# A Framework for Measuring and Predicting the Impact of Routing Changes

Ying Zhang
University of Michigan
wingying@umich.edu

Z. Morley Mao
University of Michigan
zmao@umich.edu

Jia Wang
AT&T Labs-Research
jiawang@research.att.com

***Abstract*— Routing dynamics heavily influence Internet data plane performance. Existing studies only narrowly focused on a few destinations and did not consider the *predictability* of the impact of routing changes on performance metrics such as reachability. In this work, we propose an efficient framework to capture coarse-grained but important performance degradation as a result of BGP routing events using light-weight probing. We deployed our framework across six vantage points for 11 weeks and found that the data plane experienced serious performance degradation in the form of reachability loss and forwarding loops following a significant fraction of updates affecting many destination prefixes and networks across all vantage points studied. Specifically, more than 39% of updates resulted in reachability loss, some lasting for more than 300 seconds, impacting more than 72% of probed prefixes and more than 35% of all the prefixes on the Internet.**

**We identified that more than half of the prefixes have predictable routing behavior. Based on the stationarity of the correlation between routing changes and the data plane performance, we developed a model to accurately predict the severity of the impact due to routing changes. Such a model is directly helpful for making informed decisions for improved routing schemes such as overlay routing and backup path selection.**

## I. Introduction

Internet routing dynamics directly influence the data plane, *i.e.,* the packet forwarding behavior. Previous measurement studies [1], [2], [3], [4] have already shown that routing changes can cause transient disruption to the data plane in the form of packet loss, increased delay, and forwarding loops. In this work, we enhance our understanding of the impact of routing dynamics on the data plane performance in two dimensions. First, we develop an efficient framework enabling a more comprehensive study of routing changes that are not limited to just specific prefixes as in previous studies. Second, we identify the predictability of observed performance degradation in relation to the properties of routing updates and subsequently develop a model to accurately predict the performance impact of future updates.

We use the term *data plane failures* to describe severe performance degradation on packet forwarding manifested as reachability loss or forwarding loops. Our study focuses on data plane failures primarily caused by routing changes, as understanding the impact of routing dynamics on data plane performance is critical to the deployment of real-time applications such as Voice over IP (VoIP) and moreover provides insights into improved network operations.

Routing changes on the Internet are mostly caused by failures or configuration changes. They occur quite frequently. At the interdomain level, one can easily observe more than 10 updates per second to a wide range of destinations from a large tier-1 ISP such as Sprint using publicly available BGP data from RouteViews [5]. Motivated by such active routing dynamics on the current Internet, our study develops a methodology to identify properties of updates that cause data plane failures and characterize the location, duration, and stability of these failures.

Data plane failures are often caused by inconsistent forwarding information of routers involved in routing changes [2]. During routing convergence, some routers may lose their routes [6] or have invalid routes [1]. Routing policies, timer configurations, and network topologies are just some of the contributing factors [2], [6]. For instance, transient loops can be caused by temporarily inconsistent views among routers. Persistent loops are more likely due to misconfigurations [7]. We do not attempt to identify the cause of observed failures due to lack of information but instead search for patterns to help predict the impact of routing changes on data plane performance. Such a prediction model can improve route selection.

To achieve a comprehensive characterization of many diverse routing changes, we develop an efficient and novel measurement framework deployed at each vantage point with access to real-time BGP routing updates. Light-weight probing is triggered by locally observed routing updates. The probing target is an identified live IP address within the prefix associated with the routing change. Compared to modeling or simulation based approaches [8], [9], [10] to understand the impact routing dynamics on data plane performance, our measurement-based approach does not make simplifying assumptions and provide empirical evidence of such impact.

Given that probing is triggered directly by routing updates, it may be counter-intuitive why the observed data plane performance may still be impacted by the seemingly converged route. In some cases, the routing change is still ongoing, often manifested by subsequent updates to the same destination prefix. Given the scale of the Internet, some routing changes may impact many routers and cause delayed convergence [1]. Thus, even if locally the route to a destination appears to be converged to a stable route, data plane performance may still be seriously affected. This is supported by previous work showing that BGP messages sometimes preceded observed path failures in the order of minutes [11].

We deployed our measurement framework at six geograph-

ically distinct locations with different upstream providers for a period of 11 weeks. Using our collected set of 604,925 live IPs which belong to 48% of prefixes and 53% of ASes, we analyzed 47%-55% of all observed updates corresponding to 46%-51% of observed prefixes in routing updates across different vantage points.

We summarize our main findings by including a range of results to represent all six vantage points studied.

- Many prefixes became unreachable shortly after respective routing changes. They account for 39%-45% of probed updates, covering 72%-86% of probed prefixes. These prefixes belong to 35%-42% of all announced prefixes, originating from 39%-42% of all ASes. Stub ASes are more likely impacted. Unreachable incidences are usually transient: 84%-91% of them lasting less than 300 seconds. The failure location occurs roughly equally likely along the path.
- Among the unreachable incidences, a non-negligible fraction exhibits forwarding loops. This contributes to 4%-8% of probed updates, covering 36%-51% of probed prefixes. These loops impact 17%-24% of all announced prefixes, originating from 27%-34% of all ASes. Most loops are short-lived: 60% of them lasting less than 300 seconds. Loops are more likely to appear within large ISPs.
- Given a prefix and its identified responsible AS where traceroute stops or loop occurs, we identify over 51%-54% of probed updates to be predictable for causing reachability loss, and 49%-58% for causing loops. For such prefixes, our prediction model achieves a prediction accuracy of 90% with a false positive rate of 15% for unreachable incidences and a prediction accuracy of 80% with a false positive rate of 12% for loops. In general, prefixes originating from stub ASes and smaller ISPs are more predictable; responsible ASes for such predictable prefixes also tend to be near the edge of the Internet.

Aside from measurement findings, our main contribution is a framework to efficiently measure the impact of routing dynamics on data plane performance. Based on identified inherent stability of routing changes, we develop a methodology to predict impact of future routing updates. The ability to accurately predict routing-induced data plane failures is directly useful for applications such as overlay route selection and backup path selection.

