Analysis of Routing Characteristics in the Multicast Infrastructure

Prashant Rajvaidya and Kevin C. Almeroth
Department of Computer Science
University of California–Santa Barbara
Santa Barbara, CA 93106-5110
{prash, almeroth}@cs.ucsb.edu

Abstract— **As the multicast-capable part of the Internet continues to evolve, important questions to ask are whether the protocols are operating correctly, the topology is well connected, and the routes are stable. A critical step in being able to answer these questions is to monitor the traffic and network operation. In this paper, we analyze characteristics of the multicast infrastructure over the last three years using monitoring data collected from several key routers. Specifically, we focus on analyzing two characteristics of the infrastructure: size and stability. The size analysis focuses on counting the number of connected hosts and networks, and analyzing how the size of the infrastructure has changed over past three years. Second, the stability analysis focuses on examining persistence, prevalence, and visibility of routes across the topology. From our analyses, we identify a number of problems with multicast routing and their effect on the connectivity of certain multicast networks. Moreover, we offer insight into the evolution and future of multicast in the Internet.**

I. Introduction

Multicast is a service model for scalable, many-to-many delivery of data in the Internet[1]. Scalability is achieved by allowing a source to send each packet only once regardless of the number of receivers. Network-layer multicast uses network elements (routers, switches, etc.) to replicate a packet at as many branching points in the network as necessary to reach all interested receivers. Greater efficiency is achieved by replicating each packet as close to the receiver as possible.

The first deployment of multicast took place in 1992[1], [2]. In the early years, the Distance Vector Multicast Routing Protocol (DVMRP)[3] was the dominant protocol for routing. During this time, research focused on how to build multicast forwarding trees efficiently, provide reliability, and handle congestion. As solutions to these problems have been proposed, several new protocols have been developed, implemented and deployed. In the current infrastructure, use of DVMRP in the inter-domain has been replaced by protocols like the Multiprotocol Border Gateway Protocol (MBGP)[4], Protocol Independent Multicast (PIM)[5], and the Multicast Source Discovery Protocol (MSDP)[6].

In the last decade, although multicast protocol development and implementation has come a long way, its usage has not been as widespread as originally envisioned[7]. Numerous barriers have contributed to the lack of deployment. First, deploying a new service like multicast in a heterogeneous network the size of the Internet is bound to be difficult simply due to the large number of devices that need modification and the personnel needed to manage the service. Second, deployment of network layer multicast, especially at the inter-domain level, has been observed to be a particularly hard problem for a variety of reasons[8], [9]. Finally, problems such as instability of the overall infrastructure, routing peculiarities, political barriers, and hardware lifetimes further complicate the challenge of deployment.

For multicast, challenges in assuring stability primarily stem from the fact that multicast protocols, although robust and mostly scalable individually, have several weaknesses when operating together as a system. These weaknesses have resulted in both poor connectivity as well as significant data delivery problems. As correct operation of inter-domain multicast relies on several protocols operating in harmony across the infrastructure, the instability of the system as well as its complexity increases. In spite of our knowledge of these issues, research in the analysis of protocol operation, performance and deployment has left many unanswered questions. This lack of operational know-how has made both deployment and maintenance of the infrastructure more difficult.

There is very little knowledge about how the multicast infrastructure operates. Most multicast debugging is based on the observation of operational faults[10]. As a result, several of the important questions related to the multicast infrastructure are unanswered. Some of these questions include whether the protocols are operating correctly, whether the topology is well connected, and whether routes are stable. With effective monitoring, these questions can be answered, and many of the existing problems can be identified and subsequently solved. While this belief parallels what is happening in the unicast world[11], [12], [13], very little monitoring and data analysis related research has been done for multicast.

The focus of this paper is to understand the network layer characteristics of multicast routing. The goal is to track the evolution of the multicast infrastructure, identify the existing problems, and gain insights into the cause of these problems.

While other similar research has relied on end-to-end and application layer data collection and analysis [14], [15], this paper focuses on network layer monitoring and uses data collected by observing operation inside the network. The motivation behind using network layer monitoring is that monitoring at the application layer, although relatively straightforward, is not very effective in identifying network state and troubleshooting network problems[16]. When analyzing network problems is the main concern, network layer monitoring is the only effective way to provide efficient monitoring solutions. Another factor that distinguishes our efforts from that of other research is the scope of data collection, i.e. whether data is collected through global, infrastructure-wide, monitoring or through local, intra-domain, monitoring. While other research efforts have focused on collecting data from a single location[7], [17], [18] and have assumed that it is similar to data in other networks, our work is based on data collected from multiple topologically dispersed locations. We believe that collecting data from a single location alone is not effective[19] and that monitoring on a global scale is critical in generating representative infrastructure-wide results.

In this paper we analyze multicast routing characteristics for the last three years using data that represents an aggregate view of the MBGP routing tables. This data has been collected from some of the more important network locations with the help of our global monitoring infrastructure, Mantra. The analysis presented in this paper is an evaluation of the long-term trends evident in the global multicast infrastructure. We focus on analyzing two characteristics: size and stability. With respect to size analysis, our primary contribution is to develop, apply, and then evaluate different metrics to gauge the size of the infrastructure. These metrics are: connectedness of the infrastructure, growth in deployment, and live address space. By tracking long-term changes in size, we can also analyze the evolution of the infrastructure. To assess the stability of the infrastructure we analyze the characteristics of the infrastructure as a whole as well as the individual networks. Specifically, we analyze the persistence and prevalence of routes, and the visibility of routes across the topology.

The remainder of this paper is organized as follows. In Section 2 we describe mechanisms for analyzing routing characteristics. Section 3 presents an overview of the data used in this paper and the data collection mechanism. Section 4 describes results about the size of the multicast infrastructure and Section 5 analyzes network stability results. The paper is concluded in section 6.

II. Analysis Fundamentals

Two of the most important questions we asked when beginning our analysis were: (1) what should we collect, and (2) what metrics should be the focus of our analysis. In this section, we describe these two aspects of our analysis efforts. First we describe the significance of the data available from MBGP routing tables. Second, we build a case for using *address space*, as opposed to *network count*, as the primary metric for comparison.

