Impact of Prefix-Match Changes on IP Reachability

Yaping Zhu, Jennifer Rexford
Princeton University
{yapingz,jrex}@cs.princeton.edu

Subhabrata Sen, Aman Shaikh
AT&T Labs–Research
{sen,ashaikh}@research.att.com

ABSTRACT

Although most studies of Internet *routing* treat each IP address block (or prefix) independently, the relationship between prefixes is important because routers ultimately *forward* packets based on the "longest-matching prefix." In fact, the most-specific prefix for a given destination address may change over time, as BGP routes are announced and withdrawn. Even if the most-specific route is withdrawn, routers may still be able to deliver packets to the destination using a less-specific route. In this paper, we analyze BGP update messages and Netflow traffic traces from a large ISP to characterize both the changes to the longest-matching prefix over time and the resulting effects on end-to-end reachability of the destination hosts. To drive our analysis, we design and implement an efficient online algorithm for tracking changes in the longest-matching prefix for each IP address. We analyze the BGP message traces to identify the reasons for prefix-match changes, including failures, route flapping, sub-prefix hijacking, and load-balancing policies. Our preliminary analysis of the Netflow data suggests that the relationship between BGP updates and IP reachability is sometimes counterintuitive.

Categories and Subject Descriptors

C.2.2 [**Network Protocols**]: Routing protocols; C.4 [**Performance of Systems**]: Measurement techniques

General Terms

Measurement

Keywords

BGP, Longest-matching Prefix, IP reachability

1. INTRODUCTION

Internet routing protocols, such as the Border Gateway Protocol (BGP), compute routes for each address block (or prefix) independently. However, a destination host may fall within the range of addresses covered by multiple prefixes with different mask lengths. Nesting of prefixes is quite common for a variety of reasons. For example, regional Internet registries allocate large address blocks to Internet Service Providers (ISPs), who in turn allocate smaller blocks to their customers. Customers that connect to the Internet at multiple locations may further sub-divide these address blocks to exert fine-grained control over load balancing and backup routes. ISPs may also announce multiple blocks to protect themselves from route hijacking—for example, AT&T announces 12.0.0.0/9 and 12.128.0.0/9, in addition to the 12.0.0.0/8 supernet, to prevent other ASes from accidently hijacking traffic intended for destinations in 12.0.0.0/8. Ultimately, routers forward packets based on the longest prefix (i.e., largest mask length) that matches the destination IP address. However, this "longest-matching prefix" may change over time as BGP routes are announced and withdrawn, leading to a sometimes complex relationship between BGP routing changes and IP packet forwarding.

Understanding how routing changes affect the longest-matching prefix is important for researchers and practitioners alike. Prefix-match changes can affect the accuracy of measurement results. For example, measurement studies often aggregate traffic statistics or performance results to the prefix level, based on a static snapshot of a BGP routing table. However, this kind of analysis is not robust to prefix-match changes that affect the flow of traffic to the destinations. Analyzing BGP update messages without regard to prefix nesting can also lead to misleading conclusions. For instance, a withdrawal does not necessarily imply that the destinations have become unreachable, as they may be reachable via a less-specific route. Prefix-match changes are especially important in network troubleshooting, where a mistake in aggregating or interpreting measurement data may prevent network administrators from correctly diagnosing the cause of traffic shifts, performance degradation, or lost reachability.

In this paper, we analyze the effects of BGP routing changes on the longest-matching prefix. The problem is challenging because we cannot rely on prefixes as the building block for our analysis. Instead, we design and implement an efficient online algorithm for tracking prefix-match changes for each IP address. To make our algorithm scalable, we group addresses into *ranges* that are dynamically split as smaller prefixes are announced. We apply our algorithm to BGP update traces from a large ISP and characterize the frequency and causes of prefix-match changes. We find that

IMC'09, November 4–6, 2009, Chicago, Illinois, USA.

more than 30% of BGP updates do *not* simply switch an existing prefix from one route to another: In fact, 14.8% of the BGP updates cause addresses to gain or lose reachability, and 13.0% of the updates cause addresses to switch to a different longest-matching prefix. These prefix-match changes have a variety of causes, including route flapping, sub-prefix hijacking, and failover to backup routes. To understand the effects of prefix-match changes on end-to-end reachability, we present a preliminary analysis of Netflow traces that shows that traffic sometimes continues to flow using a less-specific prefix.

