

Understanding BGP Behavior through a Study of DoD Prefixes *

Xiaoliang Zhao, Dan Massey[†]

S. Felix Wu[‡]

Mohit Lad, Dan Pei, Lan Wang, Lixia Zhang[§]

USC/ISI

UC Davis

UCLA

Abstract

BGP is the de-facto inter-domain routing protocol and it is essential to understand how well BGP performs in the Internet. As a step toward this understanding, this paper studies the routing performance of a sample set of prefixes owned by the U.S. Department of Defense (DoD). We examine how reliably the sample set is connected to the Internet and how it affects the rest of the Internet. We show that our sample set receives reliable connectivity, with the exception of a few prefixes. We also show that, on average, the sample set has minimal impact on global routing, but certain BGP features used by DoD routers result in periods of excessive routing overhead. During some stressful periods, our sample set, only 0.2% of all prefixes, contributed over 80% of a particular BGP update class. We explain how the BGP design allows certain local changes to propagate globally and amplifies the impact of our sample prefixes.

1. Introduction

BGP [6] is the de-facto inter-domain routing protocol used to provide essential reachability information in the Internet. The Internet consists of thousands of Autonomous Systems (ASes) and BGP is used to exchange reachability information between these ASes. Defects in the BGP protocol design and faults or attacks to the BGP routing infrastructure can easily lead to adverse consequences such as host unreachability, misdirected traffic, or denial of services. To enable the Internet to achieve ultimate resilience to component faults and malicious attacks, we must gain a comprehensive understanding of BGP's operation to quantify its response to faults, and its vulnerabilities to attacks.

*This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No DABT63-00-C-1027. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the DARPA.

[†]{xzhao, masseyd}@isi.edu

[‡]wu@cs.ucdavis.edu

[§]{mohit, peidan, lanw, lixia}@cs.ucla.edu

In practice, however, the sheer size of today's global Internet makes it difficult to gain an overall understanding of the global routing at once. One way to tackle the challenge is to sample the routing performance for various destinations and gain insights on the global system by conducting detailed analysis of these samples. In this paper, we examine a small sample set of BGP prefixes owned by the U.S. Department of Defense (DoD); one motivation for this selection is the relevance and critical importance of these prefixes to government activities. We used two basic measurements to evaluate the routing performance of the set. First, we consider how well these prefixes are connected to the Internet by measuring how persistently BGP provides a route leading to these prefixes. Whenever BGP fails to provide a route to these prefixes, the hosts associated with these prefixes will be unreachable from the rest of the Internet. Second, we consider how these prefixes affect the global BGP infrastructure. Routers in thousands of ASes maintain reachability information for these Defense Department prefixes through receiving update messages regarding the connectivity changes to these prefixes. We measure the global impact in terms of the number of routing updates that are associated with our prefixes.

The paper is organized as follows. Section 2 discusses related work. Section 3 describes our methodology for gathering BGP data and discusses how we selected our set of sample DoD prefixes. Section 4 considers the reachability to these prefixes and shows that connectivity between the Internet and our set of sample DoD prefixes is reliable, with the exception of one or two prefixes which exhibit poor reachability.

Section 5 considers the number of BGP updates associated with our set of sample DoD prefixes and shows that typically a smaller than average number of updates are associated with these prefixes (as compared to the whole Internet). However, there were times when our set of sample DoD prefixes contributes an excessive number of updates. Section 6 examines the excessive updates in more detail and provides an analysis of the abnormal DoD prefix behavior. The results illustrate how the local changes in a particular BGP route attribute (e.g., the AGGREGATOR at-

tribute), can trigger wide scale changes. This is an example of BGP design decision where relatively local information is propagated globally and our set of sample DoD prefixes is especially effected by this behavior. Section 7 concludes the paper.

1.1. BGP operations and terminology

To exchange routing information, two BGP routers first establish a BGP peering session which operates on top of a TCP connection. One BGP router may have multiple peers. When a new BGP session starts the two peering routers first exchange their full routing tables through a series of BGP messages. After the initial route exchanges, each BGP router sends only incremental updates to its peers for new or modified routes. Each update lists a set of reachable prefixes attached with some attributes to describe the topological changes or policy changes. A detailed description of such attributes is listed in [6]. When a BGP router discovers that it can no longer reach a destination (i.e., an IP address prefix) that it has announced to its peers previously, it sends a message to its peers withdrawing the route.

Before we present our methodology and findings, we would like to clarify two terms that are used throughout this paper, *BGP message* versus *BGP prefix update*. A *BGP message* refers to the message used by BGP peers to announce a route, withdraw a route, or manage the BGP session. In the first two cases, the message can carry one BGP route and *multiple* IP address prefixes that use the same route. To analyze the route changes for *individual* prefixes, we studied the sequence obtained by unpacking the BGP messages. These unpacked announcements or withdraws are referred to as “BGP prefix update” (or “BGP update” for brevity).

2. Related work

While BGP has been widely used in the Internet, its behavior in this real-world environment is yet to be fully understood. Labovitz, et. al. [3] studied BGP routing messages exchanged between US service providers and reported that the majority of BGP messages consisted of redundant pathological announcements. [4] further identified the origins of certain pathological behavior. They also showed that routing instability had been significantly reduced in the core network by software improvements.

Govindan and Reddy [2] studied the Internet topology and routing stability several years ago. They found that routes to prefixes were highly available and stable at that time, but the mean reachability duration for a prefix decreases with the Internet growth. The Internet has grown rapidly since this study and more recent data is needed to help better understand current Internet performance.

Paxson [5] studied the routing behavior from an end-to-end communication point of view. The results showed that Internet paths are heavily dominated by a single prevalent route. These measurements were conducted based on traceroute data. In contrast, our study uses a different data collection methodology that focuses on BGP routing updates.

Rexford, et. al. [7] studied the routing stability of popular destinations. They found popular destinations have remarkably stable BGP routes, while a small number of unpopular destinations are responsible for the majority of BGP instability. Instead of studying the popular destinations, we focus on the critical defense networks. Both studies have similar observations. Moreover, we also studied the global impact of DoD prefixes and analyzed some abnormal routing traffic related to worm attacks.