This paper is organized as follows. Section II introduces our measurement methodology. Experiment setup is described in Section III. We provide detailed data analysis on probing results in Section IV. In Section V, we present a prediction model. We discuss related work in Section VI and conclude in Section VII.

## II. Measurement Methodology

We describe our measurement methodology to enable efficient characterization of the impact of locally observed BGP routing updates on the data plane performance from the local network to the relevant destination networks.

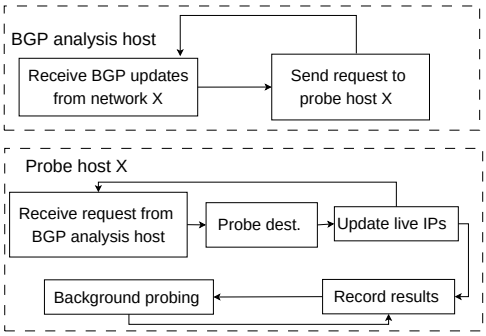


Fig. 1. Active probing architecture for vantage point X (both functionalities can be implemented on the same host).

### A. Terminology

We first introduce our terminology. We use the term *data plane* to refer to the packet forwarding behavior on the Internet. *data plane failures* describe severe data plane performance degradation in the form of reachability loss or forwarding loops. The *control plane* computes the routing state of network elements performing packet forwarding. On today's Internet, inter-domain routing involves distributed router computation within routers of different networks.

To describe probing results, we use the term *probing incidence* to mean a set of probes to the relevant destination prefix triggered by a BGP update of the prefix. Three ping requests optionally followed by a traceroute probe are sent for each prefix probed. The destination is deemed *reachable* if any ping reply returns or the traceroute response contains interface IPs belonging to the prefix. It is *unreachable* otherwise.

### B. Data Collection

There are two required data sources: control-plane BGP updates and data plane active probes. For each monitored location, local real-time BGP data are analyzed to identify probing destinations. BGP data can be obtained by setting up a monitoring BGP session using software such as Zebra [12] with a BGP router with a default-free routing table in the local network. To differentiate between unreachable destinations and blocked probes due to firewalls, we must identify at least one *live IP* that responds to ping or traceroute requests for each prefix probed. Besides active probing [13], such data can be gathered passively from various server logs, *e.g.,* Web and DNS server logs, or traffic traces.

### C. Active Probing Methodology

Figure 1 depicts the probing architecture for one vantage point consisting of a BGP analysis host identifying probe targets based on the local BGP feed and a probe host in the same network for performing probing triggered by routing updates. The list of live IPs is continuously updated. To identify persistent failures and verify live IPs' responses, background probing is done.

*1) Probing Methodology:* Unlike previous studies, our probing is designed to be light-weight to scale to many destinations covering most observed updates. Therefore, we

| Category | Tier-1 | Tier-2 | Tier-3 | Tier-4 | Tier-5 |
|---|---|---|---|---|---|
| Num of ASes (Pctg relative to all ASes in each tier) | 20 (90%) | 173 (80%) | 1092 (78%) | 1235 (80%) | 7136 (52%) |
| Num of prefixes | 3045 | 4672 | 10034 | 9424 | 16727 |
| Num of IPs | 73670 | 119136 | 134982 | 126818 | 116643 |

TABLE I

DIVERSITY OF NETWORKS COVERED BY OUR COLLECTED LIVE IPS.

focus on coarse-grained performance metrics associated with reachability. We are nevertheless limited to probing only prefixes for which we have identified a live IP. We plan to remedy this in the future.

We describe the detailed probing steps. Triggered by a routing update, three ICMP-based ping requests are first sent to the corresponding live IP. We randomly choose the IPs belonging to the given prefix and regularly update IP liveness. Three is chosen to balance the overhead and packet loss probability. If any ping reply returns, the destination is considered *reachable*. Otherwise, traceroute is performed. If the traceroute response contains an IP belonging to the probe destination prefix, the destination is deemed reachable. Otherwise, ping and traceroute probes are continuously sent after each other as soon as the previous probe finishes, until the destination or a timing limit is reached as described later.

*2) Probing Control:* Given the potential high frequency of routing updates, we take measures to avoid overloading the probe host and the destination networks probed. The resources under consideration are CPU and memory resources of the probe host, and network bandwidth of both the probe host and targets. Multiple probe hosts can be used. We make explicit trade-offs between probing coverage and consumed resources.

The first measure is to ignore routing updates caused by the BGP session reset of the monitoring session using known techniques such as [14], as such updates do not reflect true routing changes. As a second measure, we impose a limit on the *maximum probing duration* for each destination prefix. Probing is performed as long as the target is deemed unreachable until this limit is reached. Moreover, at most one IP from each prefix is probed by a single host at any time.

Probe requests may not be serviced immediately due to unfinished probing. We impose a *maximum wait time* between the time an update is received and the time its probe request is initiated, as excessive delays prevent us from effectively capturing the impact of routing changes. As future work, we plan to explore other ways to reduce probing overhead, *e.g.*, by probing based on unique AS paths.

## III. EXPERIMENT SETUP

We describe the experiment setup based on our measurement methodology.

### A. Data Collection

We set up a software router using Zebra [12] to serve as the BGP monitor to obtain live BGP feeds from six distinct locations with different upstream providers mostly in the U.S.: Michigan, Massachusetts, New York, Illinois, Washington, and Amsterdam. They belong to the PlanetLab [15] and the RON project [16]. Combining active probing [13], DNS logs, and five days of Netflow data from a Tier-1 ISP network, we collected 604,925 live IPs covering 48% of all announced prefixes and 53% of all ASes. Using the tier ranking defined in [17], where a lower tier means large ISPs and tier-5 refers to stub or customer ASes, we illustrate the diversity of collected IPs in Table I. The set is shown to cover a large percentage of ASes in different tiers. The results presented span an 11-week period from May 3 until July 19, 2006.