The rest of the paper is organized as follows. In Section 2, we briefly characterize prefix nesting based on a static snapshot of a BGP routing table. Then, Section 3 introduces our online algorithm for tracking changes in the longest-matching prefix. In Section 4, we apply the algorithm to one month of BGP updates to analyze the frequency and causes of prefix-match changes, and present our preliminary analysis of the Netflow traces. We present related work in Section 5 and conclude the paper in Section 6.

2. STATIC ANALYSIS OF PREFIX NESTING

To understand the nesting of prefixes, we analyze a BGP routing table collected from a router in a large ISP on February 1, 2009. We ignore small prefixes (with mask longer than /24) corresponding to the ISP's own routers and links, as they are not externally visible. We characterize prefix nesting from two perspectives: (i) how many prefixes cover each IP address? and (ii) what fraction of addresses covered by a prefix actually use that prefix for packet forwarding?

The light bars in Figure 1 plot the distribution of the number of prefixes covering each IP address, with a logarithmic scale on the y-axis. While 75.8% of IP addresses are covered by a single prefix, 19.7% are covered by two prefixes, and 4.0% by three prefixes; some addresses are covered by as many as *seven* prefixes. In addition, destination addresses that match multiple prefixes are responsible for a higher fraction of the traffic, relative to other destinations, as seen by the dark bars in Figure 1. These bars plots the distribution weighted by the volume of traffic collected from the same router. While 61.6% of the traffic is destined to addresses matching a single prefix, 31.3% of the traffic corresponds to two prefixes, and 6.0% to three prefixes. We see similar trends for both histograms across a variety of routers and time periods for data collected in the same ISP.

We also explore what fraction of the IP addresses covered by a prefix use that prefix for packet forwarding. We use the same routing table snapshot for this analysis, which was taken on February 01, 2009. Table 1 shows the results for five sets of prefixes, grouped by mask length. Interestingly, 17% of the /8 prefixes are not the longest-matching prefix for *any* of the addresses they cover; the 12.0.0.0/8 prefix mentioned in Section 1 is one example. In fact, 39% of the /8 prefixes handle forwarding for less than half of their addresses, as seen by summing the first three rows of the "/8" column in Table 1. For smaller prefixes (with larger mask lengths), the prefixes are responsible for a larger fraction of the IP addresses they contain. Because we filtered the prefixes with mask length larger than 24 for this analysis, the /24 prefixes are the longest-matching prefix for all of their IP addresses. We saw similar results when analyzing a routing-table snapshot taken on March 01, 2009.

The nesting of prefixes suggests that BGP update messages may change which prefix is used to forward traffic to particular destination addresses. In the following sections, we track the evolution of the longest-matching prefix to understand when and how BGP routing changes affect the forwarding of IP packets.

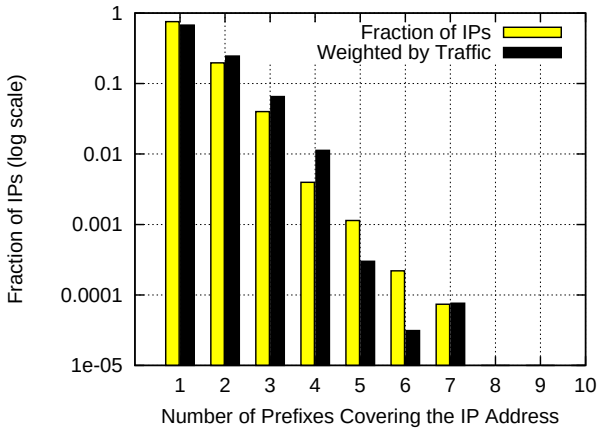

Figure 1: Distribution of number of matching prefixes (from a BGP routing table at 00:00:00 GMT Feb 01, 2009)

Fraction of IP Addresses	Prefix Mask Lengths /8	/12	/16	/20	/24
0	0.17	0.16	0.09	0.04	0.00
(0, 0.25]	0.13	0.14	0.02	0.02	0.00
(0.25, 0.5]	0.09	0.06	0.03	0.03	0.00
(0.5, 0.75]	0.09	0.06	0.03	0.04	0.00
(0.75, 0.9]	0.13	0.05	0.03	0.07	0.00
(0.9, 1]	0.39	0.53	0.81	0.80	1.00

Table 1: Prefix coverage for different mask lengths (from a BGP routing table at 00:00:00 GMT Feb 01, 2009)

3. TRACKING PREFIX MATCH CHANGES

In this section, we present an online algorithm for tracking changes in the longest-matching prefix, and the associated BGP route, for each destination IP address. We first introduce the notion of an *address range* to group IP addresses that have the same set of matching prefixes. Then, we describe our algorithms for updating the address ranges to track changes to the longest-matching prefix.