Cowie, et. al. [1] analyzes the BGP traffic during worm attacks and noticed that there were some “BGP storms”, i.e., excessive numbers of BGP updates over short periods of time. However, after looking into the BGP traffic and classifying it into different categories, [10] found that 40% of BGP storm was caused by a measurement artifact: BGP session resets at the monitoring point.

The work reported in this paper represents another step toward a comprehensive understanding of BGP performance. We measured the BGP reachability to a sample set of prefixes and explained the causes of “BGP storms” generated by this sample set during stressful network conditions.

3. Data methodology

We analyzed BGP routing updates collected by RIPE NCC[8] during several months in 2001 and 2002. RIPE NCC has eight data collection points. We selected one of these, monitoring point RRC00, and gathered data from the BGP routers listed in Table 1. Some of these routers are located in global ISPs, while others are located in regional ISPs. Geographically, these routers are located in different countries including the United States, Japan and three European countries.

The RRC00 monitoring point provides a diverse view of ASes in the U.S., Asia, and Europe, but no single vantage point provides a view of the “Internet”. Rather, each AS has its own view of the Internet and that view is dependent on the AS location, its peers, its policies, the policies of its peers, and so forth. As a result, each AS experiences different BGP routing dynamics. In this paper, the dynamics of nine routers from the ASes listed in the Table 1 are captured at the RRC00 monitoring point and this study uses these nine diverse views. However, it should be noted that results for other ASes might result in somewhat different views.

We chose our particular collection point (the RRC00 collection point) because of its diverse routers and because it receives full routing tables from ISPs. If an ISP only provides partial routing tables and later withdraws its route to a prefix, this action may indicate that an ISP has lost its route to this prefix or may indicate the ISP has simply changed routes and the new route does not match the partial export policy.

Location	ASes that RRC00's peers belong to
US	AS7018 (AT&T), AS2914 (Verio)
Netherlands	AS3333 (RIPE NCC), AS1103 (SURFnet) AS3257 (Tiscali Global)
Switzerland	AS513 (CERN), AS9177 (Nextra)
Britain	AS3549 (Global Crossing)
Japan	AS4777 (NSPIXP2)

Table 1. RRC00's peering ASes Examined in This Study

It should also be noted that BGP updates are sent to the monitoring point via multi-hop BGP connections. This allows the RRC00 monitoring point to capture the views from diverse locations, but differs from peering sessions used in operational networks. In the operational Internet, nearly all ISP peerings are through BGP routers sharing a common physical link and BGP updates are sent via TCP connection over single link/hop. In contrast, the RRC00 monitoring point peers with ISP routers via TCP connections that cross multiple hops. When the multi-hop session fails, the monitoring point reports a session state change. In nearly all cases, the same routes are re-advertised when the session to the ISP router resumes. We attribute this behavior to lower stability of the multi-hop BGP sessions. We pre-process the update files to remove the updates that are generated due to session resets. Our work in [10] discusses problems associated with multi-hop sessions and techniques for cleaning the data in more detail. Our pre-processing of BGP updates results in a clean set of BGP updates to analyze.

The data used for our analysis was collected from July 2001, September 2001, November 2001, February 2002, July 2002, and August 2002. All of the data was examined using the methods described in the following sections. Due to paper size limitations this paper only presents the results for particular months and from particular peers' point of view. Unless stated otherwise, the results for other months and other peers are generally similar to the results presented here.

3.1. Selecting DoD prefixes

Prefixes belonging to the U.S. Department of Defense (DoD) are originated from several ASes. Each AS may connect to the public Internet via different ISPs and from different topological locations. Some aspects of BGP behavior may vary depending on the origin AS. For example, an origin AS that has only one upstream ISP may be more likely to experience failure than origin AS that is multi-homed to many ISPs. We identified one AS, AS 568, which is operated by Defense Information Systems Agency (DISA) as the dominant origin AS for DoD prefixes. For presentation clarity and length consideration, our analysis focuses on this specific AS only.

To obtain the prefixes originated by AS 568, we took a snapshot of BGP routing table from Oregon RouteView server on August 23, 2002. From this routing table, we obtained 281 different prefixes originated by AS 568. In the rest of this paper, those prefixes are called the set of sample DoD prefixes or simply the DoD prefixes.

Our data shows that AS 568 covers more than 68% of the IP address space assigned to DoD. This value is obtained in the following way. We first identified the DoD owned ASes by searching for the keywords, “.mil”, “.army”, etc in the Routing Assets Database (RADb) and Internet Routing Registry (IRR), and identified a set of the DoD prefixes. We incorporated this prefix/AS information in the routing table snapshot on August 23, 2002, and finally obtained a set of 98 DoD ASes, which originate 2,573 prefixes. We then considered how much of that IP space is covered by AS568. The 98 DoD ASes (including AS568) covers about 394,949 /24 IP blocks, but AS568 itself covers about 283,949, more than 70% of the total identified DoD prefixes. The rest of the 98 ASes originates 2,292 prefixes, and 1,409 of them are more specific than the AS568 prefixes that cover the same address space. These 1,409 prefixes punched “holes” in 146 of 281 AS568 prefixes. These “holes” cover about 15,333 /24 IP blocks, less than 6% of total IP space covered by AS568. Overall, of the 394,949 /24 identified DoD IP blocks, 268,616 (68%) of them are solely covered by AS568. Therefore, we believe that the selection of AS568 is a reasonable choice for our study of DoD prefixes.

4. Reachability of DoD prefix set

For the users of a network, the fundamental concern is the network reachability, *i.e.*, whether the network users can reach the rest of the Internet and whether this network can be reached by the rest of the Internet. In this study, we define reachability in terms of BGP routing. We say a prefix is *reachable* if there is a BGP route for prefix. Similarly, we say a prefix is *unreachable* if there is no BGP route for the prefix (*i.e.* the route either never existed or existed, but

was withdrawn). In this section, we will measure how the number of unreachable DoD prefixes changes over time and examine the implications of this behavior.