A. MBGP Routing Table Details

MBGP is an extension of the Border Gateway Protocol (BGP)[20] and is the primary mechanism for exchanging inter-domain route information among multicast-enabled domains. This exchange of information takes place through a series of MBGP-peering relationships. Each such relationship can be defined as a pair of MBGP routers that have been specifically configured to exchange routing information. The network of all such peers constitutes the MBGP topology that, in the ideal case, includes the entire multicast infrastructure. The idea behind this setup is that through a series of exchanges among the peer routers, routing information from one network location will eventually propagate throughout the infrastructure.

The basic unit of routing information exchange is referred to as a *route update*. Through these updates, a multicast router advertises routing information for IP address prefixes to its peers. Also included in this update is a path that can be used to reach hosts included in the IP address prefix. A path is represented as a series of Autonomous System (AS) numbers. An IP address prefix is represented using Classless Inter-Domain Routing (CIDR) notation, i.e. a network-number and network-mask pair. Routing updates in multicast routers are either advertised in real-time, i.e. as soon as a change occurs, or at intervals of no more than 3 minutes. Every multicast router that is an MBGP peer maintains state for all networks from which it has received a routing table update. This table, referred to as the *MBGP routing table*, represents the snapshot of reachable multicast networks from the point-of-view of a particular router.

Processing information from MBGP routing tables is a basic yet powerful mechanism for analyzing various characteristics of the multicast infrastructure. It is of particular importance in analyzing the characteristics that are the focus of this paper, e.g. size and stability. Because an MBGP routing table represents all the reachable multicast networks from a router, this table provides a snapshot of the entire infrastructure from the point-of-view of that router. Therefore, measuring the amount of the infrastructure represented in this snapshot is a direct indication of size. Similarly, tracking its variations over an extended period of time is useful for analyzing the evolution of infrastructure. Finally, analysis of short-term changes is important for analyzing the stability of the infrastructure.

MBGP routing tables from individual routers can also be combined to create an aggregate view of all possible multicast networks. An aggregate view is useful because results can be generated that more closely represent the sum of all possible networks in the infrastructure and not just the view from a single router. This is because although some infrastructure-wide information is available in the routing tables of every

MBGP router, this information is usually not complete[19]. Therefore, global state is not consistent across all routers. Also, all networks are not reachable/seen by all other networks. We generate aggregate views by taking the union of all sets of networks that are reachable from a set of individual routers. The advantage of such an aggregate view is that even though a particular network may not be seen by *all* routers, it might be seen by *some*.

In an ideal case, an aggregate view should be an exact reflection of the true global picture, i.e. it should be a snapshot of all the networks connected to the topology. However, in practice, generating such a snapshot is difficult because it requires aggregating information collected from *all* routers at *exactly* the same time. Nevertheless, we believe that an aggregate view of reasonable quality can be generated if it is composed of information collected at nearly the same time from a small set of key routers that are close to the core of the infrastructure and have peering relationships with other routers in geographically and topologically diverse locations.

B. Units of Measurement

Analysis of routing tables usually involves measuring the amount of the infrastructure represented by the entries in a routing table. For example, the extent of multicast deployment can be calculated by measuring the size of the infrastructure represented by all of the entries in a routing table view. Similarly, instantaneous visibility can be calculated by measuring the fraction of the infrastructure connected at a given instance in time. Results from such calculations can be measured using one of the two basic units: network count and address space. While network count measures the number of unique networks represented by the routing table entries under consideration, address space measures the number of unique IP addresses that constitute these networks. A description of these two units and their evaluation is presented below.

- **Network Count**: Network count is the most widely used unit and is also straightforward to measure. Unfortunately, this metric has several drawbacks. As a result, it does not provide a particularly accurate measure. Most of these drawbacks stem from the fact that network count assigns equal weight to all networks regardless of their size. This presents an inaccurate estimate of the infrastructure's size. For example, although a network with a mask length of 8 can possibly have approximately 2^{24} hosts, it will be given the same weight as a network with a mask length of 32 (a single host). Assigning equal weights to all networks also makes comparing views from different routers impossible. For example, if a network has a mask length of 8 in one routing table but is represented as multiple networks with longer masks in another table, the network counts will appear different even though they are really the same. A final limitation of using network-count as a measurement unit results from anomalies in the MBGP routing tables that can cause the same portions of the infrastructure to be represented in multiple overlapping network entries. For example, if entries for networks 128.111/16 and 128.111.52/24 are both present, there will be a redundant representation for the network 128.111.52/24. Such anomalies, likely caused by mis-configuration of MBGP, can affect the accuracy of results.
- **Address Space**: Using address space as a unit for measuring the size of the infrastructure provides a simple way of weighting the routing table entries and, hence, eliminates the drawbacks associated with using network count. To calculate the amount of address space, the number of addresses in each unique network represented by MBGP entries is computed, (2^{32-net_mask}), and summed. Using address space as a unit, however, comes with its own problems. The primary problem being that address space does not represent the *actual* number of multicast capable hosts, but rather, it represents the *possible*. Therefore, address space is not an accurate measure for the actual number of hosts that are multicast capable. However, address space is still a good way of measuring *relative* size.

Unfortunately, neither network count nor address space can be used to answer the question of how many multicast capable hosts exist. In fact, this question is nearly impossible to answer–almost as hard as answering the question of how many hosts are connected to the Internet. The solution is to keep a proper perspective on exactly what these two measures can tell us. The usefulness is not in the absolute value but in the value relative to itself as it changes over the short- and long-term. Our analysis shows that when the unit of measurement is address space, the results more accurately reflect what is actually happening.

To confirm the claims we have made with regards to these units and to justify that it is better to use address space as a unit for our results, we briefly compare the change in the size of the multicast infrastructure using these two metrics. Figures 1 and 2 show the changes in the size of the infrastructure between August 1999 and June 2002 in terms of network count and address space respectively. Although we analyze these results in detail to assess connectedness and instability in the later sections, in the present context, these results provide an excellent example of the disadvantages of using network count as a unit of measurement. The key observation is that the trend between network count and address space are only loosely correlated. There are often changes in the network count that (1) do not exist, (2) exist and have roughly the same degree of change, or (3) do not fully represent the magnitude of the change. Again, the problem is basically that using only network count does not take into consideration the network mask. There are several noteworthy examples in Figures 1 and

2. For example, a large spike occurs in Figure 1 on August 2000. This spike caused the number of networks to almost double, while the corresponding increase in the amount of address space was less than 25%. A second example occurs during many of the changes in 2000 when the network count seems to be fairly stable, but the address space varies widely and frequently. Before going into more details on these and other results, we first present a more detailed description of the data we have collected and the collection process.