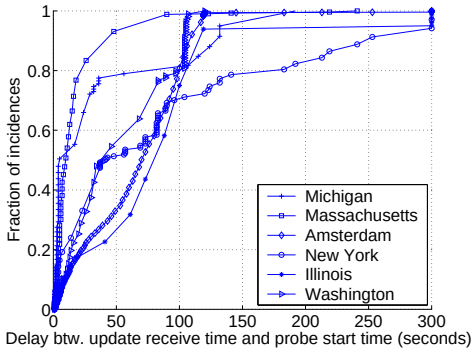


Fig. 2. Probing delay distribution for each BGP feed: Most delays are within 100 seconds.

### B. Probing Control

We limit the maximum probing duration to be 300 seconds as most BGP routing changes converge within about three minutes based on previous studies [1], [3]. Our own measurements described later in Section IV also show that about 90% of reachability problems last less than 300 seconds. To ensure our characterization captures the effect of routing dynamics on the data plane, we limit the maximum wait time to be 300 seconds. Background probing is performed to ensure each live IP is probed at least once every 300 seconds.

### C. Probing System Performance

During the 11 weeks of study, the average probing rate is only about 2 updates per second for each feed with a maximum rate of 11 updates per second. Probing duration varies from less than 10 seconds to the limit of 300 seconds.

Figure 2 plots the distribution of probing delays for each probing location. The delay is computed as the time difference between the probe time and the update receive time. The figure shows that at least 80% of updates are probed within 100 seconds for most feeds. For some locations, the delays are mostly between 50 to 100 seconds. Only 6% of updates are not probed due to the maximum wait time constraint.

To prevent aggressive probing, we measure the probing rate. We found that 80% of the difference between two consecutive probes for the same IP is larger than 300 seconds, with a minimum difference of around 100 seconds. This shows that our system did not overload the destination networks probed.

### D. Probing System Limitations

We discuss the limitations of our probing methodology to understand the potential bias introduced in our results.

| | | Incidence | Prefix | AS |
|---|---|---|---|---|
| Unreachable | Loop | 185728 (6.0%) | 21821 (23.9%) | 5024 (33.5%) |
| | Other | 1129014 (36.3%) | 66321 (72.8%) | 5802 (38.7%) |
| | All | 1314742 (42.3%) | 66883 (73.5%) | 9559 (63.0%) |
| Reachable | | 1796392 (57.7%) | 75578 (83.1%) | 14870 (98.0%) |

TABLE II
GENERAL STATISTICS OVER THE PERIOD OF 11 WEEKS

First, the data presented later correspond to probing triggered by routing announcements only. We also probed after route withdrawals, as such prefixes can still be reachable due to covering prefixes: 1.4%-2.1% of withdrawn prefixes are reachable, while less than .013% of withdrawn prefixes are unreachable despite the presence of covering prefixes. But most of them recover within 300 seconds. Second, our probe delays are mostly within 100 seconds. Thus, we focus on serious data plane failures lasting for at least 100 seconds.

The third limitation is that we do not differentiate between performance degradation due to routing changes from other possibly unrelated causes such as congestion. Given that our probing immediately follows routing updates, the observed performance degradation could also coincide with other events. However, if a destination consistently experiences performance degradation following routing changes, such degradation may likely be caused by routing dynamics.

Although our probing uses simple ping and traceroute probes, we try to overcome limitations of measurement tools. For example, we distinguish unreachable cases caused by routers disabling ICMP replies from unreachable end hosts using history information.

## IV. CHARACTERIZING DATA PLANE FAILURES

During a routing event such as link failures or recoveries, packet forwarding is likely disrupted. This is likely caused by some routers temporarily losing their routes to the destination. Moreover, even without transient failures in the control plane, *i.e.,* every router has a route to the destination, the route may not be valid due to routing inconsistency. Next, we characterize data plane transient failures using "reachability" as the performance metric. This is motivated by the fact that gain or loss reachability will cause the most severe impact on data plane performance.

### A. Overall Statistics

We conducted Internet experiments over the period of 11 weeks from May 3, 2006 to July 19, 2006. Table II shows the overall statistics. We found that 42% of probing incidences are unreachable, affecting 73.5% of destination prefixes and 63% of destination ASes probed in our experiments. In addition, about 14% of the unreachable incidences are caused by loops, affecting 24% of destination prefixes and 34% of ASes probed.

### B. Reachability Failures

*1) Destination Networks Impacted by Failures:* We classify destination ASes experiencing reachability loss according to their tiers and geographic locations. Table III shows the top 10 destination ASes which encounter the most unreachable incidences. We observe that most of them are stub ASes, *i.e.,* customer ASes. Moreover, we found that many unreachable incidences affect a small portion of destination prefixes and ASes observed in our routing updates. For example, as shown in Figure 3, 80% of unreachable incidences impact only 30% of prefixes and 10% of ASes, respectively.

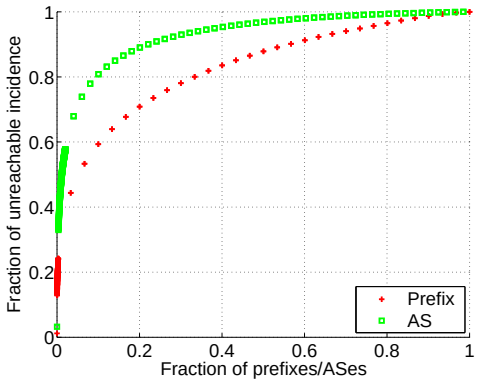


Fig. 3. Destination prefixes and ASes affected by reachability problems.

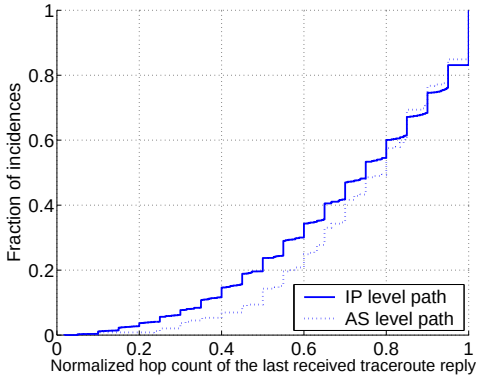


Fig. 4. Normalized hop distance btw. the source and the last received traceroute reply.