By observing the updates sent by a particular peer, we can measure reachability at that peer. Since the peers in our study provide full route tables, we can determine BGP reachability for a prefix by observing the update messages sent by the peer for that prefix. In other words, if the peer advertises a BGP route to the prefix, then we say the prefix is reachable via that peer. If the peer later sends a withdraw for that prefix, then we say the prefix is unreachable through that peer.

For our set of 281 sample DoD prefixes, Figure 1 shows the number of prefixes unreachable through peer P during August 2002. The X-axis is time during August 2002 and the Y-axis denotes the number of prefixes that could not be reached by this peer at time x . Some spikes in the graph are very narrow or single line spikes, suggesting that the prefixes were only unreachable for a very brief amount of time. There are points where over 25 prefixes are unreachable from this peer. However, our objective is to understand the behavior from the DoD prefix point view. Some spikes in the above graph may be local to the peer (or a region near the peer). To better understand the DoD prefix behavior, we need to consider how the prefixes are viewed from multiple vantage points.

4.1. Globally unreachable prefixes

Prefix reachability often depends on the peer's local viewpoint. For example, suppose AS4777 (NSPIX2 in Japan) withdraws the route to a prefix because some intermediate AS on the AS path fails and no alternate path exists. The prefix will be unreachable according to the AS4777 peer. However, the same failure may have no impact on the BGP route used by AS2914 (Verio) and the AS2914 peer will continue to declare the prefix reachable. During our study, different peers did report differing reachability states for the same prefix. We are particularly interested in cases where all nine peers in our study declared a prefix to be unreachable. Since our peers are located at diverse spots throughout the Internet, we say a prefix is *globally unreachable* if all nine peers cannot reach the prefix. A globally unreachable prefix suggests a routing failure occurred at or near the origin AS (AS 568).

Figure 2 shows the globally unreachable DoD prefixes observed during September 2001 and August 2002. The X-axis is time and the Y-axis denotes the number of prefixes that could not be reached by any of the nine peers at time x . At the worst instant on 9/23/01, there were only seven prefixes that were declared unreachable by all nine of the RRC00 peers. Some spikes in the graph are very narrow or single line spikes, suggesting that the prefixes were only un-

reachable for a very brief amount of time. However between September 8 and September 23, there are always at least 3 globally unreachable prefixes. Further studies revealed that three prefixes were withdrawn by all peers on September 8th and were not announced again until Sept 26th.

In August 2002, there are a number of narrow spikes but no long periods where a prefix was globally unreachable. The narrow spikes can imply BGP advertisement flapping activity, where a prefix was withdrawn and then announced again immediately, followed by a withdrawal and so on. We identified one prefix that experienced this type of flapping. Out of the 281 prefixes in the our sample set of DoD prefixes, very few experienced substantial reachability problems at all nine peers.

Note how Figure 1 contrasts the global view from all nine peers with the individual view from a single peer. Compared with Figure 2(b), there are more spikes in Figure 1 and the spikes are much greater. This indicates that events near that peer did impact reachability, but these same events did not reduce reachability at all other peers. For example, the single peer was unable to reach over 25 prefixes on 8/6/02 but all but one of these prefixes were reachable for at least one of the other 8 peers.

The difference between individual peer views and the global view from all nine peers simply reflects the fact that different peers have distinct AS paths that do not rely on the same failed component. One way to increase the number of disjoint paths is for the origin AS to connect with different providers. Our data show that AS 568 currently peers with four providers and thus provides a diverse set of potential paths. This could partially explain the large difference between the spikes in these two figures.

4.2. Unreachability duration

The Duration of unreachability is the time between a prefix is withdrawn and the prefix is announced again. Generally speaking, short duration unreachability may not necessarily be alarming and can possibly be due to transient problems (e.g., caused by a link failure that is quickly routed around). However, a long period of unreachability is worthy of strong concern.

Figure 3 shows the cumulative probability distribution of duration for globally unreachable prefixes in August 2002. 49 distinct prefixes were globally unreachable at least once during August 2002. Ideally, one wishes the duration of global unreachability to be very short, in reality only 16% of unreachable durations lasted less than two minutes. 40% of unreachable durations were shorter than ten minutes, a period that is long enough to be noticeable by applications and end users. Furthermore, 17% of unreachability durations were longer than one hour. The longest unreachability duration was thirty-eight hours, representing a serious net-

