Multicast Routing Instabilities

Native multicast is critical for scaling the delivery of high-bandwidth data, such as audio and video, to multiple receivers. Until recently, however, multicast routing has been unstable, thus making it difficult to ensure robust data delivery. The authors' in-depth analysis of routing instabilities in the multicast infrastructure seeks to identify underlying inconsistencies among routing views at different routers and characterize instabilities by evaluating temporal and spatial variations in these views. Results are based on multiprotocol border gateway protocol (MBGP) routing tables.

Prashant Rajvaidya and Kevin C. Almeroth
University of California, Santa Barbara

Multicast[1] is one of the best available service models for scalable, many-to-many delivery of data on the Internet. Multicast achieves scalability by letting sources send a packet only once, regardless of the number of receivers. In network-layer multicast, network elements (routers, switches, and so on) replicate packets at network branching points, delivering copies to all interested receivers. Replicating each packet as close to the receiver as possible improves efficiency. Multicast uses range from the delivery of conventional multimedia streams such as audio and video to emerging services such as desktop streaming, whiteboards, and collaborative distance learning.

As the multicast infrastructure has matured, several limitations have come to light. Although the infrastructure has recently become more robust,[2] weaknesses in the underlying protocols have historically resulted in poor connectivity and significant routing instabilities. Although the severity of multicast routing problems is well known, little research toward solving them exists. Consequently, researchers know little about how the multicast infrastructure operates, how to solve existing problems, and what other problems exist.

This article seeks to remedy this lack of knowledge by analyzing multicast routing instabilities at individual routers and comparing views between them. Our goal is to present a network-layer view of multicast routing stability and to characterize instabilities by evaluating temporal and spatial variations in the routing views. We also explore potential causes of these instabilities.

Routing Protocols

In the early 1990s, multicast existed as a tunnel-based infrastructure — the multicast backbone (MBone) — that used the Distance-Vector Multicast Routing Protocol (DVMRP). In 1997, advancements in protocol development and subsequent deployment had considerably diminished DVMRP's use at the interdomain level, and

 Published by the IEEE Computer Society 1089-7801/04/$20.00 © 2004 IEEE

the MBone had ceased to exist. Since then, native deployment has been growing steadily.[2] Interdomain routing in the current multicast infrastructure relies primarily on two protocols[3]:

- Multiprotocol Border Gateway Protocol (MBGP),[4] the interdomain route-exchange protocol; and
- Protocol-Independent Multicast (PIM),[5] the main routing protocol, which uses MBGP route information to create and manage data-distribution trees.

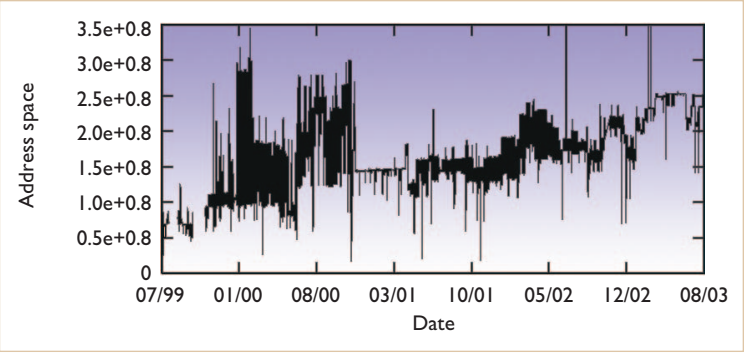

Figure 1. Number of addresses visible in an aggregate, or macroscopic, view.

Poor robustness is largely due to routing instabilities, which are not unique to multicast; they are also common in the Border Gateway Protocol.[6,7] Because MBGP is a BGP extension, its instabilities are not surprising, but they have been particularly severe[2] and have led to significant multicast reachability problems.[8] (A network's reachability is the portion of the infrastructure to which receivers can successfully send join messages.) Furthermore, multicast routing instabilities have a greater impact on data delivery because multicast data delivery occurs through dynamic sets of network links that constitute data-distribution trees. Unstable routes can force networks to reconstruct existing distribution trees, disrupting ongoing multicast data communication for multiple hosts over an extended time.