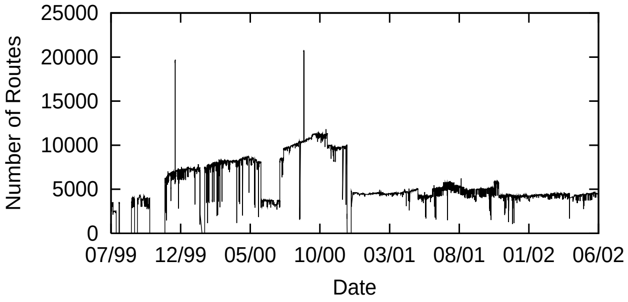

Fig. 1. Unique networks visible in the aggregate view.

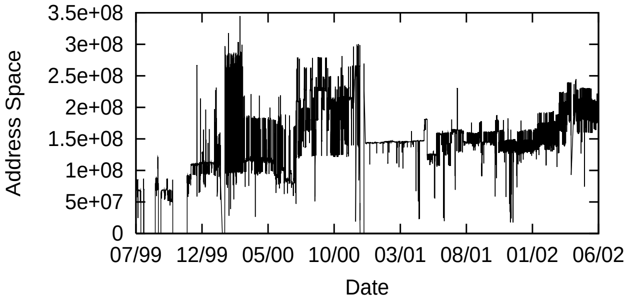

Fig. 2. Unique addresses visible in the aggregate view.

III. Data Collection

The analysis presented in this paper is based on network-layer routing data collected over a three year period using Mantra[16], [21], [19]. Data was collected by capturing state information from a set of four topologically and geographically diverse multicast routers. In this section, we present a brief overview of Mantra, detail when and from where data was collected, and discuss the utility and limitations of the data.

A. Overview of Mantra

Mantra is a system for global monitoring of the multicast infrastructure at the network-layer. It collects data directly from routers by logging into the routers and capturing their internal memory tables. The type of data collected from these routers ranges from routing tables for MBGP and DVMRP, to forwarding tables for PIM, to source announcements for MSDP.

In addition to data collection, other monitoring tasks performed by Mantra include archiving collected data, processing it to generate an aggregate view of the infrastructure, analyzing this view to generate several types of results and presenting the results via a WWW server[1].

Mantra has been in operation for approximately four years and currently monitors 16 routers. Several of these routers are key exchange points and/or border routers for transit providers. Monitoring these routers provides us with data from a topologically and geographically diverse set of network locations. The importance of these collection points in the topology and their diversity is key to collecting an accurate global view of the multicast topology. Mantra results are useful for several purposes including: real-time fault identification, building of traffic weather maps, route analysis, and analysis of host-group behavior.

B. Characteristics of Data

Data sets analyzed in this paper mostly consist of MBGP routing tables collected over a three year period, starting in August 1999, from four routers: Federal IntereXchange–West (FIXW), STARTAP, DANTE, and Oregon IntereXchange (ORIX). There are two main reasons for selecting these four routers. First, we have high-quality data from all of them. Second, routing information available at these routers is more representative of the global view because these routers are close to the core of the infrastructure and each of them has MBGP-peering relationships with routers in several other important networks. For example, *FIXW* is one of the more important multicast exchange points on the West Coast of the United States. *STARTAP* is a core router in Abilene (the Internet2 backbone), and acts as an interface between Internet2 and the commodity Internet. *DANTE* is an exchange point between the US and Dante's high speed European research backbone. Finally, *ORIX* is a router in the Northeast US that peers with several important US and international networks specifically for the purpose of monitoring.

In addition to MBGP data, we have also used DVMRP routing tables collected from *FIXW* during a two year period starting in November 1998. This data has primarily been used to analyze the transition of the multicast infrastructure from DVMRP tunnels (called the MBone) to native deployment of MBGP. DVMRP data from *FIXW* is of particular importance for this analysis because *FIXW* has been a key route exchange point for both infrastructures.

The granularity of the data sets that we have used in this paper is 15 minutes, i.e. fresh routing tables have been collected from the routers at 15-minute intervals. Another characteristic of these data sets is that the routing tables from different routers that correspond to the same time interval depict nearly concurrent snapshots of the routing infrastructure. This was

[1]See the Mantra web site at: http://www.nmsl.cs.ucsb.edu/mantra/

accomplished by synchronizing when data was collected from the routers. These concurrent snapshots have been aggregated and the results are what we have used for our analysis.

C. Limitations of Mantra Data

Mantra contains a significant amount of collection and processing robustness. However, there are always limitations in the collected data sets. Fortunately, these limitations are relatively minor as we believe they do not affect our results. The two most significant limitations are: (1) the concurrency of snapshots from different routers is not guaranteed; and (2) the granularity of data is limited to 15 minute snapshots.

- **Snapshot Concurrency**: Although different views at different routers are collected at nearly identical times, it is not guaranteed that the snapshots represented by these views correspond to exactly the same instance in time. This is the case because although Mantra instantiates the collection processes for all the routers concurrently, the actual instance when snapshot is captured can vary from one router to another.
- **Low Granularity**: Mantra is not able to record *every* time there is a route update. This is because the granularity of data collection is 15 minutes. Therefore, it is likely that our data sets miss changes triggered by fine-grained route flaps, i.e. routing table entries that change and then change back within an interval. However, because our analysis does not rely on knowing all changes, the effect is not significant. To confirm this, in July 1999 we collected MBGP data from ORIX at 5-minute intervals. The aim was to discover the number of updates that occur in a shorter interval and then estimate how many of these updates would have been missed if a longer interval would have been used instead. We estimated the number of updates that would have been missed by Mantra if data were collected at 5, 10, 15, 20, 30, or 45 minute intervals. Our results generally show that the number of missed updates increases significantly as the interval grows more than 20 minutes. However, the number of updates missed for a 15 minute interval constituted only a small portion of recorded updates.

In spite of these limitations, we feel the data sets are representative of the infrastructure for the types of analyses we perform. This is the case because the analyses presented in this paper is long term and mostly involves cumulative statistics. Therefore, the effects of both types of limitations are normalized and have negligible impact on the trends that we discover. Furthermore, inaccuracies in the aggregate view do not have significant effect on the results because our analysis primarily focuses on changes in the aggregate view over time rather than of analyzing the characteristics of a single aggregate view. Finally, data inaccuracies due to loss of granularity have an insignificant effect on our results not only because the loss of information is minimal, but, also because the loss of information in data sets from one router is compensated for by the availability of information from another.