Identifying the failure location along the path helps us understand whether the problem usually happens close to the destination networks. If failures occur near or within destination networks, multi-homing or overlay routing cannot bypass such failures. We approximate the location of a data plane failure as the IP hop where the traceroute probe stops.

Figure 4 shows that the normalized hop count is evenly distributed along both the IP level and AS level path. The hop distance is normalized by the hop count of the reachable path before the incidence. Note that the last hop of the stopped traceroute may not be where the problem resides since absence of traceroute replies may be due to firewalls or routers disabling ICMP replies. We differentiate such cases by examining whether routers in a particular AS ever replied with ICMP packets in history data. Such an AS is expected appear in the data path based on history or BGP data. Furthermore,

| ASN | Unreachable Incidences | Prefixes | AS Name | Tier | Primary Country |
|---|---|---|---|---|---|
| 25543 | 112784 (8.6%) | 34 | FasoNet-AS ONATEL/FasoNet's Autonomous System | 5 | Burkina Faso |
| 4134 | 110787 (8.4%) | 590 | CHINANET-BACKBONE No.31, Jin-rong Street | 2 | China |
| 19982 | 107709 (8.1%) | 3 | TOWERSTREAM-PROV Towerstream | 4 | United States |
| 8866 | 45840 (3.4%) | 72 | BTC-AS Bulgarian Telecommunication Company | 3 | Bulgaria |
| 9121 | 43021 (3.2%) | 423 | TTnet Autonomous System | 3 | TURKEY |
| 8011 | 41768 (3.1%) | 39 | CoreComm - Voyager, Inc. | 4 | United States |
| 22543 | 37267 (2.8%) | 16 | PIXELWEB Pixelweb | 5 | Canada |
| 4595 | 36300 (2.7%) | 8 | ICNET ICNet/Innovative Concepts | 5 | United States |
| 17974 | 35573 (2.7%) | 369 | TELKOMNET-AS2-AP PT TELEKOMUNIKASI INDONESIA | 5 | Indonesia |
| 4314 | 28951 (2.2%) | 20 | COMMNET-ASN CommNet Data Systems, Inc. | 3 | United States |

TABLE III

TOP 10 DESTINATION ASES EXPERIENCING MOST UNREACHABLE INCIDENCES.

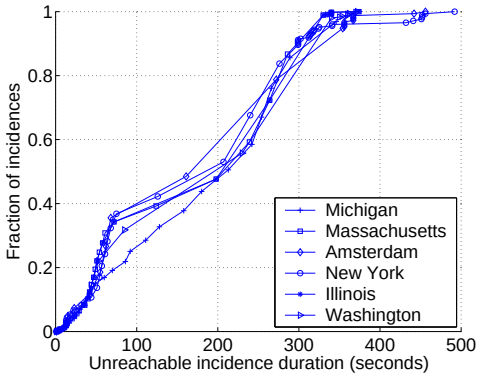


Fig. 5. Duration of unreachable incidences.

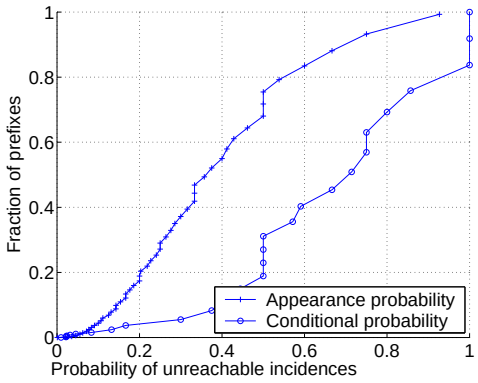


Fig. 6. Appearance probability and conditional probability (conditioned on the responsible AS) of unreachable incidences.

we can usually assume that an AS applies a uniform policy regarding ICMP for all its routers [18].

*2) Failure Duration:* We compute the duration of reachability loss to be the period starting from the time when the update is received to the time that the destination is reachable by probing. Figure 5 shows the cumulative distribution of the duration of unreachable incidences. We found that most such incidences last than 300 seconds. They are likely due to transient routing failures [6] or routing convergence delays. However, 10% unreachable incidences last longer than the maximum probing limit of 300 seconds. They may be caused by other factors such as configuration errors and path failures. The observed reachability disruption lasting a few hundred seconds is expected to have serious performance impact on real-time applications such as Voice over IP.

*3) Failure Predictability:* Routing incidences and their corresponding impact on certain destination networks can be predictable. For a given destination prefix $D$, we define the *appearance probability* of $D$ as the probability of an unreachable incidence occurring with any routing update to $D$. We define the *conditional probability* of $D$ conditioned on an AS or an AS path segment as the probability of an unreachable incidence occurring under the condition of observing a routing update to $D$ through a particular AS or an AS path segment. Moreover, we define the *responsible AS* for an unreachable incidence to be the AS where traceroute stops.

Figure 6 shows the CDF of the appearance probability and the conditional probability conditioned on the responsible AS. Around 30% of the prefixes have unreachable appearance probability of larger than 0.5. This indicates that the reachability loss is difficult to predict for most prefixes upon observing a routing update of that prefix. However, the corresponding plot for the conditional probability (conditioned on the responsible AS) is about 80%. This indicates that, given a routing update to a destination and the responsible AS, unreachable incidences can be much more predictable. By comparing the looping AS path with the normal AS path obtained from background probing, we can estimate the responsible AS's AS level hop count to the destination. 95.9% of these responsible ASes are at least one hop away from the destination. Therefore, taking alternate path might be possible to bypass the problem.

### C. Forwarding Loops

We now focus on a subset of unreachable incidences – forwarding loops, which have been widely studied [19], [20], [4], [21]. It has been shown that transient loops can be caused by inconsistent or incomplete views among routers during routing convergence [22], while persistent loops are more likely a result of configuration errors [7]. In our experiments, we identify loops in traceroute and compare path from background probing with path from triggered probing to detect persistent loops and to exclude loops caused by measurement artifacts. We find only 0.027% forwarding loops are persistent. We focus on transient loops in the rest of this section.