Our previous work toward understanding multicast's operation tracked the infrastructure's evolution, gauged the extent of deployment, and assessed the stability characteristics of individual multicast addresses.[2] From a routing stability perspective, results for aggregate visibility — an instantaneous measure of the address space visible in the routing tables of all data-collection points — are particularly interesting. Figure 1 plots visibility results for a four-year period beginning in July 1999. As the figure shows, visibility within the infrastructure is highly variable. Route stability problems are a root cause of these variations.

Microscopic Analysis

Because the results in Figure 1 are based on a macroscopic analysis of infrastructure-wide data, the figure is useful only for determining the stability of multicast routing. This aggregated view of multiple network locations misses details about individual routers' stability levels. Consequently, these results are effective neither for quantifying the degree of instability nor for understanding its causes. An in-depth study of routing instabilities, therefore, requires router-level microscopic analysis that evaluates stability trends at individual routers.

Data Characteristics

To find peculiarities in routing instabilities, we studied differences among MBGP routing tables collected between July 2000 and June 2001 using Mantra,[9] our global monitoring infrastructure. We collected data from four routers at 15-minute intervals:

- Federal IntereXchange–West (FIXW), one of the more important multicast exchange points on the US West Coast;
- Science, Technology, and Research Transit Access Point (STARTAP), a core router in the Abilene Network that acts as an interface between Internet2 and the commodity Internet;
- DANTE, an exchange point between the US and DANTE's (Delivery of Advanced Network Technology to Europe) high-speed European research backbone; and
- Oregon Interexchange (ORIX), a router in the northwest US that peers with several important US and international networks.

We used data from this time period because:

- we had high-quality data from all four routers;
- the infrastructure's visibility changed from highly unstable to quite stable during this period (see Figure 1); and
- during the high-instability period, visibility trends were similar to trends observed in data sets collected in 2003.

The data thus give us an ideal opportunity to analyze not only MBGP routing characteristics during high- and low-instability periods but also instabilities that are similar to current instabilities.

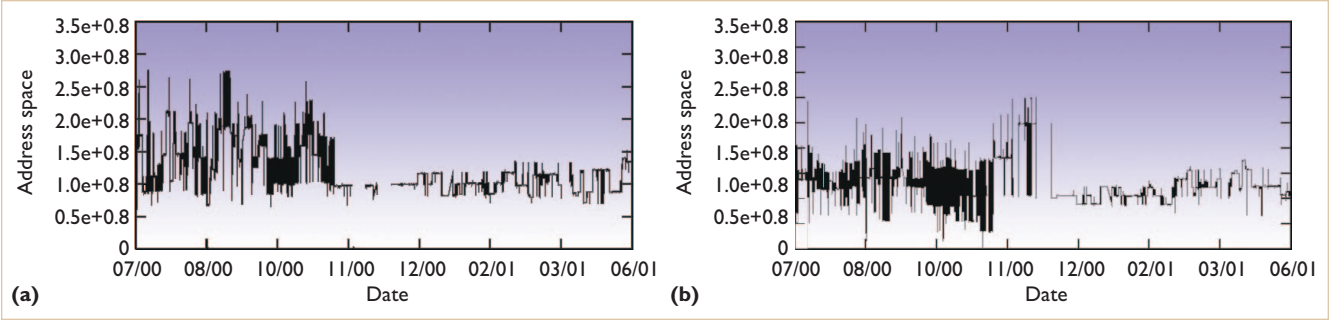

Figure 2. Router visibility results: (a) number of addresses visible at FIXW and (b) difference in visibility at FIXW and ORIX.

Address Visibility

In theory, each routing table entry identifies part of the infrastructure reachable from the router. We use address visibility to determine reachability. Visibility results, therefore, provide a straightforward way of measuring routing instabilities. This is because multicast data delivery depends on consistent and stable visibility at all infrastructure locations.