3.1 Data Structure for Address Ranges

Because of the nesting of prefixes, an IP address could match several prefixes with different mask lengths. In order to track prefix-match changes over time, we need to store information about changes to all prefixes covering the IP address. We refer to the collection of all matching prefixes for a given IP address as its *prefix set*; packet forwarding is driven by the longest-matching prefix in the set. For example, suppose a BGP routing table contains prefixes 12.0.0.0/8 and 12.0.0.0/16. Then, IP address 12.0.0.0 has the prefix set {/8, /16}. IP address 12.0.0.1 also matches the same pre-

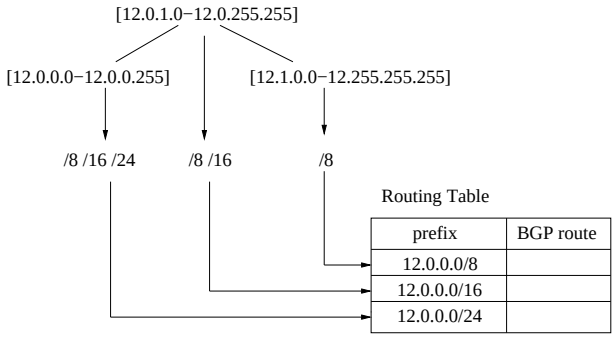

Figure 2: Storing address ranges and prefix sets for prefixes 12/8, 12/16, and 12/24

fixes. However, the prefix set for 12.1.0.1 is {/8}. Rather than tracking the prefix set for each individual IP address, we group contiguous addresses that have the same prefix set into an *address range*. For example, prefixes 12.0.0.0/8 and 12.0.0.0/16 divide the IP address space into two address ranges—[12.0.0.0, 12.0.255.255] with prefix set {/8, /16} and [12.1.0.0, 12.255.255.255] with prefix set {/8}. Note that address ranges differ from prefixes in that the boundaries of an address range are not necessarily powers of two. For instance, no single prefix could represent all IP addresses in the range [12.1.0.0, 12.255.255.255].

As we process BGP update messages, address ranges may be created, subdivided or updated. For ease of searching for the affected address range(s), we store information about address ranges in a binary tree, as shown in Figure 2. A binary tree efficiently supports all the operations we need (including inserting a new address range, lookup an address range) in an average time of $O(log\, n)$, where n is the number of address ranges. The node of the binary tree contains left-most address in the address range, and each node keeps pointer to the size of the address range and the associated prefix set. Each element of the prefix set includes a pointer to the BGP route for that prefix; to save memory, we store a single copy of each BGP route. As illustrated in Figure 2, both address ranges [12.0.0.0, 12.0.0.255] and [12.0.1.0, 12.0.255.255] have prefix 12.0.0.0/16 in the prefix set, and their prefix sets store the pointers to the route entry for 12.0.0.0/16. Note that in the figure, we only plot the pointers from the most-specific prefixes to the routing table for illustration.

3.2 Tracking Changes to Address Ranges

Next, we present an online algorithm that reads BGP table dumps or update messages as input, and tracks the changes to the address ranges and their associated prefix sets. The algorithm first determines the address range(s) covered by the prefix, perhaps creating new address ranges or subdividing existing ones. Then, for each of the associated address ranges, the algorithm modifies the prefix set as needed.

Updating address ranges: A BGP announcement for a new prefix may require creating new address ranges or subdividing existing ones. For example, suppose 18.0.0.0/16 is announced for the first time, and no earlier announcements covered any part of the 18.0.0.0/16 address space; then, our algorithm inserts a new address range [18.0.0.0,18.0.255.255], with a prefix set of {/16}, into the binary tree. As another example, suppose we have previously seen route announcements only for 12.0.0.0/8 and 12.0.0.0/16; then, the binary tree would contain [12.0.0.0,12.0.255.255] with prefix set {/8, /16}, and [12.1.0.0,12.255.255.255] with prefix set {/8}. On processing an announcement for 12.0.0.0/24, our algorithm would subdivide [12.0.0.0,12.0.255.255] into two address ranges—one with prefix set {/8, /16, /24} and another with {/8, /16}, as shown in Figure 2. Currently, our algorithm does not delete or merge address ranges after withdrawal messages. We take this lazy approach towards deleting and merging address ranges because withdrawn prefixes are often announced again later, and because we have seen empirically that the number of address ranges increases very slowly over time.