IV. Size and Evolution of the Infrastructure

The fundamental component of an assessment of the multicast infrastructure is a measurement of its relative size. This analysis is important in understanding the extent of multicast deployment. Furthermore, tracking size results over time provides a basis for evaluating how the infrastructure has evolved to its current state, and hopefully, what the near-term trend will be. This general analysis can be divided into three more specific analyses. These include:

1) **Connectedness**: raw number of multicast capable networks and addresses connected to the infrastructure.
2) **Changes in Infrastructure Size**: relative size of the address space over time.
3) **Active Address Space**: address space corresponding to stable and active multicast activity.

Detailed results from each of these analyses are presented in the remainder of this section.

A. Connectedness

Connectedness is the most basic metric for measuring the size of the infrastructure. Connectedness measures the number of networks connected to the MBGP topology at any given instance, i.e. the number of addresses represented by valid route announcements. As already shown and briefly described, Figures 1 and 2 show the changes in the connectedness of the infrastructure between August 1999 and June 2002. We now discuss these results more thoroughly.

Figures 1 and 2 indicate that connectedness within the infrastructure is highly variable. It is evident from these figures that not only are the variations frequent but the degrees of these variations are large. On a typical day, the number of routes varies anywhere from 3000 to 5000, and the amount of address space varies between 100 million to 200 million addresses. Over longer periods of time, variations can be even larger. For example, in November of 2001, while the minimum number of networks visible at any time was 1700, the maximum was 6000. Although these variations suggest that there are at least some problems in stability, we cannot identify any particular reasons for these variations based only on connectedness. In the next section we look more closely at exactly what is happening, and offer more of an explanation.

Using just the connectedness results, one conclusion we can draw based on operational experience is that actual deployment does not vary as much as the results suggest. Because connectedness only depicts the state of the infrastructure at one instance in time and because not all the networks are connected to the topology all the time, it is clear that connectedness results are an under-representation of the size of the infrastructure.

Although connectedness is not an accurate metric for measuring the size of the infrastructure, long term trends in the results in Figures 1 and 2 can be useful in providing a general overview about how the size of the infrastructure has varied over the last three years. One of the most prominent long term trends is that the size of the infrastructure has a "floor". On a typical day, the number of visible networks does not typically go below 3000 and the address space does not go below 100 million. One conclusion that can be drawn is that over the last three years certain networks have a high degree of persistence. In fact, if we ignore the noise caused by frequent variations, the common set of consistent networks and address space has been growing. For the address space, this growth has resulted in a doubling over the last three years. As we will discuss later in this section, there is strong evidence that deployment has indeed increased in the last two years.

In addition to its usefulness in calculating connectedness, information in a routing table can also be used to analyze the distribution of network masks. While shorter masks represent larger numbers of addresses, longer masks represent fewer addresses. On an average day, about 68% of network masks are in the range from 24 bits to 32 bits, 18% are in the range of 16 bits to 23 bits, and 14% are in the range of 8 bits to 15 bits. These results indicate that not many large domains have multicast enabled in all of their constituent networks. It can be further inferred that a majority of multicast networks are very small networks that are connected to the topology at the edges. It is likely that these networks either represent small stub domains or are small multicast capable subnets within large domains.

We have further examined the long-term trend in the evolution of the network mask distribution. We have found that in comparing the second quarter results for the years 2000 and 2002, the network mask distribution has shifted towards mid-size networks, i.e. those with the mask-length 16-23. This suggests multicast is evolving to the point where it is being deployed in larger-sized networks. We have also observed that new multicast networks being added have longer network masks than average. This suggests only small additions are being made. Unfortunately, we have also observed that lost multicast networks have shorter masks, i.e. multicast is being turned off in large networks.

B. Evolution of the Multicast Infrastructure

Over the last three years, a significant number of networks have started announcing that they are reachable for multicast. Ideally, all these networks should be visible in the aggregate view and the connectedness of the infrastructure should have been increasing over the last three years. However, while some new networks and address space can be seen in Figures 1 and 2, these results can not be used to measure growth. Because they only provide an instantaneous count, it is difficult to determine what part of the count represents new addresses, old addresses, and addresses only advertised for a short period of time. Therefore, growth of the infrastructure, as measured by the number of unique addresses that have ever been seen in a routing table, is a better representation of the growth of the infrastructure.

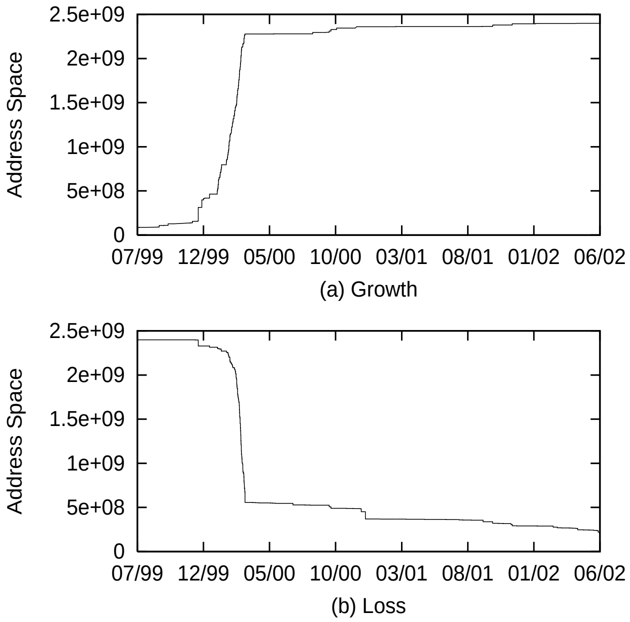

Fig. 3. Growth and loss of the MBGP-reachable address space in the aggregate view.

Figure 3(a) shows the *growth* of the multicast infrastructure over the last three years. This graph shows a line that increases each time a new address or group of addresses is first announced via an MBGP route announcement. Growth results show that there is a clear rise in the size of the infrastructure and that the amount of address space has grown by nearly 50 times over the last tree years, from 50 million to about 2.5 billion addresses.