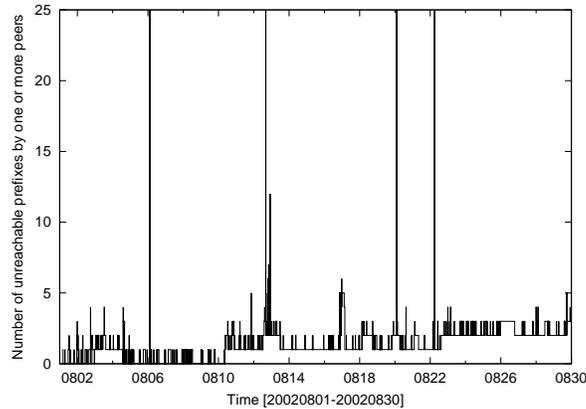


Figure 1. Unreachable Prefixes At Peer P - August 2002

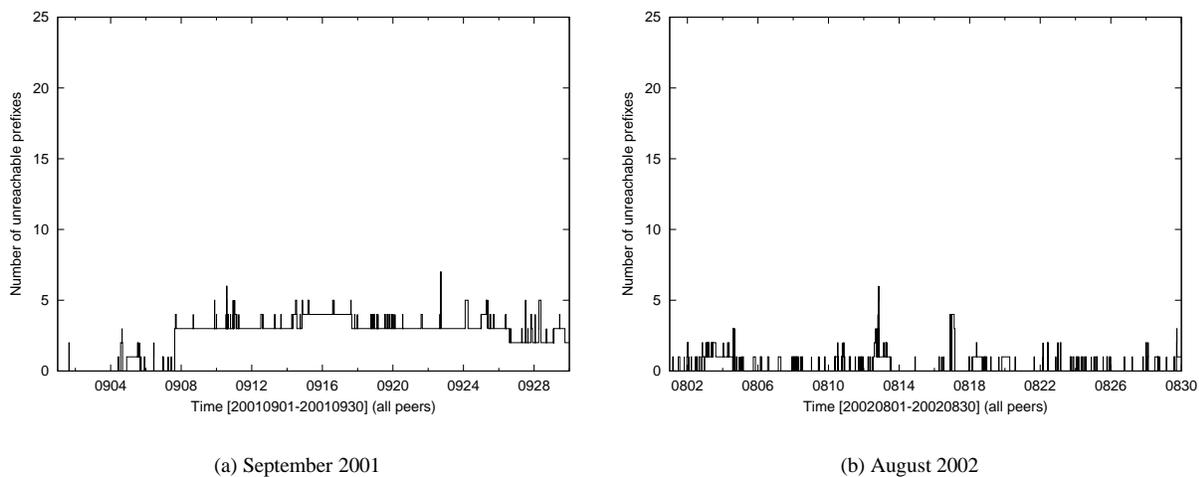


Figure 2. Globally Unreachable DoD prefixes

work connectivity outage.

5. Global impact of the DoD prefix set

In this section, we compare our set of sample DOD prefixes with Internet prefixes as a whole and examine prefix behavior in terms of both BGP update volume and BGP update types. In BGP, the route to a prefix should be announced once and then reannounced only if there is a change in some attribute associated with the route. Thus ideally, a prefix would have a stable route that is announced once and no additional updates would be sent for the prefix. In practice, [7] shows that routes to some popular prefixes tend to be quite stable and these prefixes contribute only a few updates to the volume of BGP updates seen in the

global infrastructure. But other less popular prefixes can be less stable and these unstable prefixes contribute a disproportionately large number of updates to the global BGP infrastructure.

The number of updates for the set of DoD prefixes, on an average, was no more than the number of updates sent by Internet prefixes on most days of normal activity. In some cases the number of updates contributed by our set of sample DOD prefixes was even fewer than average. However, during network stress events such as the Nimda worm attack, our set of sample DoD prefixes behaved much worse than the Internet as a whole. In section 5.1, we start by looking at the number of updates generated for our set of sample DoD prefixes and compare this with the average number of updates generated for all the Internet prefixes. In section

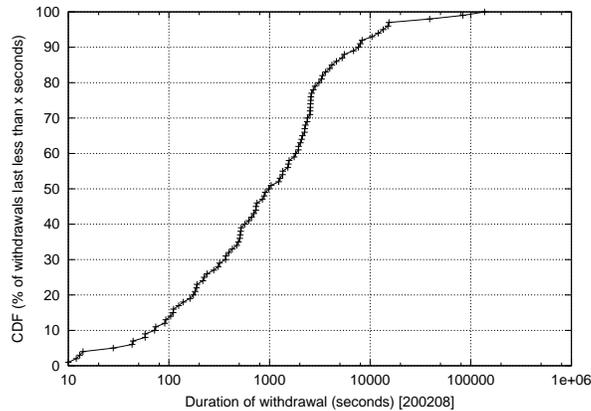


Figure 3. Unreachability Duration of August 2002

5.2, we follow up with a study of the distribution of the number of updates per prefix. Finally, in section 5.3, we conclude this comparison between DoD and the Internet as a whole, with a classification of the updates into different types.

5.1. Update counts

Our set of sample DoD prefixes consists of 281 prefixes, while a typical backbone router contains BGP routes to over 100,000 prefixes. Thus the total number of updates involving our small set of sample DoD prefixes should be only a small fraction of the total number of BGP updates seen in the Internet. Our objective is to determine whether this set of 281 prefixes is sending its proportionately “fair share” of updates.

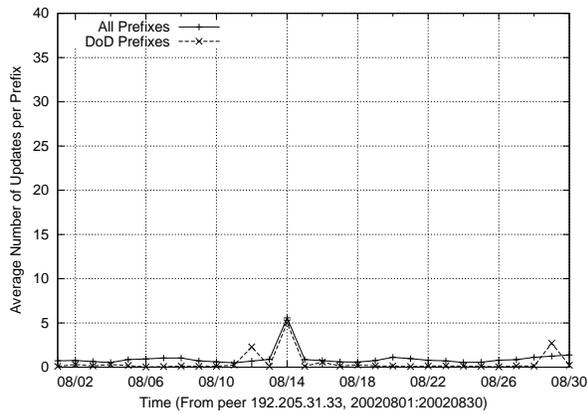
Figure 4 shows the number of updates per-prefix sent during August 2002 and September 2001 as seen from ISP A’s point of view. August 2002 is a typical month and similar views are seen in other months and from other peers. September 2001 was selected since the Nimda worm attack occurred during this month, and this attack is known to have had an adverse impact on the Internet. Figure 4 shows that our set of sample DoD prefixes consistently generated fewer updates than other Internet prefixes. However, there are a few noticeable spikes where the number of updates per DoD prefix is substantially higher than that of other Internet prefixes. For example, on September 18, 2002 (the day of the Nimda attack), there were nearly forty updates per DoD prefix, but less than five updates per Internet prefix. In the following sections, we will provide an explanation for the spikes in DoD prefix updates.

5.2. Update count per prefix (CDF)

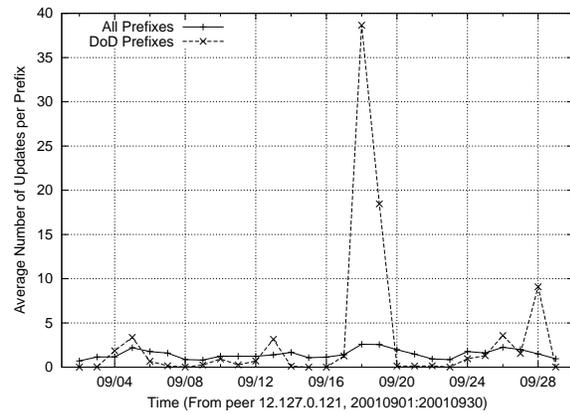
In this section, we plot the cumulative distribution of the number of updates sent for a prefix during a month (Figure 5, 6 and 7). The X axis is the number of updates accumulated over a month (log scale) and the Y axis shows the percentage of prefixes sending less than or equal to the corresponding number of updates. In each graph, we show two distribution curves, one belonging to our set of sample DoD prefixes and the other belonging to all the Internet prefixes.