Updating prefix set for address ranges: Continuing with the example in Figure 2, suppose the route for 12.0.0.0/16 is withdrawn. Then, the algorithm would determine that both [12.0.0.0-12.0.0.255] and [12.0.1.0,12.0.255.255] have /16 removed from the prefix set. For addresses in [12.0.1.0-12.0.255.255], the withdrawal would change the longest matching prefix to the less specific 12.0.0.0/8.

4. DYNAMICS OF PREFIX-MATCH CHANGES

In this section, we apply our algorithm to BGP update messages collected for the month of February 2009 from a top-level route reflector in a tier-1 ISP backbone. We first determine the frequency of the BGP updates which affect the longest-matching prefix for IP address ranges. Then, we study four main categories of prefix-match changes, based on the origin ASes (i.e., the AS that introduces the prefix into BGP) of the two prefixes and how often the more-specific prefix is available. Finally, we present a preliminary analysis of Netflow data to understand the impact of prefix-match changes on end-to-end reachability.

4.1 Frequency of Prefix-Match Changes

The BGP update messages from the top-level route reflector give us a view of BGP routing changes seen at a large Point-of-Presence (PoP) in the ISP backbone. Our algorithm starts by reading a BGP table dump taken at the beginning of the month, followed by the stream of BGP update messages. We filter duplicate update messages, including those sent after resets of our monitoring session [1] to the route reflector. We also filter updates caused by route flapping, where a prefix is repeatedly announced and withdrawn for a long period of time. As in previous work [2], we group update messages for the same prefix that occur with an interarrival time of less than 70 seconds, assuming these updates are part of the same BGP convergence event. Since most convergence events last less than five minutes [2], we assume longer events correspond to persistent flapping, and remove these flapping updates from further analysis. This filtered 25,120 BGP updates caused by route flapping, which account for 0.21% of the total number of BGP updates in that month.

We find four main categories of BGP update messages, as summarized in Table 2:

Updating a route for an existing prefix: Just under 70% of the update messages are announcements that merely change the route for an existing IP prefix, as indicated by the first row of the table. These update do not affect the longest-matching prefix used for forwarding data packets.

Category	% Updates
Same prefix, route change	69.5%
Gain reachability	7.4%
Lose reachability	7.4%
More-specific prefix	6.5%
Less-specific prefix	6.5%
No impact announcements	2.3%
No impact withdrawals	0.2%

Table 2: Classification of BGP update messages

Gaining or losing reachability: Another 14.8% of messages either add or remove the only prefix that covers some range of IP addresses. Half are withdrawal messages that leave these addresses with *no* matching prefix, and the other half are announcements that allow these addresses to go back to having a matching prefix.

Changing the longest-matching prefix: Another 13.0% of messages cause some addresses to change to a different longest-matching prefix. Half are withdrawal messages that force these addresses to match a less-specific prefix, and the other half are announcements that allow these addresses to match a more-specific prefix.

Affecting a prefix that is not used for forwarding: The remaining 2.5% of update messages either add or remove a prefix that is not the longest-matching prefix for *any* IP addresses[1]. These prefixes are supernets like 12.0.0.0/8 that correspond to an address space that is completely covered by more-specific prefixes like 12.0.0.0/9 and 12.128.0.0/9.

Analysis of BGP update messages for a different time period (namely, March 2009) led to very similar results. In the rest of this section, we focus on the 13.0% of BGP update messages that cause prefix-match changes.

4.2 Characterization of Prefix Match Changes

To analyze the prefix-match changes, we also account for the effects of route flapping, and filtered 25,120 BGP updates (0.21% of the total BGP updates in February 2009) caused by route flapping, and left us with 1,278,552 prefix-match changes for the month of February 2009 for further analysis.

