure is*

$$E(|P^c|) = \frac{1}{L_c}((N-1)(N_c M_c + N_e) - L_e(M_c + 1) - \beta N_e(N_e - 1))$$

Theorem 4.3. *The total number of triggered updates summed over all core link failures is*

$$U^c = (N_c - 1)L_e M_c + N_e L_e - \sum_{u \in V_c} d_e(u)^2$$

Again, simulations results confirms that our analytical results are valid in small networks [1]. Figure 3 shows the calculation results for large networks, from which we have the following observations.

Observation 4. *When $M_c > 1$, core link failures tend to affect more routes and trigger more updates than edge link failures.*

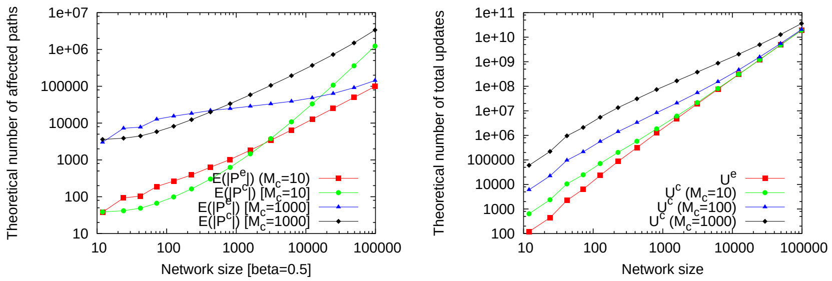
As M_c increases, core links carry more prefixes than edge links. Therefore, when a core link fails, more routes are affected, and consequently, more updates are triggered.

4.2 Multiple Links between Core Nodes

In reality, large ISPs connect with each other at multiple places, such as exchange points and private peering points, resulting in multiple links between two core nodes. To extend our model, we assume there are $K \geq 1$ links between every pair of core nodes.

When a core link fails, there are two ways to re-construct the routes. In Figure 4, there are three core nodes, A, B and C. A has two routers, $A.a$ and $A.b$; B and C also have two routers each. There are two links between every pair of core nodes as shown in the figure. Now, $A.a$ has two paths to reach B, $(A.a, B.a)$ and $(A.a, A.b, B.b)$. Links between two nodes are called "external links," and links within one node are called "internal links."

[1] Complete results can be found in [13].



(a) Average Affected Routes

(b) Total Updates

Figure 3: Large Networks (M_c is the number prefixes originated by each core node)

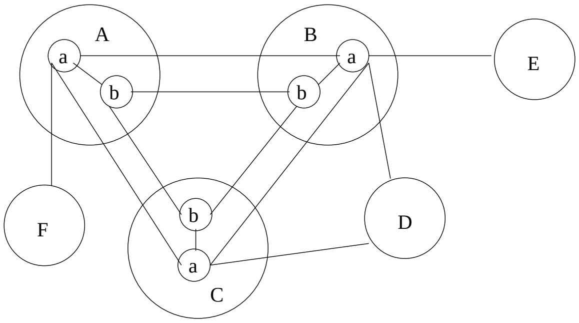


Figure 4: An example of multiple links between core nodes

When link $(A.a, B.a)$ fails, $A.a$ has to switch to the internal link $(A.a, A.b)$ in order to reach B. This change is a "local" change because there is no new path information for other nodes. However, in order to reach D, $A.a$ has two choices. One is going through the internal link $(A, a, A.b)$ first, then through links $(A.b, B.b, B.a, D)$. This choice keeps the original path (A, B, D) at node level and doesn't trigger any update. The other choice is going through an external link $(A.a, C.a)$ first, then through the link $(C.a, D)$. This changes the original path to (A, C, D). Both choices give paths with 3 node-hops. In reality, a BGP router usually tries to direct traffic out of its network via nearest exit. In our example, this means $A.a$ will choose the external link $(A.a, C.a)$ to reach D. In BGP terms, this is often phrased as "prefer paths learned from eBGP peers over those learned from iBGP peers," which results in "hot potato routing." Therefore, we adopt the following policy.

- When a node u has multiple paths to reach the same prefix, u will always prefer external links over internal links.

Multiple links between core nodes does not affect the results of affected nodes and affected routes for edge link failures. For other metrics, we have the following results.

Theorem 4.4. *When $K > 1$, the average number of affected nodes after a core link failure is*

$$E(|V^c|) = 2 + \frac{1}{L_c K}(\alpha\rho(N_e - 1)N_e^{''})$$

where $N_e^{''}$ is the number of multi-homed edge nodes, ρ is the percentage of edge-core-core-edge paths that destine to multi-homed edge nodes, and α is the percentage of unique prefixes edge-core-core-edge of such edge-core-core-edge paths.