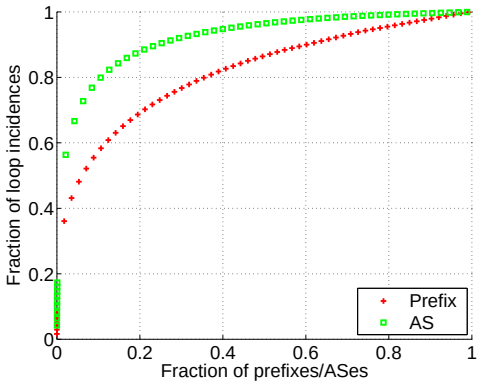


Fig. 7. Destination prefixes and ASes affected by forwarding loops.

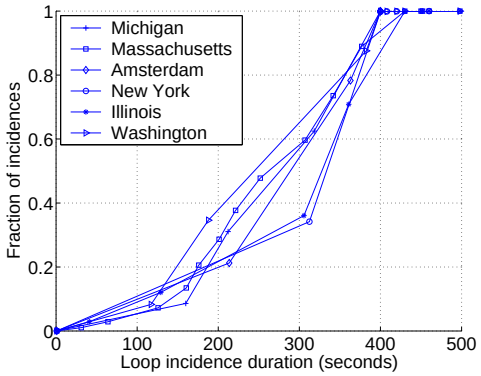


Fig. 8. Duration of loop incidences.

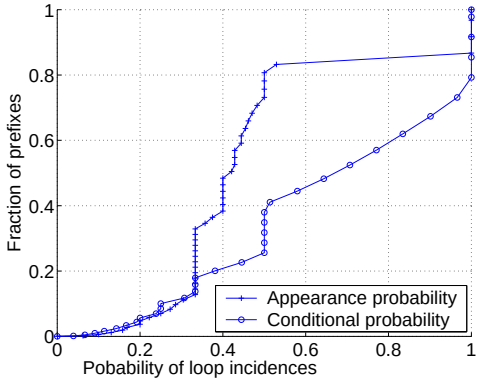


Fig. 9. Appearance probability and conditional probability (conditioned on the responsible AS) of loop incidences.

*1) Destination Networks Impacted by Loops:* Figure 7 shows the fraction of destination prefixes and ASes impacted by forwarding loops. Similar to unreachable incidences, we observe that the distribution of loop incidences across destination prefixes and ASes are very skewed. For example, top 10% of prefixes and ASes observed in our routing updates experience 60% and 80% of forwarding loops, respectively.

For each loop incidence, we consider the ASes where the loop occurs as the *responsible ASes*. We observe that 98% of the loop incidences are intra-AS loops, *i.e.,* the IPs involved in the loop are within one AS. Table IV shows the top 10 responsible ASes for loop incidences. Interestingly, we observe that most of these ASes are tier-1 ASes. This is because large ASes in the core of the Internet have more complicated routing policies, potentially more complex routing dynamics, and larger network diameters translate to longer delays for propagating updates. All these factors can cause more transient failures within such networks [6], [19].

*2) Loop Duration:* We measure the loop duration as the time period from the receipt of the routing update until when probes can reach the destination without experiencing loops. Figure 8 shows about 70% loops last less than 350 seconds. Note that for loops lasting longer than 300 seconds from the first probe, we overcome our maximum probing duration of 300 seconds by background probing to such long-lasting loops to determine whether they are persistent loops. We found that only 0.0027% loop incidences are persistent loops, 74% of which occur close to the destination networks. In addition, we observe that the vast majority of loops involves a small number of IP level hops. For example, 81% of loops involve two IP addresses.

*3) Loop Predictability:* Similarly, we study how predicable loop incidences are. The appearance probability and conditional probability (conditioned on the responsible AS) of loop incidences are shown in Figure 9. Only around 20% of prefixes have appearance probability of more than 0.5, indicating that loop incidences are difficult to predict for most prefixes based simply on the presence of any update to the prefix. However, 75% of prefixes have conditional probability (conditioned on responsible AS) of more than 0.5. This illustrates that, given a routing update to the prefix and the responsible AS, loop incidences can be much more predictable.

## V. Failure Prediction Model

How well does a routing update indicate the occurrence of a data plane failure? Can we detect the presence of such a failure based on observed routing updates? To answer these questions, we develop a statistical prediction model to infer the probability of a data plane failure given a routing update. As observed in our experiments, the predictability of failure incidences given routing updates across all prefixes follows a bi-modal distribution: some prefixes are highly predictable, while others are not. In this section, we focus on prefixes which are more predictable as analyzed in Section IV. We first present our prediction model and then verify the model via supervised learning. Finally, we discuss applications of the prediction model.

### A. Prediction Model

In this section, we derive a model for predicting whether a failure incidence $Y$ occurs upon observing a routing update $R$ to a given destination.

*1) Model:* We use the random variable $Y$ to represent the data plane observation: $Y = 1$ if there is a failure in the data plane, and $Y = 0$ otherwise. We use the random variable

| ASN | Loop Incidences | Destination Prefixes | Responsible AS Name | Tier | Primary Country |
|---|---|---|---|---|---|
| 701 | 34457 (18.6%) | 2059 | ALTERNET-AS UUNET Technologies, Inc. | 1 | United States |
| 1239 | 33998 (17.7%) | 2013 | SPRINTLINK Sprint | 1 | United States |
| 3356 | 32674 (17.6%) | 1998 | LEVEL3 Level 3 Communications, LLC | 1 | United States |
| 7018 | 27971 (15.1%) | 1587 | ATT-INTERNET4 AT&T WorldNet Services | 1 | United States |
| 174 | 21060 (11.3%) | 1149 | PSINET PSINet Inc. | 1 | United States |
| 2914 | 13612 (7.3%) | 787 | VERIO Verio, Inc. | 1 | United States |
| 4134 | 13362 (7.2%) | 534 | CHINANET-BACKBONE No.31, Jin-rong Street | 2 | China |
| 6453 | 13106 (7.0%) | 746 | TELEGLOBE-AS Teleglobe Inc | 1 | United States |
| 3549 | 12267 (6.6%) | 850 | GBLX Global Crossing | 1 | United States |
| 3561 | 12087 (6.5%) | 691 | CWUSA Cable & Wireless USA | 1 | United States |

TABLE IV

FORWARDING LOOP INCIDENCES IN THE TOP 10 RESPONSIBLE ASES.