Looking at the remaining measurement data, we notice that most addresses ranges have a single prefix that serves as the longest-matching prefix the vast majority of the time. In fact, 95.2% of the address ranges have a prefix they use more than 90% of the time, and 98.7% have a prefix they use more than 60% of the time. We apply a threshold of 60% to identify the *dominant* prefix for each address range, and analyze the prefix-match changes that cause an address range to *stop* using its dominant prefix. This leaves us with 688,914 prefix-match changes to analyze (which is 53.9% of the total prefix-match changes). For some address ranges, these events involve the brief announcement (and subsequent withdrawal) of a more-specific prefix; for others, these events involve the brief withdrawal of the dominant prefix and the temporary use of a less-specific route. As such, we classify prefix-match changes in terms of whether the dominant route is more-specific or less-specific than the other (briefly used) prefix. To understand the possible reasons for the prefix-match changes, we also compare the origin ASes of the old and new prefixes. This leaves us with four cases, as summarized in Table 3. Note that the more-specifc and less-specific prefix match mentioned in the table are for the briefly used prefixes.

Same origin AS, more-specific prefix: About 13.6% of the prefix-match changes involve a brief announcement of a more-specific prefix with the same origin AS as the dominant prefix. We suspect that these prefix-match changes are caused by temporary route leaks, where the more-specific prefix is announced inadvertently due to a configuration mistake that is fixed relatively quickly (e.g., within a few hours or at most a day or two).

Same origin AS, less-specific prefix: About 58.4% of the prefix-match changes involve a brief withdrawal of the dominant prefix that leads to the temporary use of a less-specific route with the same origin AS. We suspect that these prefix-match changes are caused by multi-homed ASes that announce both prefixes for a fine-grained load balancing. For example, a multi-homed stub AS connected to two providers may announce 15.0.0.0/17 to one provider and 15.0.128.0/17 to the other, and the supernet 15.0.0.0/16 to both. The more-specific 15.0.0.0/17 prefix would be withdrawn whenever the link to the first provider fails, and the less-specific 15.0.0.0/16 would remain because the route is also announced via the second provider.

Different origin ASes, more-specific prefix: Only 2.9% of the prefix-match changes involve a brief announcement of a more-specific prefix from a different origin AS. We suspect some of these announcements correspond to "subprefix hijacking" caused by a configuration mistake or a malicious attack. For example, during the infamous hijacking of YouTube in February 2008 [3], Pakistan Telecom mistakenly announced 208.65.153.0/24, a subnet of YouTube's 208.65.152.0/22 address block. Another cause could be an ISP that inadvertently misconfigures a route filter that is supposed to block small address blocks announced by one of its customer ASes.

Different origin ASes, less-specific prefix: About 25.1% of the prefix-match changes involve a brief withdrawal of the dominant prefix that leads to the temporary use of a less-specific route with a different origin AS. We suspect that these prefix-match changes occur when a customer AS fails, but its provider does not. For example, suppose a provider that announces 12.0.0.0/8 has allocated 12.1.1.0/24 to one of its customers. If the customer fails, the customer's route for 12.1.1.0/24 is withdrawn, while the provider's 12.0.0.0/8 route remains.

In our ongoing work, we are analyzing these four cases in greater detail, to understand the causes of the prefix-match changes and the resulting impact on end-to-end reachability.

4.3 Joint Analysis with Traffic Data

In this subsection, we present a joint analysis with the traffic data from the same router to understand the effects of prefix-match changes on end-to-end reachability.

In practice, active and passive measurements are two approaches to infer data-plane reachability. With active mea-

[1]In this category, we see more announcements than withdrawals—a seemingly odd phenomenon we intend to investigate further. We suspect that, over time, some ASes introduce additional supernet routes as part of configuring backup routes.

Origin ASes	Prefix Match	#Events	Possible Explanations
Same	More-specific	94,121 (13.6%)	Route leak
Same	Less-specific	402,006 (58.4%)	Load balancing, failover to a backup route
Different	Less-specific	172,596 (25.1%)	Customer failure
Different	More-specific	20,191 (2.9%)	Sub-prefix hijacking, announcement of a new customer route

Table 3: Four classes of prefix-match events and their possible causes

surement, tools like ping and traceroute generate ICMP packets for the destination host or routers along the path to respond. However, active measurement tools can not accurately infer IP reachability, because: (i) ICMP packets may be filtered by middle boxes such as NAT and firewalls and (ii) many routers do not generate ICMP responses, or rate-limit the responses. Finally, active measurement often imposes heavy measurement overhead by sending many ICMP packets to monitor blocks of IP addresses over short time intervals. Instead, we use the passive measurement of IP flows, which are sampled and collected as Netflow records, for our analysis of end-to-end reachability.