This theorem implies that increasing the connectivity between core nodes will decrease the average number of affected nodes after core link failures. It is because that the edge nodes' dependency on core links are now shared by K links. In the extreme case that $K \to \infty$, $E(|V^c|) \to 2$, meaning only the two incident core nodes are affected.

ρ is a topology-dependent factor determined by how edge nodes connect to core nodes. α is determined by the routing preference of edge nodes. Given the same number of paths between edge nodes, if every edge node chooses to load-balancing its traffic, i.e., using different core links to reach different destinations, α will increase. As a result, the average number of affected nodes will increase, since every core link failure will affect nodes that do load-balancing.

Theorem 4.5. *When $K > 1$, the average number of affected routes after a core link failure is:*

$$E(|P^c|) = \frac{1}{L_c K}(K(N_e N_c - L_e) + \rho(N_e - 1)N_e^{''})$$

When $K \to \infty$, $E(|P^c|) \to (2M_c + \frac{N_e N_c - L_e}{L_c})$. This shows that if core nodes continue to increase the number of links among them, the average number of affected paths will approach to a constant value determined by the overall topology.

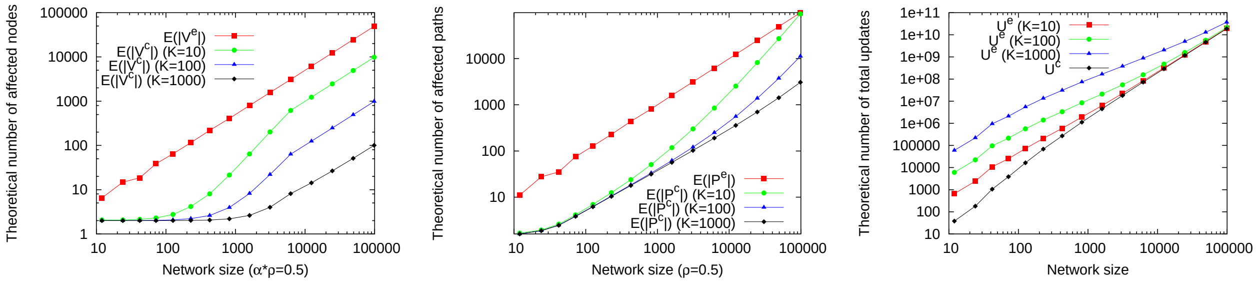
Theorem 4.6. *When $K > 1$, the total number of updates summed over all edge link failures is:*

$$U^e = (N_c - 1)L_e K + N_e L_e - L_e$$

Theorem 4.7. *When $K > 1$, the total number of updates summed over all core link failures is:*

$$U^c = N_e L_e - \sum_{u \in V_c} d_e(u)^2$$

Figure 5 shows the trends in large networks.



(a) Average Affected Nodes

(b) Average Affected Routes

(c) Total Updates

Figure 5: Large Networks (*K* is the number links between two core nodes)

Observation 5. *When $K > 1$, edge link failures affect more nodes and routes than core link failures.*

This is because some core link failures are "local changes", which do not change routing paths at node level nor trigger any update.

Observation 6. *When $K > 1$, edge link failures trigger more updates than core link failures.*

The dominant term of U^e is $(N_c - 1)L_eK + N_eL_e$. K plays a role in U^e because the information about edge link failure will be propagated through all K links between affected core nodes. For U^c, because multiple links keep core nodes connected, it cancels updates due to path changes between two core nodes. Therefore, the K factor amplifies routing overhead when edge links fail, but reduces routing overhead when core links fail.

Observation 7. *N_eL_e becomes the dominant factor on total number of updates in very large networks, for both edge link failures and core link failures.*

Comparing Figure 2(c), 3(b), 5(c), all values tend to converge to the same value when the network size is very large. It is because the dominant term in all equations is N_eL_e when N_e and L_e are growing exponentially.

5. RELATED WORK

A number of previous work studied BGP routing dynamics through data measurements. Labovitz *et. al* [15] found several pathological behaviors by examining BGP update logs and analyzed the possible origins of such behavior in [16]. Other studies examined the BGP dynamics during stressful events such as worm attacks. [17] reported correlation between the surge of BGP traffic and worm activities. In [18], Wang *et. al* showed that worm attacks impacted some edge networks, which could be possible causes for BGP update surge. Another study [19] also showed that worm attacks affected some edge networks like Department of Defense (DoD) networks.

The limitation of data measurement is partially addressed in [20], which showed how difficult it is to interpret operational data. Griffin *et. al* [9] built a formal model to examine BGP dynamic behavior to study especially BGP divergence problem. This model is further extended in [21], which studied the theoretical bound of routing convergence time. This work also used similar routing policies and hierarchical network topologies as used in this paper. Our work also exploits the power of formal analysis, but with a quite different focus. We are interested in different dynamics caused by edge and core link failures. This paper represents an initial step toward a deep understanding of the relation between network failures and routing dynamics.