$R$ to represent routing updates with AS path $x_1, \ldots, x_n$. The model is built based on observations of <failure $Y$, routing update $R>$ pairs in the history data.

In our model, we use a direct acyclic graph (DAG) to represent all the paths for each destination prefix. Each node in the graph represents an AS. In addition, we assume that failures are independent. To determine whether a data plane failure will occur, *i.e.,* $Y = 1$ given a routing update $R$, we compute the data plane failure likelihood ratio.

$$\Lambda(Y) = \frac{P(Y = 1|R; D)}{P(Y = 0|R; D)} \quad (1)$$

where $P(Y = 1|R; D)$ is the conditional probability of data plane failure occurrence given a routing update $R$ for prefix $D$, and $P(Y = 0|R; D)$ is the conditional probability of no data plane failure occurrence given a routing update $R$ for prefix $D$. We say that a data plane failure occurs if $\Lambda(Y) > \lambda$, where $\lambda$ is a decision threshold which determines false positive and negative rate.

Given an update $R$ with the AS path $x_1, x_2, \ldots, x_n$, if a failure occurs in $x_b$, then the ASes along the path can be classified to three categories:

- ASes $x_1, \ldots, x_{b-1}$ appearing in the path before $x_b$ are "good" AS nodes;
- AS $x_b$ is a "bad" AS node, also known as the responsible AS.
- ASes $x_{b+1}, \ldots, x_n$ appearing after $x_b$ in the path are "unknown" AS nodes.

Therefore, the probability of AS $x_i$ being a bad node for destination $D$ can be computed as

$$P(Y = 1|x_i; D) = \frac{BadCount(x_i)}{TotalCount(x_i)} \quad (2)$$

where $BadCount(x_i)$ is the number of occurrences AS $x_i$ appears as a bad node for destination $D$, and $TotalCount(x_i)$ is the total number of occurrences AS $x_i$ appears in the path for destination $D$.

Thus, given a routing update $R$ with AS path $x_1, \ldots, x_n$ for destination $D$, the probability that $R$ will cause a data plane failure is

$$P(Y = 1|R = x_1, x_2, \ldots, x_n; D) = 1 - \prod_{i=1}^{n}(1 - P(Y = 1|x_i; D)) \quad (3)$$

Similarly, the probability that $R$ will not cause a data plane failure is

$$P(Y = 0|R = x_1, x_2, \ldots, x_n; D) = \prod_{i=1}^{n}(1 - P(Y = 1|x_i; D) \quad (4)$$

After computing the failure likelihood ratio $\Lambda(Y)$, we use the receiver operating characteristic (ROC) in signal detection theory [23] to decide the value of $\lambda$. ROC curves are commonly used to evaluate prediction results. In particular, the ROC of a predictor shows the trade-off between selectivity and sensitivity. A curve of false positives ratio (false alarms) versus true positive ratio (detection accuracy) is plotted while varying a sensitivity or threshold parameter. In our experiment, given $\Lambda(Y)$, we determine the ratio of false positive, $P_{FP}$, and the ratio of detection accuracy $P_{AC}$, with varying values of $\lambda$.

*2) Validation:* We evaluate both false positive ratio and false negative ratio of our prediction model. A false positive refers to the case where our prediction model predicts a data plane failure given a routing update, while there is no failure observed in our experiment. A false negative refers to the case where our prediction model fails to predict a data plane failure given a routing update. As we have observed in Figures 6 and 9, some prefixes are more predictable than others. The poor predictability on certain prefixes could be explained by inherent non-stationary properties associated with certain failures, or by the limited visibility from the vantage points of our experiments. Next, we analyze the predictability across different prefixes and focus on the set of more predictable prefixes to further evaluate our prediction model.

We repeat the following experiments 10 times. We first divide the data set into the training set and the testing set. In particular, we randomly sample 50% of the entire observations as the training set and compute the failure likelihood ratio for all the routing updates in the test set. During the training process, we only consider observations that appear at least $k$ times for a given prefix and a responsible AS. In our experiment, we choose $k = 3$. As a result, we discard 5.6% of observations. Using $k = 4$ will increase the prediction accuracy by 0.34%, while discarding 1.3% of observations.

Next, we compute the average $\Lambda(Y)$ of each prefix for both unreachable and reachable incidences based on our observations. Figure 10 shows that the prediction accuracy is limited considering all observed prefixes. Given $\lambda = 1$,

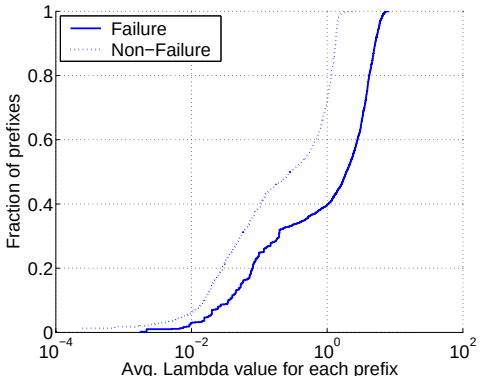


Fig. 10. Average value of $\Lambda$ for each prefix

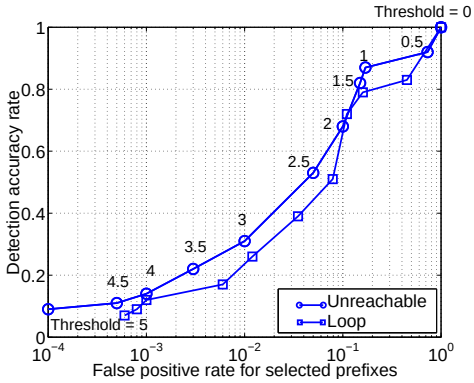


Fig. 11. Receiver operating characteristics for selected subset of prefixes.