As a starting point for our study, we analyzed address visibility from the viewpoint of the FIXW, STARTAP, DANTE, and ORIX routers. Ideally, all routers would have stable and identical visibility results. On the contrary, our results showed two key trends:

- Temporal variations. Visibility at each router varied frequently and to high degrees during the analysis period.
- Spatial variations. Visibility is inconsistent across the four routers, and the address space available in the respective routing tables differs.

Temporal and spatial variations likely inhibit multicast operation. The presence of temporal variations indicates that routers cannot always reach some portions of the network, making data delivery unreliable. The presence of spatial variations indicates that some addresses are visible to only a subset of routers – evidence of severe network reachability problems.

What do these variations mean for end users? Consider a multicast audio conference that includes several participants across domains. The presence of temporal variations implies poor audio quality, and the presence of spatial variations implies that each group member could see a different set of participants. Our study of router address visibility shows that significant spatial and temporal variations exist and that, during the analysis period, address visibility was neither consistent across routers nor stable at any router we measured.

Figure 2 shows two sets of study results. Figure 2a plots the visibility results for FIXW, and Figure 2b plots the difference between visibility at FIXW and ORIX. We calculate differences by subtracting the visibility at FIXW from that at ORIX.

Figure 2 confirms an important conclusion we have reached using the aggregate visibility analysis: prior to November 2000, frequent temporal and spatial variations made the multicast infrastructure's robustness quite poor. After November 2000, stability increased as temporal variations decreased. Although spatial variations were still prevalent, they were significantly smaller. Our visibility analysis results also show that not only are some routers more stable – that is, the degree and frequency of temporal variations are smaller – but their visibility is also consistently higher. As with unicast,[7] multicast routing is more stable at certain routers.

Although visibility results provide useful insights into multicast routing's stability, they are more suitable for observing general trends. In addition, it is difficult to reach specific conclusions or quantify the extent of spatial and temporal variations using the results. Because this quantification can help clarify instabilities' extent and effect, we analyze routing stability from the viewpoint of individual addresses.

Address-Level Analysis of Variations

Extending the concept of reachability, we define a reachability rank for categorizing addresses based on whether each of the four network locations identified earlier knows the next hop to which group join messages can be sent. This lets us rank addresses on a scale of 1 to 4 (that is, the number of routers from which the address is visible) and provides a basic yet powerful mechanism for quantifying spatial variations. Our goal is to measure the population of addresses for each rank and then track variations to further analyze routing instabilities. Figure 3 plots the summary of rank