Figure 5 shows that, from August 2, 2002 to August 30, 2002, about eight updates were generated for the best 10% of our DoD prefixes, roughly the same as that seen for the Internet prefixes. At the other extreme, at least twenty-five updates were generated for the worst 10% of DoD prefixes, while for the whole Internet, at least seventy-five updates were generated for the worst 10% of the prefixes. Less than twenty updates were sent for 80% of DoD prefixes, while less than twenty updates were sent for 60% of all Internet prefixes from August 2, 2002 to August 30, 2002. Overall our set of sample DoD prefixes were not among either extreme end of Internet prefixes and seemed to have an average performance for August 2002 (a typical month in our study).

Figure 6 shows the distribution for the month of September 2001 (from Sept 2, 2001 to Sept 30, 2001). Comparing this graph with figure 5, we see that the updates generated by our sample set of DoD prefixes was much more than the number generated by Internet prefixes. This behavior significantly differs from that of August 2002. The cumulative distribution curve for our DoD set is worse than the Internet prefix curve and implies that the number of updates generated for even the best DoD prefixes in the month of September 2001 was much more than the Internet in general. Comparing this curve with 5, we see that the September 2001 DoD curve is much more worse than that for August 2002.



(a) August 2002



(b) September 2001

Figure 4. Avg. Number of Updates Per Prefix - Viewed from ISP A

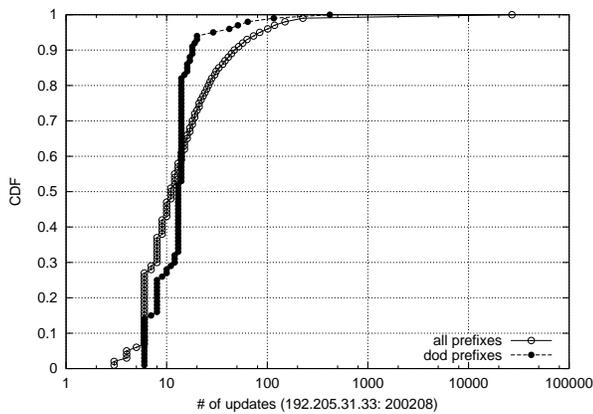


Figure 5. CDF viewed from ISP A for August 2002

One reason behind this aberration is the Nimda worm attack that took place on September 18, 2001. The DoD prefixes appear to have been impacted more by the Nimda attack than the Internet as a whole. An similar event was the Code Red attack that occurred in July of 2001, and again the update curve for our DoD set is worse compared to the Internet as a whole, as shown in Figure 7.

5.3. Update classification

The previous sections show that the DoD prefixes behave well most of the time, but there are some spikes, where the updates generated for our DoD set is far more than its proportional share of updates implying that the DoD prefixes

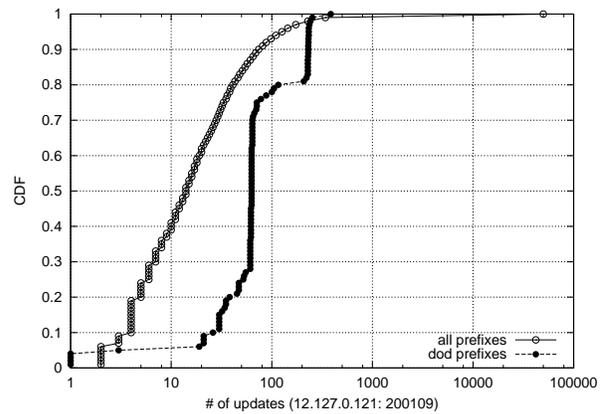


Figure 6. CDF viewed from ISP A for September 2001

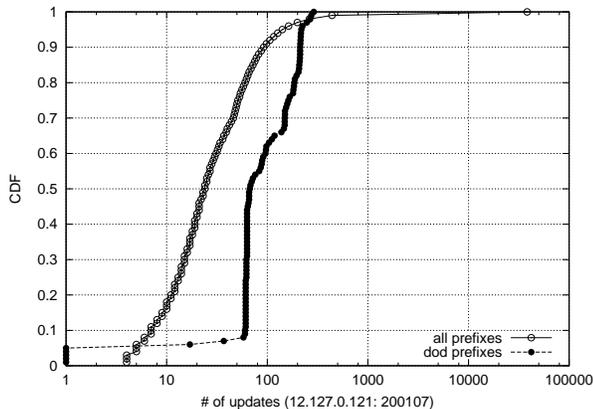


Figure 7. CDF viewed from ISP A for July 2001

appear to be more sensitive to events such as Code Red and Nimda. To better understand the behavior of our sample set of DoD prefixes, we examined the type of updates being sent. Our work in [10] defines the update class hierarchy (shown in Figure 8) that is based on the timing of an update and its relationship to previous updates. We examine the type of updates sent for our set of sample DoD prefixes and compare this with the type of updates sent for the Internet as a whole.

The update classes are defined as follows:

- A BGP peer may send an update to announce a previously unreachable prefix becoming reachable, and such update is classified as *New Announcement*.
- If a BGP peer sends an update to a currently reachable prefix, but the update contains the exactly same information as it previously sent, such an update is called a *Duplicate*.
- If a BGP peer sends an update to a currently reachable prefix, and the update replaces any of the attributes other than AS PATH, such an update is an *SPATH Implicit Withdraw*, or *SPATH* for short.
- A *DPATH Implicit Withdraw* is an update which replaces the the AS PATH attribute.
- A BGP peer may send a *Withdrawal* to withdraw a previously reachable prefix.

Figure 9 and Figure 10 compare our set of sample DoD prefixes with Internet prefixes in August 2002 and September 2001 respectively. Note that a large percentage of DoD prefix updates are SPATH updates, but this type of updates is normally only a small percentage of total updates for normal Internet prefixes. In addition, most of the spikes in DoD

updates involve an increase in SPATH updates. To understand the DoD behavior, the next section examines SPATH updates in detail.