6. CONCLUSION

As an initial step toward understanding the role of various influential factors in routing dynamics, in this paper we use formal analysis to examine the impact of the location of link failures on routing instability. Our simplified network model differentiates link locations into two classes, edge links and core links. Our analytical results readily show the impact of different link failure locations as measured by the number of affected nodes, the number of affected paths, and the number of updates triggered, identifying the effect of the location of failed links. In addition, our results pinpoint out dominant factors which determine the number of updates as the network size grows.

In the process of achieving the above results we also gained experience in how to combine an analytical approach with simulation validation, to gain deep insights of a large-scale system. As our next step we plan to further verify the initial results through simulations with Internet-like topologies. During this exercise we will study the occurrence of slow convergence in order to capture its impact in our formal model. We also plan to further extend our topology model to better capture the essence of the Internet topology to make the results more applicable to real networks. Furthermore, intra-domain routing protocols can impact routing dynamics as well because of their interactions with inter-domain routing protocols. We can extend our model to define intra-core links and study their routing impacts.

7. ACKNOWLEDGMENT

We thank reviewers for their thoughtful and valuable comments.

8. REFERENCES

[1] G. Iannaccone, C. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, "Analysis of Link Failures in an IP Backbone," in *Proceedings of ACM IMW 2002*, October 2002.

[2] X. Zhao, D. Pei, D. Massey, and L. Zhang, "A Study on Routing Behavior of Latin America Networks," in *IFIP/ACM Latin America Networking Conference*, 2003.

[3] Mohit Lad, Xiaoliang Zhao, Beichuan Zhang, Dan Massey, and Lixia Zhang, "Analysis of BGP Update Surge during Slammer Worm Attack," in *5th International Workshop on Distributed Computing (IWDC)*, 2003.

[4] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos, "On Power-Law Relationship of Internet Topology," in *Proceedings of ACM SIGCOMM*, 1999.

[5] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the internet hierarchy from multiple vantage points," in *Proceedings of the IEEE INFOCOM*, June 2002.

[6] L. Gao, "On inferring automonous system relationships in the internet," *IEEE/ACM Transactions on Networks*, vol. 9, no. 6, 2001.

[7] Tian Bu, Lixin Gao, and Don Towsley, "On Characterizing Routing Table Growth," in *Global Internet*, 2002.

[8] Geoff Huston, "Analyzing the Internet BGP Routing Table," *The Internet Protocol Journal*, mar 2001.

[9] T. Griffin and G. Wilfong, "A Safe Path Vector Protocol," in *Proceedings of IEEE INFOCOMM*, March 2000.

[10] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed Internet Routing Convergence," in *Proceedings of ACM Sigcomm*, August 2000.

[11] D. Pei, X. Zhao, L. Wang, D. Massey, A. Mankin, F. S. Wu, and L. Zhang, "Improving BGP Convergence Through Assertions Approach," in *Proceedings of the IEEE INFOCOM*, June 2002.

[12] A. Bremler-Barr, Y. Afek, and S. Schwarz, "Improved BGP Convergence via Ghost Fluching," in *Proceedings of the IEEE INFOCOM*, April 2003.

[13] Xiaoliang Zhao, , Beichuan Zhang, Dan Massey, Anderis Terzis, and Lixia Zhang, "The Impact of Link Failure Location on Routing Dynamics," Tech. Rep. 04-820, USC, May 2004, URL: `http://www.cs.usc.edu/Research/TechReports/04-820.zip`.

[14] "The SSFNET Project," http://www.ssfnet.org.

[15] C. Labovitz, G. Malan, and F. Jahanian, "Internet Routing Instability," in *Proceedings of ACM Sigcomm*, September 1997.

[16] C. Labovitz, G. Malan, and F. Jahanian, "Origins of Internet Routing Instability," in *Proceedings of the IEEE INFOCOM*, 1999.

[17] J. Cowie, A. Ogielski, B. J. Premore, and Y. Yuan, "Global routing instabilities triggered by Code Red II and Nimda worm attacks," Tech. Rep., Renesys Corporation, Dec 2001.

[18] L. Wang, X. Zhao, D. Pei, R. Bush, D. Massey, A. Mankin, S. Wu, and L. Zhang, "Observation and Analysis of BGP Behavior under Stress," in *Proceedings of the ACM IMW 2002*, October 2002.

[19] X. Zhao, M. Lad, D. Pei, L. Wang, D. Massey, A. Mankin, S. Wu, and L. Zhang, "Understanding BGP Behavior through a Study of DoD Prefixes," in *Proceedings of the IEEE DISCEX III*, 2003.

[20] Tim Griffin, "What is the sound of one route flapping?," in *Network Modeling and Simulation Summer Workshop*, June 2002.

[21] D. Obradovic, "Real-time Model and Convergence Time of BGP," in *Proceedings of the IEEE INFOCOM*, June 2002.