In this analysis, our aim is to find counterexamples to the conventional understanding of reachability changes: (i) when a prefix is withdrawn, the IP addresses it covers become unreachable and (ii) if a prefix has a BGP route in the routing table, then the covered IP addresses are reachable. To counter the first conventional wisdom, we show that even if a prefix is withdrawn, the IP addresses could still be reachable via a less-specific prefix, corresponding to the second row in Table 3. For the second conventional wisdom, we illustrate that even when the routing table contains a route to a prefix, the IP addresses covered by this prefix might not be reachable, especially if the route is a less-specific prefix of the network provider.

To perform our analysis, we consider the prefix-match changes given in Table 3. For each change, we compute traffic volume from all the Netflow records of the affected address ranges in five minute bins around the time of the prefix-match change. This allows us to understand the impact of the prefix-match change on reachability. Our expectation is that if an address range becomes unreachable, the traffic volume would drop to a very low level. While a comprehensive analysis is part of ongoing work, we here present two examples that allow us to counter the two conventional wisdoms mentioned above.

The Netflow records are collected at the incoming interfaces at most of the core routers in the tier-1 ISP. In order to make sure that the traffic changes are caused by the routing changes, we selectively use the Netflow records for traffic that leaves the ISP at the same PoP where we collected the BGP routing updates. We used Netflow records from February 18-27, 2009 for the joint analysis. There are two stages of sampling during the collection of the Netflow records: packet sampling at the rate of 1/500 and smart sampling at the threshold of 80,000,000 bytes [4]. Because of both stages of sampling, correction has to be done to estimate the actual number of bytes or packets.

Figure 3 shows the traffic volume for an address range that changes to a less-specific prefix with the same origin AS. Specifically, the address range changed from a /20 to /17 prefix for about half an hour on February 18, 2009. As the traffic volume curve shows, the destinations in the address range continued receiving about the same amount of traffic. Note that the traffic volume drops to zero at some points, meaning that no packets were captured by Netflow during that 5-minute period of time. Because of the correction for sampling, the corrected bytes at each 5-minute interval tend to be multiples of the smart sampling threshold. But this example still shows that even if the most-specific prefix is withdrawn, the less-specific prefix may still be used to deliver the traffic.

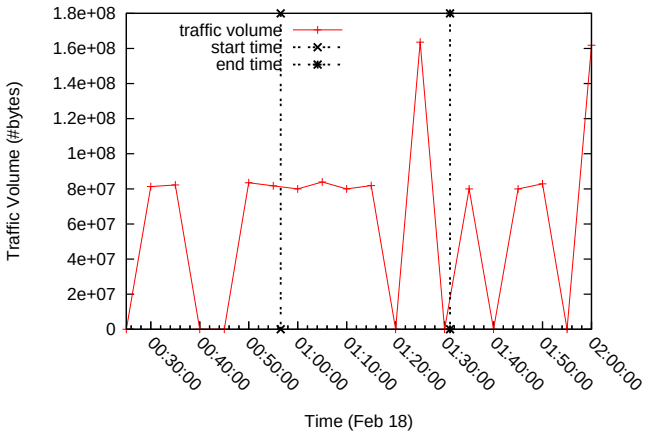

Figure 3: Although a prefix was withdrawn, the IP addresses it covered were still reachable (Feb 18, 2009).

Figure 4 shows the traffic volume for another address range, which changes to a less-specific prefix from a *different* origin AS, corresponding to the third row in Table 3. In this case, the address range changed from a /20 to /9 prefix for about 15 minutes. As illustrated in the figure, although a less-specific prefix is available in the routing table, the traffic volume dropped to zero. If we take the sudden drop of traffic as evidence of the address range becoming unreachable, this example illustrates the point that the existence of a prefix in the routing table does not necessarily imply that the prefix is reachable. Of course, it is possible that the destinations in the address range were still reachable in this example, but because of low volume, the traffic was not sampled by Netflow. As part of our ongoing work, we are working on techniques for detecting big changes in traffic volume using Netflow records, and analyzing the fields in these records, such as the number of packets and TCP flags, to better infer when a collection of IP addresses have become unreachable.

