DIMACS Center
Rutgers University

Report on DIMACS Workshop on Multicasting: Architecture, Algorithms, and Applications

Fred Roberts

1 Introduction

The DIMACS Workshop on Multicasting: Architecture, Algorithms, and Applications was held on May 2-4, 2001 at Rutgers University and was attended by 40 people. It was organized by Micah Adler, University of Massachusetts, and D.Frank Hsu, Fordham University.

2 Overview

Multicasting plays an important role in the design, development, operation, and application of next generation networks that rely on the efficient delivery of packets to multiple destinations across a multifaceted and multi-domain network. Due to the advent of broadband, wireless and web-based system design technologies, it has become possible and feasible to design and construct large scale, heterogeneous and complex wireline and wireless communication networks that can support multimedia conferencing, streaming media distribution, distributed data sharing, distance learning, "push" oriented application, and QoS for wired-cable and wired-wireless applications. However, these technology advancements and applications and the convergence of computing, telecommunications and information also open up many challenging problems and issues for both the theory community and practitioners.

The purpose of the workshop was to bring together many of the best researchers and practitioners to present and discuss the recent evolution in the subject area, to interact on emerging issues of common interest, and to set directions and possible standards for future research of and implementation on multicasting networks and their infrastructures.

This workshop covered the following topics:

- IP Multicasting; concept, motivation, standardization, "Host Group" mod el, scoping and forwarding, IGMP for end stations.
- Multicast routing protocols and algorithms; RPM, source-based trees (DVM RP, PIM- dense mode, MOSPF), shared trees (Core-Based Trees(CBT), PIM-sparse mode), Internet multicast routing (Mbone vs. M-GBP or BGMP/GUM?), fast algorithms to compute the multicast tree.
- Multicast transport protocols, RMP and RAMP, interoperability frameworks, expanding-ring searches.
- Multicast Congestion Control; layering schemes such as RLM, multicast at the router and/or switch level.

- Topology inference and network monitoring using multicasting.
- Multicast in wireless systems, mobile computing, ad hoc networks.
- Case Studies and other issues; implementation cases, multicast in the ent erprise, fairness in multicasting, pricing of multicasting.

3 Speakers and Abstracts

Compression Using Efficient Multicasting, Micah Adler, University of Massachusetts

Many multiprocessor systems have the ability to broadcast and/or multicast information efficiently. However, this ability is often overlooked when designing algorithms for these systems. In this talk, we introduce a new compression technique that uses efficient multicasting to significantly reduce the amount of information communicated during parallel and distributed computation, resulting in significantly faster algorithms for Fast Fourier Transforms and sorting on shared memory parallel models with limited bandwidth. These algorithms demonstrate the importance of taking advantage of efficient multicasting. The compression technique uses a new, natural variant of Ramsey theory, which may be of independent interest. (Joint work with Tom Leighton)

Multicast in Optical Networks, J-C. Bermond, University of Nice Sophia-Antipolis

Motivated by wavelength division multiplexing in all-optical networks, we consider the problem of finding a set of paths from a fixed source to a multiset of destinations, which can be coloured by the fewest number of colours so that paths of the same colour do not share an arc. We prove that this minimum number of colours is equal to the maximum number of paths that share one arc, minimized over all sets of paths from the source to the destinations. We do this by modeling the problems as network flows in two different networks and relating the structure of their minimum cuts. The problem can be efficiently solved (paths found and coloured) using network flow techniques. (This lecture covers work by B. Beauquier, P.Hell and S. Perennes)

Fault-Tolerant and Time-Step Optimal Multicasting in Mesh Multicomputers Using Faulty-Block-Information, Dr. Xiao Chen, Southwest Texas State University

We propose a fault-tolerant and time-step optimal multicast algorithm for mesh multicomputers based on the concept of the extended safety level which is a vector associated with each node to capture fault information in the neighborhood. In the approach each destination is reached through a minimum number of hops. In order to minimize the total number of traffic steps, three heuristic strategies are proposed. This approach can be easily implemented by pipelined circuit switching (PCS). A simulation study is conducted to measure the total number of traffic steps under different strategies. Our approach is the first attempt to address the fault-tolerant and time-step optimal multicast problem in mesh muticomputers based on limited global information with a simple model and succinct information.