6. An analysis of abnormal DoD prefixes behavior

In the previous section, we showed that our set of sample DoD prefixes generally behaved well when compared with the Internet as a whole. More specifically, on most days the average number of updates sent for DoD prefixes was lower than the average number of updates sent for general Internet prefixes. However, there are notable spikes when the DoD prefixes performed worse than the Internet as a whole. Figure 4 shows the average number of updates sent during August 2002 and Sept 2001 and in this section we examine the spikes where the average number of DoD updates exceeds that of general Internet prefixes.

During August 2002 there were three noticeable spikes in the number of DoD prefix updates on 8/14/02, 8/12/02, and 8/29/02. Note the spike on 8/14/02 is a bit dramatic, but it is not specific to the DoD prefixes. By comparing Figure 4(b) with Figure 9, we observe that the 8/14/02 spike was from an increase in duplicate updates sent by the peer being monitored. The number of duplicate updates for DoD prefixes increased proportionately and the average number of DoD updates still remained below the average number of Internet updates. While this is interesting from the monitored peer's perspective, we are primarily interested in the spikes where the DoD prefixes behaved differently.

During the spikes on 8/12/02 and 8/29/02, the average number of updates generated by Internet prefixes remained relatively normal but the average number of DoD prefix updates increased dramatically. By again looking at Figure 4(b) and Figure 9, we see this increase in DoD updates consists of an increase in SPATH updates. In September 2001, there are spikes in the average DoD prefix updates on 9/13/01, 9/18/01, and 9/19/01 and again all of these spikes consist of SPATH updates. A fourth spike on 9/4/01 is from a combination of SPATH and duplicate updates (see Figure 10). Only one spike on 9/28/01 does not correspond to an increase in SPATH updates. The data from other months is similar and in most cases, an increase in the number of DoD prefix updates corresponds to an increase in SPATH updates. Therefore, we would like to understand the class of SPATH updates sent by the DoD prefixes.

6.1. The DoD SPATH updates

An SPATH update indicates that the BGP route to the prefix still uses the same AS path (SPATH), but some other route attribute has changed. In other words, an SPATH update does not convey new information about the topologi-

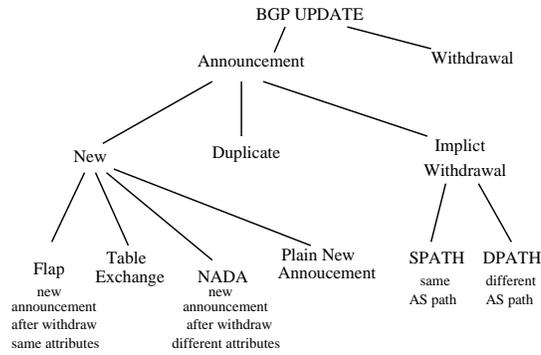
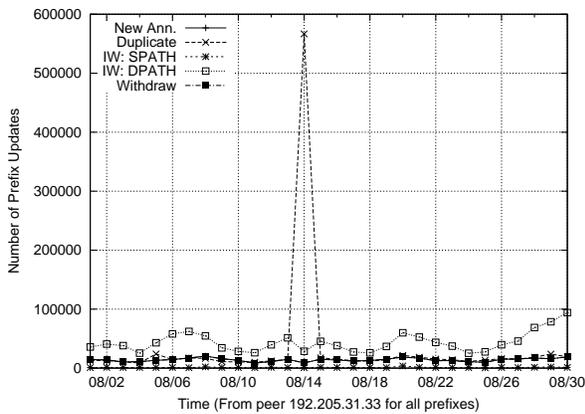
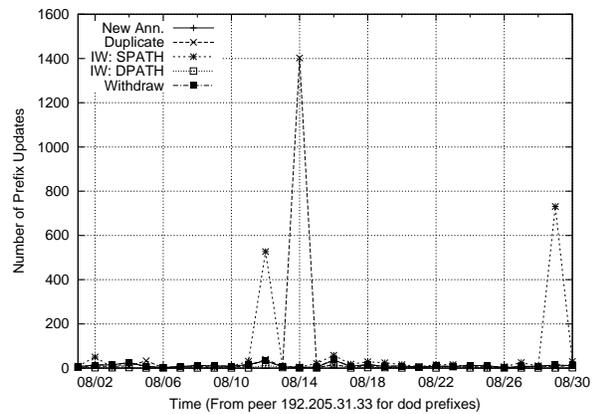


Figure 8. BGP Update Class Hierarchy



(a) ISP A's view of Internet prefixes



(b) ISP A's view of DoD prefixes

Figure 9. ISP A's view of update classes in August 2002

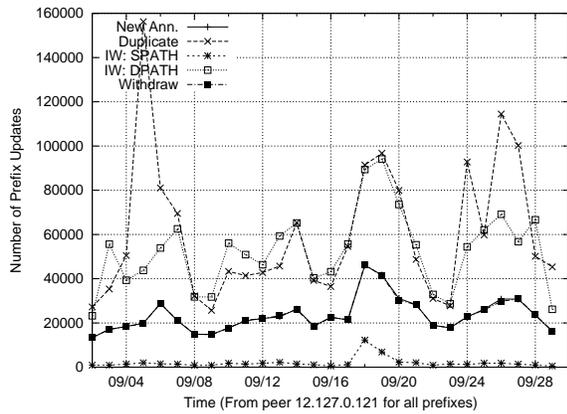
cal path of ASes used to reach the prefix, but it does contain new information about some other route attribute. Examples of attributes other than the AS path include ORIGIN, MED, AGGREGATOR, and other attributes. The changed attribute may be either non-transitive or transitive. If the changed attribute is non-transitive, then the SPATH update conveys local information between two directly connected BGP peers and the SPATH update does not propagate beyond the two directly involved peering ASes. However, if the changed attribute is transitive then the new information must be propagated to all Internet routers that use the route.