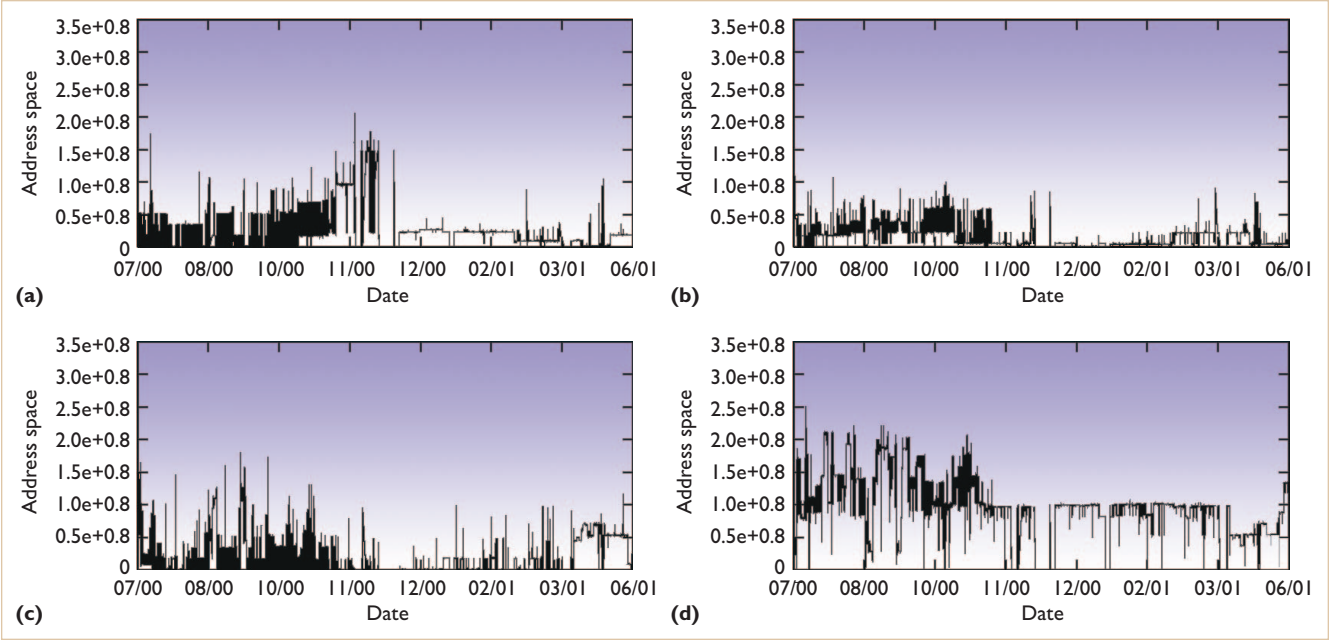

Figure 3. Address reachability rank. The graphs plot the number of addresses visible at (a) one router, (b) two routers, (c) three routers, and (d) four routers.

results for our analysis and quantifies the number of addresses seen by different numbers of routers.

Ideally, each address would be visible at all routers, with a corresponding reachability rank of 4. Rank results show that this is not the case, however, and that address reachability is poor. Throughout the analysis period, no more than two-thirds of the aggregate space was visible at all four routers, and reachability between July and November 2000 was the worst. Furthermore, for 14 percent of the data points, less than 20 percent of the address space was visible at all four routers.

Results in Figure 3 further confirm that some routers have consistently higher visibility. Each graph contains instances in which the number of addresses stays above a certain value. For example, from December 2000 to February 2001, the number of addresses with rank 1 did not go below 30 million. Similarly, between February and March 2001, the number of addresses with rank 2 did not go below 30 million. A close examination of the data shows that during this period, a constant set of 22 million unique addresses was exclusively visible at STARTAP. The results also show that stable routing doesn't necessarily mean good reachability. Even with few or no temporal variations, addresses might not be visible at all routers. Results from after November 2000 highlight this fact: despite increased stability, several addresses had rank values of less than 4. In January 2001, for example, there were hardly any temporal variations, but all four routers could reach less than 75 percent of the aggregate address space. Although this isn't particularly bad compared to results from before November 2000, the 25 percent of addresses that not all routers can be a source of serious reachability problems.

Reachability results provide more information about the source of instabilities. A closer look at addresses with changing ranks reveals that the same set of about 120 million addresses caused most of the instability during the period before November 2000. A detailed analysis shows that the routers could reach all of these addresses only through a small set of links. Specifically, these 120 million addresses were routed through two important transit links, which were not present after November. Most of the addresses were lost from the infrastructure; the rest became reachable through a different set of links; and about 45 million new addresses became visible. The end result was a significant increase in stability.

Infrastructure Health

To evaluate the infrastructure's health as a whole, we focus on global routing robustness in terms of address consistency and route stability. Essentially, we aggregate reachability rank results to evaluate the severity of spatial and temporal variations. Although previous results allowed analysis of temporal and spatial variations from router and address