5. RELATED WORK

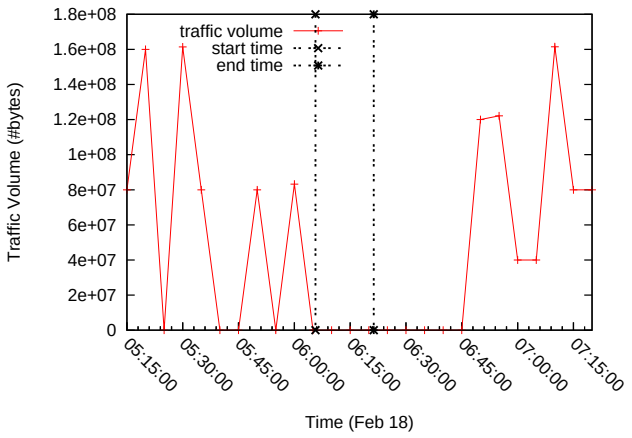

Figure 4: After the BGP withdrawal, the IP addresses matched a less-specific prefix; still, no traffic to these destination IP addresses was observed until after the more-specific prefix was announced again (Feb 18, 2009).

Our paper relates to earlier studies that used BGP measurement data to analyze the relationship between IP prefixes [5, 6, 7, 8]. For example, the work on BGP policy atoms [5, 6] showed that groups of related prefixes often have matching AS paths, even when viewed from multiple vantage points; typically, a more-specific prefix had different AS paths than its corresponding less-specific prefix [5]. Other researchers analyzed BGP table dumps to understand the reasons why each prefix appears in the interdomain routing system, and the reasons include delegation of address space to customers, multihoming, and load balancing [7, 8]. Our results in Table 3 present a similar classification scheme, though focused on the *changes* in the longest-matching prefix rather than a static analysis of a BGP table dump.

Our work also relates to earlier analysis of BGP routing dynamics [9, 10, 11, 2]. These studies analyzed announcement and withdrawal message for each destination prefix, and group related BGP update messages to identify BGP convergence events and route flapping. Whereas these studies treated each IP prefix independently, our analysis of BGP update dynamics focuses on the *relationship between nested prefixes*. Still, we draw on the results in these earlier studies when selecting thresholds for identifying phenomena such as BGP path exploration and route flapping. Our paper also relates to measurement studies of prefix hijacking and particularly *subprefix* hijacking [12, 13] that triggers a change in the longest-matching prefix. However, our study considers a wider range of causes of prefix-match changes.

Previous studies have also characterized IP reachability through direct or indirect observations of the underlying data-plane paths used to forward packets [14, 15, 16, 17, 18, 19, 20]. Most of these studies involve active probing (using ping, traceroute, or custom tools) [14, 15, 16, 17], sometimes triggered by passive observations of reachability problems [14, 15]. Other work has focused on analysis of passively collected traffic measurements (such as Netflow data or Web server logs) to detect possible routing changes or reachability problems [18, 19, 20]. In contrast, our paper has focused primarily on how the longest-matching prefix, used in packet forwarding, changes over time. That said, these previous studies are quite relevant to our ongoing analysis of the Netflow data to understand the impact of these prefix-match changes on end-to-end reachability.

6. CONCLUSION

In this paper, we analyze BGP routing changes that affect the longest-matching prefix used for packet forwarding. We find that prefix-match changes are relatively common, accounting for more than 13% of BGP update messages. Ignoring these prefix-match changes can lead to misleading conclusions for researchers and practitioners alike. A BGP withdrawal does not necessarily imply that IP addresses have become unreachable, if the route for another (less-specific) prefix can successfully deliver the traffic. A BGP withdrawal can also make a previously unreachable destination reachable again, if the withdrawal marks the end of a subprefix-hijacking event. Or, a withdrawal may have no impact at all on packet forwarding, if all the IP addresses match more-specific prefixes. These distinctions can only be made by understanding the nesting of prefixes and tracking changes in the longest-matching prefix over time. Our joint analysis with the Netflow data illustrates the cases where the relationship between BGP updates and IP reachability could be counterintuitive.