Of interest is that from November 1999 until March 2000, growth in the address space has been in spurts. Since March 2000, growth has been more gradual. Our belief about why this has happened is that growth early in the collection period was a result of the sporadic introduction of several large networks. However, after March 2000, only smaller networks have joined the topology. Further investigation into this behavior reveals that most of the growth until 2000 was, in fact, not really growth in the deployment at all. Instead, it was simply a switch from the use of DVMRP to the use of MBGP. Although MBGP deployment started in the early part of 1999, for almost a year both the DVMRP and MBGP topologies co-existed and DVMRP remained the dominant routing protocol. The rate of transition from DVMRP to MBGP increased only at the end of 1999 when several larger ISPs stopped using DVMRP and deployed MBGP. This transition was mostly completed by March 2000.

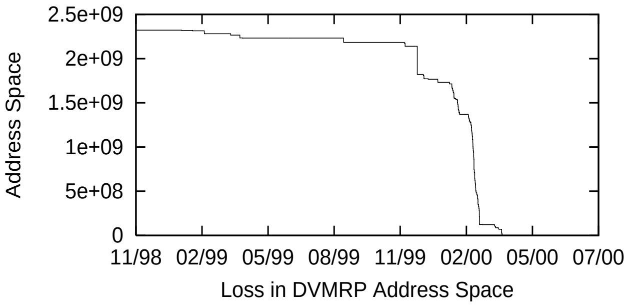

Fig. 4. Loss of the DVMRP-reachable address space as seen at FIXW.

To further substantiate this conclusion, Figure 4 shows loss in the address space reachable via DVMRP at FIXW between November 1998 and July 2000. The graph in this figure takes as its starting point the total number of unique addresses that we have observed to be advertised via DVMRP during the collection period. The graph then decrements at the point in time when each address stops being advertised. This figure confirms that the use of DVMRP started to decline in February of 1999 and almost completely ceases by March 2000.

An interesting observation can be made by comparing the range of values in Figure 3(a) to those in the connectedness results in Figure 2. This comparison shows that there is an order of magnitude more addresses that have been advertised than can typically be found in the MBGP table. In the last three years the number of addresses that are connected at each measurement point is between 100 and 150 million, but, the *total* number of addresses seen in the same period is about 2.4 billion. Assuming that all 2.4 billion addresses are valid, more than 95% of the infrastructure was not connected and, hence, not reachable. Obviously many of the advertisements that were made during the collection period were not long-lived. Countering the growth caused by MBGP announcements received by new networks is the departure, never to return again, of networks from the infrastructure.

Therefore, the "growth" of the infrastructure is offset by networks that stop announcing their MBGP routes. These networks represent "loss" in the infrastructure. While examining growth of the infrastructure shows how rapidly multicast is being deployed, it can only be accurately interpreted by counter-balancing with the networks that have been lost. Figure 3(b) shows the *loss* of networks in the infrastructure. Like Figure 4, this graph takes as its starting point the total number of addresses announced via MBGP over the last three years ($2.4x10^9$). The graph then decrements this number by the number of addresses at the point in time when an MBGP advertisement is observed for the last time. We can know the "last advertisement time" because we have all of the data up to the end of the collection period.

The basic idea is that growth shows the cumulative number and timing of newly advertised address space while loss shows when address space stops being advertised. The ideal scenario for the deployment of a new service like multicast is that growth is exponential and loss does not happen. However, results in Figure 3(b) show considerable loss. Ideally, this curve should be a straight line running along the value of $2.4x10^9$ addresses. This would show that an advertisement encompassing every address was received till the end of the collection period. Unfortunately, this is not the case and many networks that once announced that they were multicast capable no longer do so. Although this has been gradual recently, large loss was noticed for a period of about 4 months starting in December of 1999. The most significant drop occurred around March 2000. As discussed above, this was the period when many networks transitioned from DVMRP to MBGP. There are three likely reasons for this. First, as the popularity of DVMRP declined, fewer service providers were willing to offer tunnels. Some networks would have lost their only option for multicast service. Second, a number of network providers attempted to transition from DVMRP to MBGP but found that they could not provide a robust service and so turned it off. Finally, it is also possible that some unicast routes were accidentally and wrongly advertised as multicast routes. This was observed to have happened a significant number of times when DVMRP was still a dominant protocol.

C. Active Address Space

None of the results above give a particularly accurate representation of the size of the *active* infrastructure. The aggregate address space, shown in Figure 2, likely underestimates the size because it is only a snapshot in time that may miss a temporarily unadvertised range of addresses. The growth graph in Figure 3 overestimates the size because it includes all addresses ever advertised via MBGP. In this section, we develop a new measure, *active address space*, that provides a more accurate estimate of the size of the infrastructure. Again, it is important to mention that this is not an actual count of active multicast hosts, but rather a more accurate, but still relative, estimate of the size of the multicast infrastructure.

To eliminate short-term aberrations and to provide a better estimation of "active addresses", we first define an active address as one that has been announced at least once prior to the current measurement period and will be announced at least once after. The creation of this term is somewhat arbitrary. Obviously there are many possible definitions of what an "active address" is. However, our goal is simply to analyze our data with a stronger bias towards persistent addresses.

Figure 5 shows the active address space represented in the aggregate view over the last three years. These results show that when measured in terms of the number of active addresses, the extent of multicast deployment has almost doubled in the last three years. This is consistent with both our intuition and the (qualified) results presented in previous sections. While this seems like a positive result, it should be considered

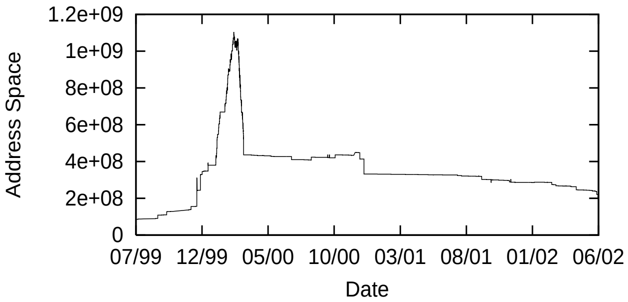

Fig. 5. Aggregate view of the active multicast address space.

only after taking the trend, which we discuss below, into consideration.

Another relevant point about Figure 5 is that for a brief period of time in March 2000, the size of the infrastructure was very high and was more that 5 times of its current value. This can be explained using the growth and loss results that we have just described. The sharp rise in the number of active addresses corresponds to the transition of the infrastructure from DVMRP to MBGP. The sharp fall, corresponds to the fact that most of these new addresses were soon lost. Except for this short interval, the infrastructure's size has remained fairly stable and is mostly in the range of 30 million to 40 million addresses.