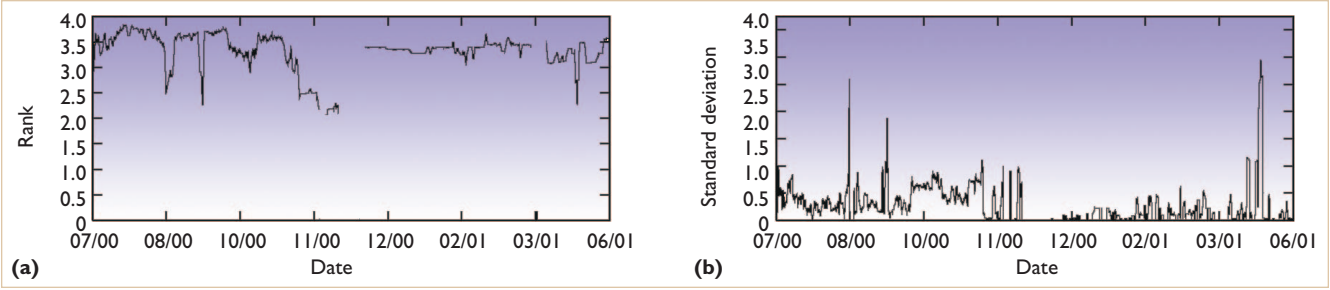

Figure 4. Metrics for determining the infrastructure's health: (a) composite rank and (b) rank deviation. Both measures show a fair amount of stability.

perspectives, the results we present here help us draw conclusions from the infrastructure perspective.

We developed two metrics for evaluating infrastructure health:

- Composite rank averages spatial variations in the infrastructure during a 24-hour sliding window.
- Rank deviation estimates temporal variations within the 24-hour window.

We calculate composite rank as a one-day running average of the values measured every 15 minutes from each of the four routers. Although composite rank normalizes all the high-frequency variations and estimates an address's average reachability within a 24-hour period, rank deviation shows the degree of variation.

Figure 4 presents results for the two metrics. Ideally, composite rank would be 4 and rank deviation would be 0. However, the maximum composite rank for the analysis period is 3.8 (generally, composite rank is approximately 3.5 but is high between July and November 2000). Between July and November 2000, rank deviation is consistently high — mostly between 0.2 and 0.4. Such values imply that the four routers do not consistently see much of the address space over a 24-hour period. This level of instability would seriously interfere with proper multicast group operation.

After November 2000, however, rank deviation decreases considerably and is occasionally zero. Throughout the analysis period, low composite rank and high rank deviation instances also occur. Instances in which composite rank is low and deviation is high are particularly interesting because normally a significant part of the infrastructure is visible (that is, the composite rank is high). A low composite rank combined with high deviation implies that almost none of the infrastructure is stable (and usable). Hence, these periods indicate significant infrastructure-wide problems.

Analyzing the results in Figure 4 requires a better understanding of the nature of temporal variations caused by instabilities because the variations determine how often and for how long an address has a certain rank. We classify temporal variations based on the address-ranking changes they effect:

- frequent but short-lived changes and
- infrequent but long-lived changes.

We use rank deviation to measure these variations. High rank deviation indicates a frequently changing address rank.

Results in Figure 4 confirm two of our earlier conclusions. First, instability decreased after November 2000. Whereas rank deviations were consistently higher than 0.2 between July and November 2000, they were more sporadic and usually low after that.

Second, some spatial variations are consistent and long-lived. We know this because composite rank never approaches the ideal value of 4. Further, the drop in composite rank after November 2000 shows that the proportion of addresses not visible at all four routers increased after that time.

By analyzing trends observed in Figure 4, we can make important conclusions about the state of multicast before and after November 2000. Previously, we saw that 120 million addresses (about 50 percent of the address space) were unstable. A closer look at the data shows that the addresses were usually visible at either all four routers or none of them. Although these addresses were visible in a subset of routers while they transitioned from being visible at all routers to being visible at none, and vice versa, such instances were short-lived.

Composite rank during this period is therefore high because most of the visible addresses have rank 4 — that is, they are considered reachable by all four routers from which data was collected. However, these results do not show the addresses with rank 0

(those not seen by any router). Because these addresses are not seen by any router from which data was collected, they are not considered to be part of the infrastructure (though they might have been in the past and might be again in the future). Although network outages were present after November 2000, they were shorter and less frequent.

Composite rank and rank deviation results also show that during high instability periods, all four routers could reach the addresses for only short intervals. We base this conclusion on the observation that when rank deviation is higher than 0.7, composite rank is low — that is, the visibility of unstable addresses drops significantly during severe instabilities. Further investigation of the cause of these high deviations shows that they result from router-specific instability problems similar to those we discuss in the following sections.