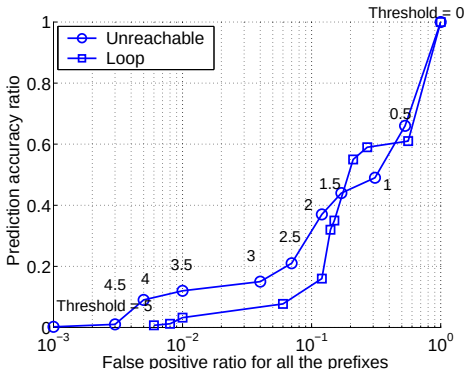


Fig. 12. Receiver operating characteristics for all probed prefixes.

61% of prefixes are predictable (*i.e.,* $\Lambda > 1$) for all failure incidences and 72% of prefixes are predictable (*i.e.,* $\Lambda < 1$) for all non-failure incidences.

Given a prefix and its identified responsible AS, we identify over 51.2%-54.3% of probed updates to be predictable for causing reachability loss and 48.9%-57.5% for causing loops across all six vantage points. The corresponding figures for probed prefixes are 58.7%-67.5% and 53.2%-55.8%, respectively. These destination prefixes account for 28.1%-32.4% of announced prefixes originating from 27.4%-31.9% of all ASes. 3.8% and 5.1% of such destination ASes are tier-1 and tier-2 ASes respectively. The figures for tier-3, tier-4, and tier-5 ASes are 22.4%, 23.6%, and 45.1%, respectively. The set of responsible ASes for unreachable and loop incidences consists of 10.8% tier-1 ASes, 11.9% tier-2 ASes, 19.7% tier-3 ASes, 21.2% tier-4 ASes, and 36.4% tier-5 ASes. This shows that prefixes from the edge of the Internet are more predictable and most responsible ASes are also from the edge.

Figure 11 shows the receiver operating characteristics curve of predicting the incidences in the test set. The false positive ratio is shown in $x$-axis and the prediction accuracy ratio is shown in $y$-axis. We observe that, by varying $\lambda$, our prediction model achieves different degrees of accuracy. For example, with $\lambda = 1$ (*i.e.,* if $\Lambda(Y) > 1$, failure is predicted to occur), our model can achieve 89.8% prediction accuracy with 14.5% false positives for unreachable failures and 79.9% accuracy with 12.3% false positives for loops on the subset of prefixes selected above. This observation implies that the prediction model built on history observation can be used to predict future failures on certain prefixes. Figure 12 shows the corresponding curves for all prefixes. Given $\lambda = 1$, our model achieves 51% and 60% prediction accuracy for unreachable failures and loops with false positive ratio of 21% and 18%, respectively. This is consistent with our observation in Figure 10 that the predictability in general is limited. However, compared to existing work [24] on predicting data plane performance degradations with only 50% prediction accuracy with 60% false positives, our model is much more accurate.

### B. Discussion

In this section, we discuss potential applications of the measurement framework and the prediction model. First, the measurement infrastructure provides a platform of measuring the impact of routing changes on data plane performance. Detecting control plane changes and predicting corresponding data plane disruptions provide additional information for best route selection. By examining the predicted data plane performance among all available routes, the least impacted route can be selected as the best route to reduce the likelihood and degree of data plane performance degradation.

Let us use an example to illustrate how the route selection process can be improved for overlay routing. Suppose a given destination prefix can be reached via multiple overlay nodes. When a failure is predicted in AS $A$ based on the observation of a routing update of the destination prefix, we can select the next hop for a path avoiding $A$ to reach the destination. Compared to random next hop selection proposed in [25], this selection process is more deterministic and has a higher chance of avoiding data plane failures.

## VI. Related Work

A significant number of measurement studies have been conducted to examine the impact of routing changes on data plane performance degradation [11], [2], [6], [21], [26], [27], [28]. For example, [28] focused on the stability of the path

between two ISPs by artificially injecting routing failures. The duration and location of end-to-end path failures are studied and correlated with BGP routing instability in [11]. Recent work [2] analyzes how routing events affect end-to-end Internet performance and explores the root cause of data plane performance degradation. Our work focuses on exploring the coarse-grained performance degradation in terms of reachability caused by routing changes. We measure the data plane performance via active probing triggered by routing updates.

Data plane transient failures are also widely studied in [6], [21], [19], [20], [4]. It has been shown that transient loops can be caused by inconsistent or incomplete views among routers during routing convergence [22], while persistent loops are more likely a result of misconfiguration and can be explored to create flooding attacks [7]. In [29], light-weight data plane countermeasures are used to detect routing protocol and data plane attacks, which can be used in the routing architecture. Our work uses a wide range of measurements in analyzing the impact of the data plane failures triggered by routing updates.

To mitigate forwarding failures, previous research [25], [30] use reactive routing to discover and bypass the failure. Resilient Overlay Networks (RON) [31], [32] and PlanetLab [15] provides a platform to re-route the packets in an overlay network. Our paper provides a prediction model that helps select best routes which are least likely impacted by failures.

## VII. Conclusion

In this paper, we develop an efficient framework to measure and predict data plane performance degradation as a result of routing changes. Using this framework, we conducted a large scale Internet measurement study and characterized data plane performance upon receiving a BGP routing update. Our experiments and analysis cover a large portion of the announced prefixes and ASes on the Internet. We observe that the data plane performance of a certain set of prefixes is highly predictable. We further develop a statistical model which can accurately predict the severity of potential data plane failures based on observations of routing updates for a given prefix. We show that our model is very useful in a number of applications such as route selection in an overlay network.

## References


[1] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed Internet Routing Convergence," in *Proc. ACM SIGCOMM*, 2000.