The most alarming observation derived from these results is that the number of addresses has been consistently decreasing since July 2001. In fact, since December 2000, there is no indication of any relative increase in the number of active addresses. Whatever little growth that has occurred in the last two months in Figure 3 seems to have been counteracted by a corresponding increase in loss. This leads us to the conclusion that there has been little real growth in the overall size of the multicast infrastructure in the last year.

The fact that the multicast infrastructure has been decreasing in size over the last year needs to be dissected carefully, especially in comparing it to the claim mentioned above that the size of the infrastructure has actually doubled in the last two years. While there has been overall growth, even in the number of stable addresses, over a longer period of time, the recent trend has been a decrease in size.

Considering the relative changes in size that have occurred, we have attempted to understand what specific kinds of losses we are seeing. Our conclusion, further justified using results from the next section, is that much of the decline is due to the elimination of non-multicast-capable address ranges from the MBGP table. This is especially true during the April 2000 spike shown in Figure 5. During the transition from DVMRP to MBGP, many addresses were advertised as being multicast capable when in fact they were not. As ISPs learned to correctly configure MBGP, these inaccuracies were removed. Therefore, Figure 5 shows both a spike and then a return to a more normal routing table size.

D. Discussion

Our analyses of size and the change in size over time answers some important questions but creates a new set. Before identifying and then answering these new questions, it is worthwhile to summarize the most important conclusions of our analysis so far. These include:

1) Active address space is a more representative metric than network count for measuring the size of the infrastructure.
2) The multicast capable address space has almost doubled in the last three years. However, recent growth has been slow or non-existent.
3) Most of the growth has been countered by an equal or larger amount of loss.
4) By the end of the first quarter of 2000, use of DVMRP was replaced with inter-domain MBGP.
5) Connectedness results show that there is significant variability in the number of consistently advertised networks and addresses.

While these results tell us a great deal about the state of multicast, they miss answering a key question. While we know that there are many multicast capable hosts (represented by network advertisements), we do not have a quantitative understanding of whether there is a difference in *stability* among these addresses. In addition other questions now become important including: once an address is advertised, what is its lifetime?; are addresses that have been part of the infrastructure longer more stable?; and is the *floor* that we see really composed of the same set of stable addresses or is rotating among a set of relatively unstable addresses?

V. Stability of Multicast Routing

The correct operation of MBGP and robust multicast routing are two keys to the success of multicast and its ongoing deployment. Nevertheless, in the previous section we have seen that the connectedness of the global infrastructure is highly unstable. We have also seen that there is a set of consistently stable addresses, i.e. addresses that have been advertised since the beginning of their lifetime and continue to be announced even today. An important focus then becomes the unstable addresses. Are these addresses "consistently unstable" or do these addresses start as unstable and then either become stable or disappear completely? Answering these kinds of questions and analyzing the stability of the infrastructure is the focus of this section.

Stability of the infrastructure can be defined as its ability to deliver data packets to all the multicast capable hosts consistently and efficiently. Using the aggregate view data that we have collected, we are able to evaluate stability using three measures. These include:

1) **Infrastructure Visibility**: the fraction of active addresses (see Section IV.C) that are part of the infrastructure at any given point in time.
2) **Address Lifetimes**: time from when a multicast address is first advertised to the time when it is last advertised.
3) **Address Prevalence**: fraction of an address's lifetime in which it is reachable.

We analyze infrastructure visibility for the entire collection period using the complete data set, i.e. from August 1999 to June 2002. However, lifetime and prevalence analyses have been conducted only for the addresses that have been active since April 2000. We exclude results for addresses that were no longer advertised after April 2000 in order to eliminate the effects of the infrastructure transition from DVMRP to MBGP.

A. *Infrastructure Visibility*

In addition to using the results in Section IV as an estimate of size, they are also very useful in providing a first order indication of the stability of the infrastructure. The inconsistency in the number of addresses in Figure 2 shows that only part of the address space is connected to the global infrastructure at any given time. Comparing the connectedness results with those for active address space (Figure 5) show that not all of the active addresses are part of the infrastructure at any given time. Therefore, a good portion of the address space has poor reachability. We define this property as the stability characteristic, *visibility*, and analyze its change over time.

Infrastructure visibility can be defined as the percentage of *active address space* that is connected to the infrastructure at any point in time. While connectedness results in Figure 5 show that the absolute number of reachable addresses, the visibility results are measured as a percentage. We use "active address space" as opposed to "absolute address size" since it more accurately shows the percentage relative to current address size instead of to every address ever seen. If we had used total number of addresses (Figure 3(a)), our results would have been much worse. And if we had used the number of addresses as measured at a specific point in time (Figure 2), the results would have been much better. However, as discussed in Section IV, neither of these are as good a relative measure as the results shown in Figure 5.

Ideally, infrastructure visibility should be 100% at each measurement point. However, this is usually not the case. Figure 6 plots the visibility of the multicast infrastructure over the last three years. These results show that infrastructure visibility has been very poor. In fact, until the third quarter of 2001, the visibility was less than 50%. This implies that during this period half of the address space was not reachable at any given time. However, the visibility has become more consistent and has been as high as 90% in the last several months. While these are good relative results, the bottom line is that there is still a significant degree of volatility with variations between 70% and 95%.

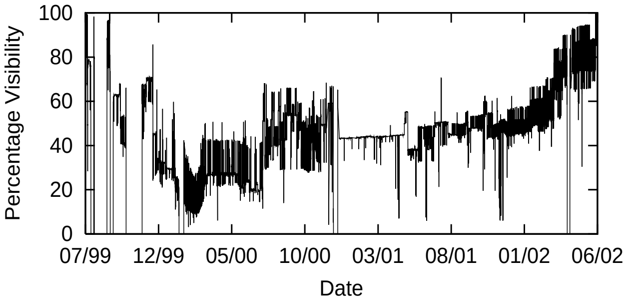

Fig. 6. Visibility of the multicast infrastructure.

An interesting observation can be made by comparing results in Figure 6 with those in the Figure 5. This comparison shows that in the last year there has been both a decrease in the active address space, and an increase in visibility. Improvements in MBGP configurations mentioned previously are what we believe is the key reason for improved visibility. As the size of the multicast address space declines, the percentage of stable addresses increases and, hence, visibility increases. Further support for our conclusions requires answering three additional questions: (1) how long are addresses advertised?; (2) how unstable were the addresses that we no longer hear from?; and (3) has stability also increased in the addresses we do hear from? These questions cannot be answered just by looking at visibility results. In the rest of this section we look at additional results to answer these questions.