Results in Figure 4 also help us distinguish widespread routing instabilities from those caused by network outages affecting only specific routers. Throughout the analysis period, certain prominent spikes in rank deviation correspond to decreases in composite rank. One such incident occurred in August 2000, in which the spike's magnitude and correspondence to an equivalent decrease in composite rank prove that the problem was specific to the DANTE router or its links. In addition, these results provide important insight about multicast operation during the transition period in November 2000 and show that the transition was not abrupt but rather occurred over approximately three weeks.

Reasons for Instabilities

Our results show that not all addresses are visible at all routers and, more importantly, that instabilities can persist in the infrastructure for a long time. Because extended instability is not due to short-term routing outages, we investigated characteristics of MBGP routing tables in an attempt to identify the correlation between instabilities and MBGP operation. We found that such a correlation indeed exists and classify reasons for instabilities into three categories: MBGP routing table redundancies, protocol bugs, and improper configurations.

MBGP Routing Table Redundancies

Much of the instability prior to November 2000 came from redundancy in the MBGP routing tables, which occurs when there are multiple possible paths to a network. That is, routing tables' entries for the large network and its constituent networks result in an address-space overlap. In MBGP, such redundancies let router administrators advertise preferred paths for smaller networks while providing a common path for the parent network. Because this practice is uncommon in multicast, however, redundant entries are rarely necessary and typically result from erroneous route advertisements.

To estimate the effect of redundancy on stability, we use the results in Figure 5 (next page), which compares the total number of MBGP entries at ORIX (Figure 5a) and the number of entries with at least some overlap (Figure 5b). As the figure shows, the number of redundant entries was consistently high before November 2000, constituting more than 67 percent of the total. After that time, however, fewer than 10 percent of the total entries were redundant. The correspondence between the drop in redundant entries and the drop in instability in November 2000 clearly demonstrates a strong correlation between redundancy and stability. Results for the other three routers show similar correlations.

Our earlier investigation into the cause of stability problems further supports this conclusion. At the beginning of May 2000, we noticed that most redundant routes originated from an autonomous system (AS) belonging to a major ISP. We contacted the network operator responsible for multicast to discuss the problem. Within an hour of appropriate configuration changes at just a few routers, most redundant routes were eliminated. The corrections reduced the MBGP table's size by more than 50 percent and substantially improved stability. Unfortunately, the MBGP configuration interface causing the problems apparently persisted and many other sites made the same configuration mistakes, resulting in the reappearance of redundancies and instabilities in about two months.

Protocol Implementation Bugs

Although it causes instability in MBGP routing tables, developers sometimes consider redundancy to be a useful protocol features. However, redundancies leading to instability are evidence of problems in the protocol's implementation. Further investigation reveals that in certain popular router implementations, redundant entries significantly increase processing requirements when creating multicast distribution trees or advertising routes to peers. Both events occur frequently, creating greater demands on the router's CPU, causing the router to malfunction,[6] and thus increasing instability.

Improper Configurations

Improper or inefficient router configurations are

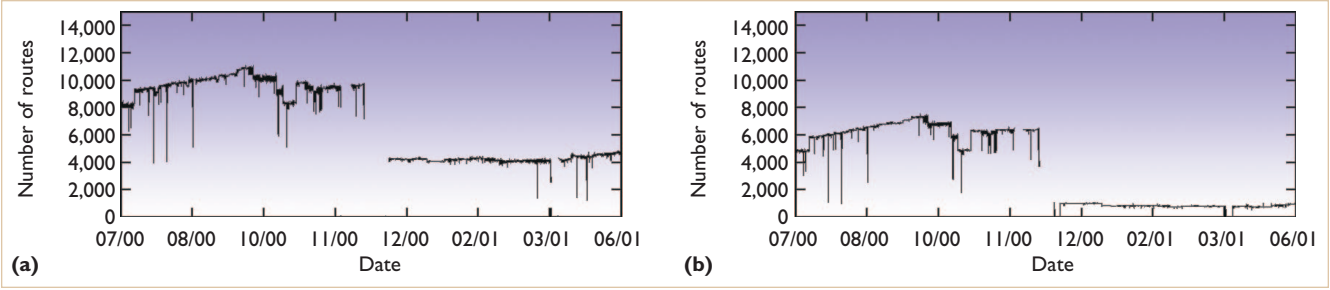

Figure 5. MBGP entries at the ORIX router: A comparison of (a) the total number of entries with (b) the number of redundant entries shows that before November 2000, a large number of routes had some overlap.

the most common cause of spatial variations. Two such configuration instances are particularly interesting because they clearly demonstrate how even simple attempts at configuration and aggregation cause problems.