Virtually all DoD SPATH updates change transitive attributes. In particular, Figure 11 shows that virtually 100% of DoD SPATH updates indicate a change in the optional and transitive AGGREGATOR attribute. A BGP router that performs route aggregation may add the AGGREGATOR attribute to list the router's AS number and IP address.[6] In this case, the AS path used to reach the DoD prefix did

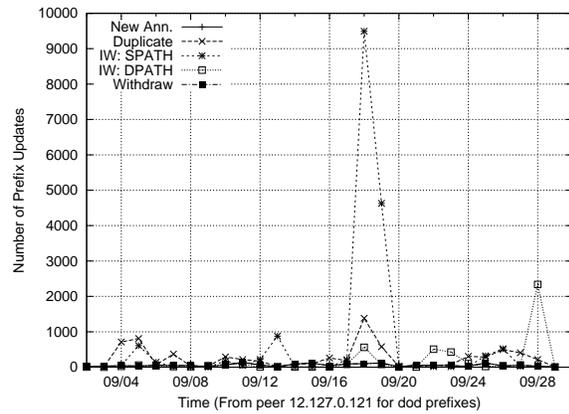
not change, but the AGGREGATOR attribute did change and any change in the AGGREGATOR must be propagated to every Internet router that uses the route. Without the changes in AGGREGATOR attributes, our sample set of DoD prefixes would not have generated a higher than average number of updates. In particular, without the AGGREGATOR attribute changes during the Nimda worm attack, the large spike in DoD prefix updates on 9/18/01 would not have occurred.

6.2. SPATH changes due to the AGGREGATOR attribute

The changes in AGGREGATOR attribute generated a large number of updates for DoD prefixes and these changes had to be propagated to the global Internet. To better understand the purpose and use of the AGGREGATOR attribute, the following figures present a slightly simplified view of



(a) ISP A's view of Internet prefixes



(b) ISP A's view of Dod prefixes

Figure 10. ISP A's view of update classes in September 2001

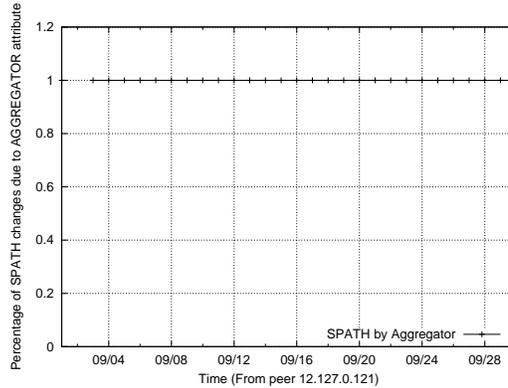


Figure 11. SPATH Updates Due to AGGREGATOR

what we believe occurred to our sample set of DoD prefixes.

In Figure 12(a), routers $R1$ and $R2$ in AS A both announce a BGP route for prefix p . To AS B , the path used to reach p consists of B, A , but AS B can select either router $R1$ or router $R2$ as a next hop. AS A can use the MED attribute to indicate whether $R1$ or $R2$ as the preferred next hop from AS A 's perspective. By changing the MED attribute, AS A can change $R1$ or $R2$ as the preferred next hop. Each change in the MED results in a new SPATH update from AS A , but to AS B the path always remains as B, A and no new update needs to be propagated to AS C . During events such as the Nimda attack, the edge links associated with $R1$ and $R2$ may vary more frequently and could result in a larger number of SPATH updates (changing MED) set from AS A to AS B . But note that this change in MED is a local optimization between AS A and AS B and

does not propagate beyond AS B .

In Figure 12(b), routers $R1$ and $R2$ also perform aggregation before advertising prefix p' and the router performing the aggregation lists its IP address in a transitive AGGREGATOR attribute. Again, AS B can use either $R1$ or $R2$ to reach p' and the AS path is always B, A regardless of whether $R1$ or $R2$ is used as a next hop. However, the AGGREGATOR attribute varies depending on whether AS B selects $R1$ or $R2$. Any local event that causes a change between $R1$ and $R2$ is reflected as a change in AS B 's view of the AGGREGATOR attribute (i.e. the attribute value varies between AGGREGATOR $R1$ and AGGREGATOR $R2$). Even though the AS path used to reach the prefix has not changed and remains B, A , the change in the AGGREGATOR attribute must be propagated to AS C in a SPATH update. Furthermore AS C must propagate this change to its neighbors and so forth.

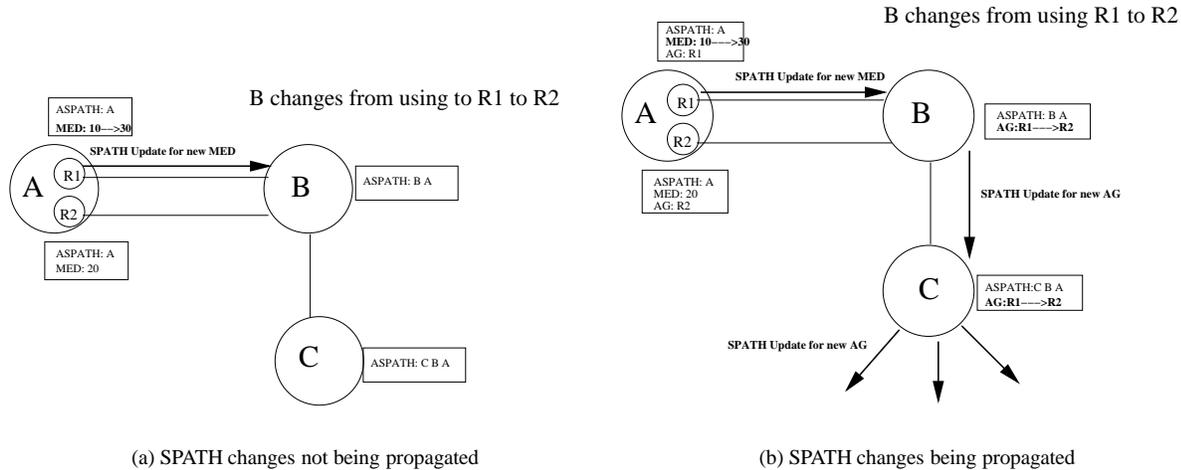


Figure 12. Illustration of SPATH changes

In particular, suppose an event such as Nimda causes oscillations in AS *B*'s choice of nexthop to reach a prefix originated by AS *A*. The resulting changes are no longer localized to AS *A* and *B*. Any change in the local choice of *R1* or *R2* results in a global change to the AGGREGATOR attribute at every Internet BGP router whose AS path includes AS *B*. Note also that the usefulness of the AGGREGATOR attribute diminishes as one moves further from AS *A*. To some distant AS, it matters little whether the aggregation at AS *A* was performed by router *R1* or *R2*, but this information still requires a BGP update at the distant router.