B. *Lifetime of Multicast Addresses*

In Figure 5 we have shown results corresponding to the number of live addresses. Another metric to consider is the "lifetime" of an address. The lifetime of a multicast address can be defined as the duration of the interval between the first and the last time the address is carried in an MBGP route announcement. Figure 7 plots the distribution of lifetimes for the addresses seen since August 1999[2]. Ideally, most of the addresses should have very long lifetimes. However, these results show that this is not the case. The average lifetime is only 537 days. Furthermore, only about 46% of the addresses have a lifetime greater than two years. Furthermore, 23% of the addresses have lifetimes less than one year. Finally, about 6% of the address space has a lifetime less than one day. The most interesting observation is that a significant portion of the address space, more than 7%, has a lifetime equal to the entire data collection period. This implies that a reasonable number of networks have been connected to the infrastructure at least since the start of native MBGP deployment.

It is important to note that lifetime results only give absolute numbers and do not provide a very good metric for measuring

[2]As mentioned at the start of this section, to eliminate the effects of the infrastructure transition from DVMRP to MBGP, only those addresses are accounted for that have been present in the routing tables at least once since April 2000.

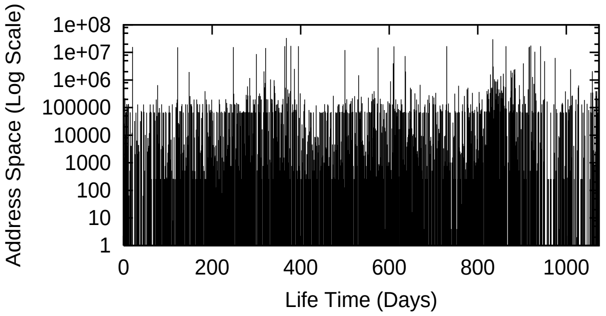

Fig. 7. Lifetimes of multicast addresses.

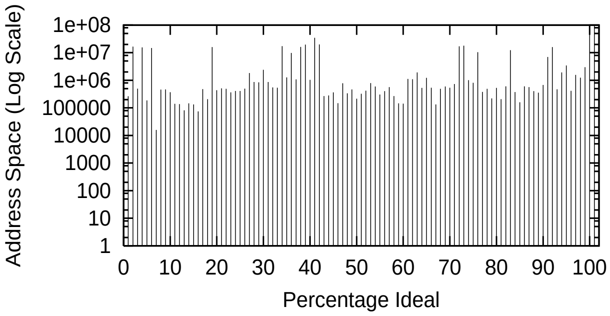

Fig. 8. Lifetimes of multicast addresses as percentage of ideal.

stability. This is because these results do not take into account how long ago an address was first seen. For example, an address range that was first advertised a week ago can only have a maximum lifetime of one week. What is important to know is whether an address stayed in the infrastructure from the time it was first seen to the end of the data collection period. However, as Figure 3(b) shows, there is a good deal of loss before the end of the collection period. Therefore, it is important to compare *ideal* lifetimes with *actual* lifetimes. Figure 8 shows the distribution for address lifetimes when represented as a percentage of the ideal lifetime. Ideally, all these percentages should be 100%. However, this is not the case. Only 50% of the addresses have a ratio greater than 90%. There is some reasonable amount of stability though: more than 31% of the addresses have a ratio of 100%. However, because the infrastructure has not shown much growth lately (Figure 5), many of these addresses not only have a high percentage but also have long lifetimes. Of particular interest is that most of these addresses correspond to networks that have been actively using multicast. When multicast is used, the infrastructure is maintained and is therefore more stable.

C. Prevalence of Address Space

Multicast address *prevalence* is defined as the fraction of a lifetime in which the address is reachable. Analysis of address prevalence is very important because lifetime results are not a particularly good metric for the stability of the infrastructure. In other words just because an address has a lifetime that is long, does not necessarily mean that it was visible at all times.

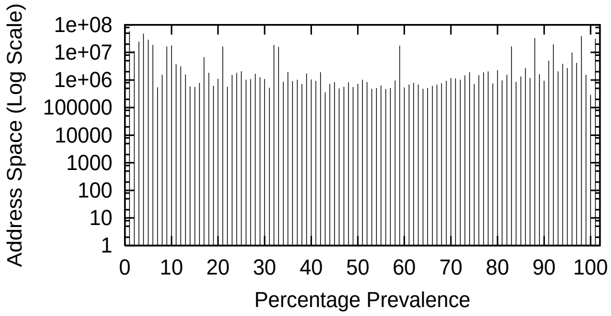

Fig. 9. Prevalence of multicast addresses.

In fact, as we have already shown, both connectedness as well as the visibility of the infrastructure is unstable. Therefore, although some addresses have long lifetimes, their prevalence is likely to be poor.

Figure 9 illustrates the distribution of prevalence values for the address space seen since August 1999. The best case is for an address to have 100% prevalence, i.e. it was seen at every single measurement point during its lifetime. However, as seen in these results, this is not the case very often. Less than 21% of the address space has a prevalence greater than 90%. The most alarming observation is that more than 40% of the address space has a prevalence less than 20%, i.e. these addresses were seen only once in every five measurement points. Figure 9 also shows that there is a concentration of address space at both ends of the x-axis. This trend is consistent with other results we have shown. We can now also say that there are some entries that are consistently stable, and some entries that are consistently unstable. The main reason for the instability in the connectedness and visibility results can be traced to this second set of addresses.

Even though there is a concentration of addresses near the upper end of the prevalence graph, a majority of addresses have low prevalence. Even worse is that the short-lived addresses, which could theoretically have a higher prevalence, do not. Our reasoning is that the likely scenario is that a network administrator tries to deploy multicast, cannot get it to work well (low prevalence), and quickly gives up (short lifetime).

To carry this analysis one step further, Figure 10 plots the prevalence percentage versus address lifetime. While no overall correlation can be made about most of this data, there are several noteworthy trends. For almost all the lifetime durations, prevalence is fairly uniformly distributed. This seems somewhat counter-intuitive since it suggests that prevalence is independent of lifetime. Our intuition (as given above) is that poor performing networks will more quickly stop participating. This intuition should likely be adjusted to reflect the higher patience levels exhibited by users of new services in research networks.