Configurations related to MBGP route-filtering policy can cause spatial variations. A network administrator can define rules at a router to filter MBGP route categories. Recently, for example, a major ISP configured its routers to filter all routing entries whose network prefix was greater than /22. Because many multicast networks are small (and thus their network prefixes are larger than /22), the routers dropped several legitimate multicast routes. Thus, the configuration was improper because, although route filtering is a valid tool for limiting unicast BGP table size, it is not valid for multicast.

Inconsistent MBGP route aggregation at different levels in the topology can also cause spatial variations. In other words, whereas some routers advertise a single block of MBGP routes, others aggregate routes from multiple small networks and advertise a single MBGP route representing a large network. If the aggregated routes do not represent sets of consecutive IP addresses, however, the address space visible at the routers will differ.

Conclusion

Although some networks remain unstable, educating network administrators about the instability's causes and the use of compatible route advertisement policies will continue to increase multicast stability. With better stability, multicast could become an integral part of the Internet.

References

1. S. Deering and D. Cheriton, "Multicast Routing in Datagram Internetworks and Extended LANs," *ACM Trans. Computer Systems*, May 1990, pp. 85–111.
2. P. Rajvaidya and K. Almeroth, "Analysis of Routing Characteristics in the Multicast Infrastructure," *Proc. IEEE Infocom*, IEEE Press, 2003, pp. 1532-1542.
3. K. Almeroth, "The Evolution of Multicast: From the MBone to Interdomain Multicast to Internet2 Deployment," *IEEE Network*, Jan./Feb. 2000, pp. 10-20.
4. T. Bates et al., "Multiprotocol Extensions for BGP-4," Internet Eng. Task Force, RFC 2283, Feb. 1998; www.ietf.org/rfc/rfc2283.txt.
5. S. Deering et al., "PIM Architecture for Wide-Area Multicast Routing," *IEEE/ACM Trans. Networking*, Apr. 1996, pp. 153-162.
6. C. Labovitz, F. Malan, and G. Jahanian, "Internet Routing Instability," *IEEE/ACM Trans. Networking*, vol. 6, Oct. 1998, pp. 515-528.
7. J. Rexford et al., "BGP Routing Stability of Popular Destinations," *Proc. ACM Internet Measurement Workshop*, ACM Press, 2002.
8. K. Sarac and K. Almeroth, "Monitoring Reachability in the Global Multicast Infrastructure," *Proc. Int'l Conf. Network Protocols* (ICNP), IEEE CS Press, 2000, pp. 141-150.
9. P. Rajvaidya, K. Almeroth, and K. Claffy, "A Scalable Architecture for Monitoring and Visualizing Multicast Statistics," *Proc. IFIP/IEEE Int'l Workshop on Distributed Systems: Operations Management* (DSOM), LNCS 1960, Springer-Verlag, 2000, pp. 1-12.

Prashant Rajvaidya is a PhD candidate in the Department of Computer Science at the University of California, Santa Barbara. His current research interests include network monitoring, the study of multicast topology, analysis of interdomain multicast routing protocols, and multicast security. He has an MS in computer science from UCSB. Contact him at prash@cs.ucsb.edu.

Kevin C. Almeroth is an associate professor at the University of California, Santa Barbara. His main research interests include computer networks and protocols, multicast communication, large-scale multimedia systems, and performance evaluation. He has a PhD in computer science from the Georgia Institute of Technology. He is chair of the Internet2 Working Group on Multicast and is a member of the ACM and the IEEE. Contact him at almeroth@cs.ucsb.edu.