In our ongoing work, we want to connect our analysis of prefix-match changes with the effects on end-to-end reachability in the data plane. Given the practical limitations of active probing, we plan to investigate how much information we can infer from passive traffic measurements, whether Netflow data (as in our preliminary analysis) or fine-grained packet traces. Our long-term goal is to find ways to extract the maximum amount of useful information from passively-collected measurement data. We believe the analysis in this paper is an important first step in that direction.

7. ACKNOWLEDGMENTS

We thank Changhoon Kim and Haakon Ringberg for their valuable feedback in the early stages of this work, as well as Alexandre Gerber and Carsten Lund for answering questions regarding the Netflow data set. We are also grateful to Olivier Bonaventure, Alex Fabrikant, Elliott Karpilovsky, Kobus van der Merwe, and the anonymous reviewers for their comments and suggestions.

8. REFERENCES

[1] B. Zhang, V. Kambhampati, M. Lad, D. Massey, and L. Zhang, "Identifying BGP routing table transfers," in *Proc. ACM SIGCOMM Workshop on Mining Network Data (MineNet)*, August 2005.

[2] J. Wu, Z. M. Mao, J. Rexford, and J. Wang, "Finding a needle in a haystack: Pinpointing significant BGP routing changes in an IP network," in *Proc. Networked Systems Design and Implementation*, May 2005.

[3] Rensys Blog, "Pakistan hijacks YouTube." `http://www.renesys.com/blog/2008/02/pakistan_hijacks_youtube_1.shtml`.

[4] N. Duffield, C. Lund, and M. Thorup, "Learn more, sample less: Control of volume and variance in network measurement," *IEEE Transactions in*

Information Theory, vol. 51, no. 5, pp. 1756–1775, 2005.

[5] A. Broido and k. claffy, "Analysis of RouteViews BGP data: Policy atoms," in *Proc. Network Resource Data Management Workshop*, 2001.

[6] Y. Afek, O. Ben-Shalom, and A. Bremler-Barr, "On the structure and application of BGP policy atoms," in *Proc. Internet Measurement Workshop*, pp. 209–214, 2002.

[7] T. Bu, L. Gao, and D. Towsley, "On characterizing BGP routing table growth," *Computer Networks*, vol. 45, pp. 45–54, May 2004.

[8] X. Meng, Z. Xu, B. Zhang, G. Huston, S. Lu, and L. Zhang, "IPv4 address allocation and BGP routing table evolution," *ACM Computer Communication Review*, January 2005.

[9] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed Internet routing convergence," *IEEE/ACM Trans. on Networking*, vol. 9, pp. 293–306, June 2001.

[10] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP misconfiguration," in *Proc. ACM SIGCOMM*, August 2002.

[11] J. Rexford, J. Wang, Z. Xiao, and Y. Zhang, "BGP routing stability of popular destinations," in *Proc. Internet Measurement Workshop*, November 2002.

[12] J. Karlin, S. Forrest, and J. Rexford, "Autonomous security for autonomous systems," *Computer Networks*, October 2008.

[13] M. Lad, D. Massey, D. Pei, Y. Wu, B. Zhang, and L. Zhang, "PHAS: A prefix hijack alert system," in *Proc. USENIX Security Symposium*, 2006.

[14] M. Zhang, C. Zhang, V. Pai, L. Peterson, and R. Wang, "PlanetSeer: Internet path failure monitoring and characterization in wide-area services," in *Proc. Operating System Design and Implementation*, 2004.

[15] N. Feamster, D. Andersen, H. Balakrishnan, and M. F. Kaashoek, "Measuring the effects of Internet path faults on reactive routing," in *Proc. ACM SIGMETRICS*, 2003.

[16] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson, "User-level path diagnosis," in *Proc. Symposium on Operating System Principles*, October 2003.

[17] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an accurate AS-level traceroute tool," in *Proc. ACM SIGCOMM*, August 2003.

[18] F. Wang, L. Gao, O. Spatscheck, and J. Wang, "STRID: Scalable trigger-based route incidence diagnosis," in *Proc. IEEE International Conference on Computer Communications and Networks*, August 2008.

[19] P. Huang, A. Feldmann, and W. Willinger, "A non-intrusive, wavelet-based approach to detecting network performance problems," in *Proc. Internet Measurement Workshop*, November 2001.

[20] V. N. Padmanabhan, L. Qiu, and H. Wang, "Server-based inference of Internet performance," in *Proc. IEEE INFOCOM*, March 2003.