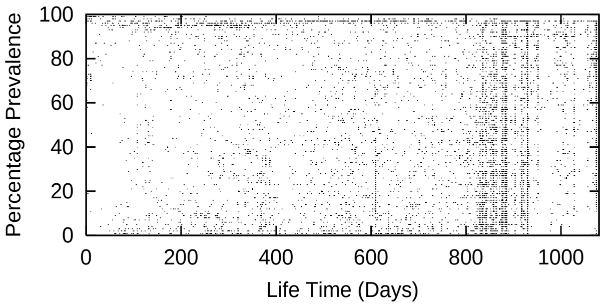

Fig. 10. Prevalence of multicast addresses versus lifetimes.

VI. Conclusions

In this paper we have presented our work on the long-term analysis of multicast routing information collected through MBGP routing tables. This data has been collected from a topologically and geographically diverse set of routers using *Mantra*. Our analysis has focused on investigating the size of the infrastructure from a routing point-of-view and on the stability of these routes. Our analysis has further refined these two areas to include: (1) measuring the extent of deployment, (2) studying the growth and loss of multicast over the last two years, (3) examining and quantifying the instabilities that exist in multicast routing, and (4) evaluating the trends in multicast deployment.

Our bottom-line conclusion is that the multicast infrastructure is slowly but steadily shrinking. However, this is not to say that new networks are not being added, to the contrary. New networks and new address space continue to be added, but the number of addresses no longer advertised is slightly larger than this growth. One very positive note though is that the stability of the infrastructure that remains is much improved. Our historical analysis has shown that there was tremendous instability when inter-domain multicast was created, DVMRP dropped, and MBGP deployed. In fact, much of the instability was attributable to causes other than the transition itself, e.g. mis-configuration problems, protocol bugs, etc. This transition happened more than two years ago and since that time, the infrastructure has continued to stabilize, i.e. the routes that are announced are advertised consistently.

The end result is that when the infrastructure was unstable, many network providers were unwilling to deploy multicast as a service, or did so only for a short period of time. But, the infrastructure has recently stabilized significantly, due in large part to fixed bugs and operational experience. The real question now is whether this stability will be able to reverse the trend of decreasing size.

Acknowledgment

The authors would like to thank k claffy and the Cooperative Association for Internet Data Analysis (CAIDA) for their support and feedback on our work on Mantra.

References

[1] K. Almeroth, "The evolution of multicast: From the MBone to inter-domain multicast to Internet2 deployment," *IEEE Network*, January/February 2000.

[2] S. Casner and S. Deering, "First IETF Internet audiocast," *ACM Computer Communication Review*, pp. 92–97, July 1992.

[3] D. Waitzman, C. Partridge, and S. Deering, "Distance vector multicast routing protocol (DVMRP)," Internet Engineering Task Force (IETF), RFC 1075, November 1988.

[4] T. Bates, R. Chandra, D. Katz, and Y. Rekhter, "Multiprotocol extensions for BGP-4," Internet Engineering Task Force (IETF), RFC 2283, February 1998.

[5] S. Deering, D. Estrin, D. Farinacci, V. Jacobson, G. Liu, and L. Wei, "PIM architecture for wide-area multicast routing," *IEEE/ACM Transactions on Networking*, pp. 153–162, Apr 1996.

[6] D. Meyer and B. Fenner, "Multicast source discovery protocol (MSDP)," Internet Engineering Task Force (IETF), draft-ietf-msdp-spec-*.txt, November 2001.

[7] K. Almeroth, "A long-term analysis of growth and usage patterns in the Multicast Backbone (MBone)," in *IEEE Infocom*, Tel Aviv, ISRAEL, March 2000.

[8] C. Diot, B. Lyles, B. Levine, and H. Kassem, "Requirements for the definition of new IP-multicast services," *IEEE Network*, January/February 2000.

[9] K. Sarac and K. Almeroth, "Supporting multicast deployment efforts: A survey of tools for multicast monitoring," *Journal of High Speed Networking–Special Issue on Management of Multimedia Networking*, March 2001.

[10] D. Massey and B. Fenner, "Fault detection in routing protocols," in *International Conference on Network Protocols (ICNP)*, Toronto, CANADA, November 1999.

[11] V. Paxson, "End-to-end Internet packet dynmaics," *IEEE/ACM Transactions on Networking*, vol. 7, no. 3, pp. 277–292, June 1999.

[12] V. Paxson, J. Mahdavi, A. Adams, and M. Mathis, "An architecture for large-scale Internet measurement," *IEEE Communications*, vol. 36, no. 8, pp. 48–54, August 1998.

[13] S. Kalidindi and M. Zekauskas, "Surveyor: An infrastructure for Internet performance measurements," in *Internet Global Summit (INET)*, San Jose, California, USA, June 1999.

[14] K. Sarac and K. Almeroth, "Monitoring reachability in the global multicast infrastructure," in *International Conference on Network Protocols (ICNP)*, Osaka, JAPAN, November 2000.

[15] R. Chalmers and K. Almeroth, "Modeling the branching characteristics and efficiency gains of global multicast trees," in *IEEE Infocom*, Anchorage, Alaska, USA, April 2001.

[16] P. Rajvaidya and K. Almeroth, "A router-based technique for monitoring the next-generation of internet multicast protocols," in *International Conference on Parallel Processing (ICPP)*, Valencia, Spain, September 2001.

[17] B. Mah, "Measurements and observations of IP multicast traffic," Tech. Rep. UCB/CSD-93-735, University of California at Berkeley, March 1993.

[18] T. Munzner, E. Hoffman, K. Claffy, and B. Fenner, "Visualizing the global topology of the MBone," in *IEEE Symposium on Information Visualization*, San Francisco, California, USA, October 1996.

[19] P. Rajvaidya and K. Almeroth, "Building the case for distributed global multicast monitoring," in *Multimedia Computing and Networking (MMCN)*, San Jose, California, USA, January 2002.

[20] Y. Rekhter and T. Li, "A border gateway protocol 4 (BGP-4)," Internet Engineering Task Force (IETF), RFC 1771, March 1995.

[21] P. Rajvaidya and K. Almeroth, "A scalable architecture for monitoring and visualizing multicast statistics," in *IFIP/IEEE International Workshop on Distributed Systems: Operations & Management (DSOM)*, Austin, Texas, USA, June 2000.