[2] F. Wang, Z. M. Mao, J. Wang, L. Gao, and R. Bush, "A Measurement Study on the Impact of Routing Events on End-to-End Internet Path Performance," in *Proc. ACM SIGCOMM*, 2006.

[3] Z. M. Mao, R. Bush, T. G. Griffin, and M. Roughan, "BGP beacons," in *Proc. ACM SIGCOMM Internet Measurement Conference*, 2003.

[4] A. Sridharan, Sue.B.Moon, and C. Diot, "On the Correlation between Route Dynamics and Routing," in *Proc. ACM SIGCOMM Internet Measurement Conference*, October 2003.

[5] "University of Oregon Route Views Archive Project." `http://www.routeview.org`.

[6] F. Wang, L. Gao, J. Wang, and J. Qiu, "On Understanding of Transient Interdomain Routing Failures," in *Proc. International Conference on Network Protocols*, 2005.

[7] J. Xia, L. Gao, and T. Fei, "Flooding Attacks by Exploiting Persistent Forwarding Loops," in *Proc. ACM SIGCOMM Internet Measurement Conference*, 2005.

[8] D. Pei, L. Wang, D. Massey, S. F. Wu, and LixiaZhang, "A Study of Packet Delivery Performance during Routing Convergence," in *Proc. of IEEE International Conference on Dependable Systems and Networks (DSN)*, 2003.

[9] L. Z. Beichuan Zhang, Daniel Massey, "Destination Reachability and BGP Convergence Time," in *Proc. of IEEE Globecom, Global Internet and Next Generation Networks*, 2004.

[10] X. Zhao, B. Zhang, A. Terzis, D. Massey, and L. Zhang, "The Impact of Link Failure Location on Routing Dynamics: A Formal Analysis," in *Proc. of ACM SIGCOMM Asia Workshop*, 2005.

[11] N. Feamster, D. G. Andersen, H. Balakrishnan, and M. F. Kaashoek, "Measuring the Effects of Internet Path Faults on Reactive Routing," in *Proc. ACM SIGMETRICS*, Jun 2003.

[12] "GNU Zebra-routing software." `http://www.zebra.org`.

[13] A. Zeitoun and S. Jamin, "Rapid Exploration of Internet Live Address Space Using Optimal Discovery Path," in *Proc. Global Communications Conference*, 2003.

[14] B. Zhang, V. Kambhampati, M. Lad, D. Massey, and L. Zhang, "Identifying BGP Routing Table Transfers," in *Proc. SIGCOMM Mining the Network Data (MineNet) Workshop*, August 2005.

[15] "PlanetLab." `http://www.planet-lab.org`.

[16] D. Andersen, H. Balakrishnan, M. Kaashoek, and R. Morris, "Resilient Overlay Networks," in *Proc. Symposium on Operating Systems Principles*, 2001.

[17] L. Subramanian, S. Agarwal, J. Rexford, and R. H. Katz, "Characterizing the Internet hierarchy from multiple vantage points," in *Proc. IEEE INFOCOM*, 2002.

[18] Neil Spring and Ratul Mahajan and David Wetherall, "Measuring ISP Topologies with Rocketfuel," in *Proc. ACM SIGCOMM*, 2002.

[19] U. Hengartner, S. Moon, R. Mortier, and C. Diot, "Detection and analysis of routing loops in packet traces," in *Proc. ACM SIGCOMM Internet Measurement Workshop*, 2002.

[20] V. Paxson, "End-to-end routing behavior in the Internet," in *Proc. the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, vol. 26,4 of *ACM SIGCOMM Computer Communication Review*, ACM Press, 1996.

[21] D. Pei, X. Zhao, D. Massey, and L. Zhang, "A Study of BGP Path Vector Route Looping Behavior," in *Proc. IEEE International Conference on Distributed Computing Systems*, 2004.

[22] Z. Zhong, R. Keralapura, S. Nelakuditi, Y. Yu, J. Wang, C.-N. Chuah, and S. Lee, "Avoiding Transient Loops through Interface-Specific Forwarding," in *Proc. IFIP/IEEE IWQoS*, June 2005.

[23] J. P. Egan, *Signal Detection Theory and ROC Analysis*. New York: Academic Press, 1975.

[24] A. Bremler-Barr, E. Cohen, H. Kaplan, and Y. Mansour, "Predicting and Bypassing End-to-End Internet Service Degradations," in *Proc. ACM SIGCOMM Internet Measurement Conference*, 2002.

[25] K. P. Gummadi, H. V. Madhyastha, S. D. Gribble, H. M. Levy, and D. Wetherall, "Improving the Reliability of Internet Paths with One-hop Source Routing," in *Proc. Symposium on Operating Systems Design and Implementation*, 2004.

[26] S. Agarwal, C. Chuah, S. Bhattacharyya, and C. Diot, "Impact of BGP Dynamics on Intra-Domain Traffic," in *Proc. ACM SIGMETRICS*, 2004.

[27] J. Li, R. Bush, Z. M. Mao, T. Griffin, M. Roughan, D. Stutzbach, and E. Purpus, "Watching Data Streams Toward a Multi-Homed Sink Under Routing Changes Introduced by a BGP Beacon," in *Proc. Passive and Active Measurement Workshop*, 2006.

[28] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed internet routing convergence," in *Proc. ACM SIGCOMM*, 2000.

[29] I. Avramopoulos and J. Rexford, "Stealth probing: Efficient data-plane security for IP routing," in *Proc. USENIX Annual Technical Conference*, 2006.

[30] M. Zhang, C. Zhang, V. Pai, L. Peterson, and R. Wang, "PlanetSeer: Internet Path Failure Monitoring and Characterization in Wide-Area Services.," in *Proc. Symposium on Operating Systems Design and Implementation*, 2004.

[31] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris, "Resilient Overlay Networks," in *Proc. Symposium on Operating Systems Principles*, 2001.

[32] D. Andersen, A. Snoeren, and H. Balakrishnan, "Best-Path vs. Multi-Path Overlay Routing," in *Proc. ACM SIGCOMM Internet Measurement Conference*, 2003.