6.3. SPATH updates during code Red/Nimda

The largest spike in DoD updates occurred on 9/18/01. According to the SANS (SysAdmin, Audit, Network, Security) Institute, the scanning activity of the Nimda worm dramatically increased at approximately 1PM GMT on September 18, 2001, and abated in the following hours[9]. In Figure 4(b), one can see a large spike of BGP updates received by the monitoring point around the 9/18/01 and this spike clearly dwarfs all other activities shown in the figure. A similar observation is obtained from July 2001 data when the Code Red worm spread.

Figure 10(b) shows that on September 18 and 19, 2001, SPATH updates accounted for a large percentage of total DoD prefix updates. In fact, 65.4% (16,079 out of 24,578) of all updates were generated during those two days. During these two days, 87.63% (14,090 out of 16,079) of the updates were SPATH updates. By contrast, SPATH updates account for only 38.53% (3,275 out of 8,499) updates seen during other days. Similar observations were obtained from the July 2001 data during the Code Red worm attack. Over-

all, the number of SPATH updates is particularly high during events such as Code Red and Nimda.

In addition, note from Figure 13 that although the set of DoD prefixes only account for less than 0.2% of all Internet prefixes, the set of sample DoD prefixes generated over 80% of all Internet SPATH updates observed in our study during the Nimda worm attack period. These SPATH updates reflect changes in the AGGREGATOR attribute for the DoD prefixes and the transitive nature of this attribute propagated local changes to the global Internet. Even without the attack, the DoD prefixes also typically contribute a disproportionately high number of Internet SPATH updates.

7. Conclusion and future work

To understand how well the BGP protocol design works in reality, we focused on analyzing the routing performance for a sample set of IP prefixes owned by the U.S. Department of Defense (DoD). We used two basic measurements for routing performance: the reachability to each prefix as seen by BGP, and the impact of the BGP updates that each prefix contributed to the rest of the Internet.

Through the analysis of BGP log data over the last two years we observed the connectivity between the Internet and our set of sample DoD prefixes was reliable and that multi-homing is an effective way to handle temporary failures of individual providers. However we also noticed the exception of a few prefixes which suffer poor reachability. In addition, we observed that, although our set of sample DoD prefixes typically contributed no more updates than the Internet prefixes as a whole, our prefix set occasionally generated excessive BGP update messages. This behavior

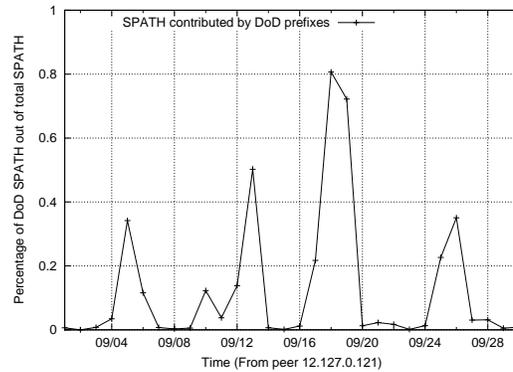


Figure 13. Percentage of SPATH caused by DoD prefixes

resulted from the DoD routers' use of the BGP's optional AGGREGATOR attribute and was especially severe during stressful periods such as the worm attacks.

During the Nimda worm attack, our set of sample DoD prefixes contributed nearly 40 updates per prefix while the Internet as a whole contributed only 3 updates per prefix. Furthermore, our set of sample DoD prefixes contributed to 80% of the total SPATH updates in that period. Our study shows that the AGGREGATOR attribute in BGP, when enabled, allows local routing changes to trigger excessive new update messages that propagate globally to the entire Internet, a behavior which may not have been intended by the protocol design. How to design a protocol that can scale well in a large, dynamic network remains a research challenge; one insight gained from this study is that we must prevent local changes from triggering global messages.

References

- [1] J. Cowie, A. Ogielski, B. J. Premore, and Y. Yuan. Global routing instabilities triggered by Code Red II and Nimda worm attacks. Technical report, Renesys Corporation, Dec 2001.
- [2] R. Govindan and A. Reddy. An analysis of internet inter-domain topology and route stability. In *INFOCOM*, pages 850–857, 1997.
- [3] C. Labovitz, G. R. Malan, and F. Jahanian. Internet routing instability. In *Proceedings of the ACM SIGCOMM '97*, pages 115–26, Cannes, France, September 1997.
- [4] C. Labovitz, G. R. Malan, and F. Jahanian. Origins of internet routing instability. In *Proceedings of the IEEE INFOCOM '99*, pages 218–26, New York, NY, March 1999.
- [5] V. Paxson. End-to-End Routing Behavior in the Internet. *IEEE/ACM Transaction on Networking*, 5(5):601–615, 1997.
- [6] Y. Rekhter and T. Li. Border Gateway Protocol 4. RFC 1771, SRI Network Information Center, July 1995.
- [7] J. Rexford, J. Wang, Z. Xiao, and Y. Zhang. BGP routing stability of popular destinations. In *ACM SIGCOMM Internet Measurement Workshop 2002*, Nov. 2002.
- [8] RIPE. Routing Information Service Project. <http://www.ripe.net/ripenc/pub-services/np/ris-index.html>.
- [9] N. System Administration and S. I. (SANS). Nimda worm/virus report. <http://www.incidents.org/react/nimda.pdf>.
- [10] 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*, Oct. 2002.