Supporting Fast Intra-Domain Handoffs and Paging using Multicast in the Dynamic Mobility Agent (DMA) Architecture, Subir Das, Archan Misra, Anthony McAuley and Sajal K. Das (Speaker), University of Texas at Arlington

In this talk we will first propose a new architecture, called Dynamic Mobility Agent (DMA) architecture, for third and fourth generation (3/4G) wireless cellular networks. The objective is to reduce the latency of intra-domain location updates and the mobility related signaling traffic. Then we will discuss the base protocol that provide fast intra-domain handoffs by using a duration-limited, proactive packet 'multicasting' scheme. We will quantify the expected buffering requirement of the proposed multicasting scheme for typical 3/4G network characteristics and compare it with alternative IP-based fast handoff solutions. We will also present a paging scheme under the base protocol that replicates the current cellular paging structure. Our paging mechanism supports generic paging strategies and can significantly reduce the mobility-related IP signaling load.

Our scheme allows the use of either mobile node (MN) initiated or base station (BS) initiated handoff triggers and is applicable to multiple future link-layer technologies. Moreover, the IP-based fast handoff technique provides a secure fast handoff solution that does not assume the existence of specific layer-2 authentication functions or require the adjacent BSs to be aware of each other's identity. Finally, the IP-layer paging mechanism, which allows an idle MN to be located even though it does not perform IP-layer registration/configuration at every change in the subnet (BS), makes the scheme relatively independent of the radio technology.

Reducing the Multicast Routing State, Michalis Faloutsos, UC Riverside

IP multicast suffers from scalability problems for large numbers of multicast groups. In mulitcast routing requires from each router to keep forwarding state for every multicast tree passing through it. Therefore, the state grows with the number of multicast groups. In this work, we propose an approach to reduce multicast forwarding states. In our approach, multiple groups are forced to share a single delivery tree. We discuss the advantages and the imlementations issues of our approach, and conclude that it is feasible and promising. We then propose metrics to quantify state reduction and analyze the bounds on state reduction of our approach. Finally, we use simulations to verify our analytical bounds and quantify the state reduction. In parallel, we study properties of multicast trees and the behavior of the members. The initial simulations suggest that our method can aggregate the routing state significantly.

Issues in Scalable Multicast Protocols, Art Farley, Virginia Lo,(Speaker) Andrzej Proskurowski,(Speaker) Daniel Zappala, University of Oregon

We investigate two related ideas supporting development of scalable multicast protocols. They concern new network substructures (generalizing the standard distribution trees) and multiple "dissemination centers" ("multiple cores"). Our investigations include both theoretical analysis and experimental evaluation.

In particular, we will present some or all of the following (time-permitting):

- Definition, distributed protocols for construction, and experimental evaluation of shared network substructures (spanners and stanners).
- Theoretical analysis of shared multicast trees for fixed set of sources
- Experimental evaluation of multiple core trees and a distributed protocol for core placement ("k-domination")
- Experimental evaluation of SSM proxies, an application-level protocol that implements any-source multicast on top of a source-specific infrastructure.

Computing Connectivity of Cayley Networks, Shuhong Gao, Clemson University

The connectivity of a network measures the number of disjoint paths between any two nodes, or from one node to several other nodes in the network. It also measures the fault tolerance of the network. Cayley networks (or Cayley graphs) are constructed from finite groups. They have many nice properties that make them excellent candidates for interconnection networks for distributed and parallel computer systems. In this talk, we show how to compute the connectivity of any given Cayley network that is connected (i.e. has only one component). Our algorithm depends only on local information, that is, the degree of a node and a small neighborhood of the node within a certain distance. The algorithm is deterministic and has a polynomial time in term of the degree. Note the number of nodes in the network is generally exponential in the degree, so our method is efficient in a strong sense.

Service Location Protocols and Architecture for ad-hoc Networks, Guillermo GuiChal, Georgia Tech

Ad-hoc networks are infrastructureless networks formed on-the-fly by devices with wireless communication capabilities. They have bandwidth and device power limitations, as well as the need to deal with mobility. If client - server interactions are to be present in such networks, a method for performing service location is needed. When two nodes that are several hops apart communicate, resources on all the intermediate nodes are wasted, so it is better to find services that are "closer" to the clients.

Existing service location protocols do not seem suitable for mobile wireless environments. An underlying, low-level architecture that addresses network access issues could enable existing protocols to work in an ad-hoc environment. We have evaluated distributed and centralized architectures for service location and examined service availability and message overhead performance using simulations.

Building Dynamic VPN Multicast Trees Using Active Networks, Yaser Haggag and Sampalli Srinivas, Dalhousie University

Virtual Private Networks (VPN) have emerged as an important solution to security threats (e.g. eavesdropping, data manipulation, repudiation) surrounding the use of public networks for private communications. VPNs provide strong security by integrating a set of security components such as authentication, encryption, access control and session management to proof the identity of both communicating parties, to hide the exchanged data, to only restrict access to legitimate parties and to insure efficient operations of the established secure sessions respectively. Furthermore, VPNs have been proven to be a commercial competitor to the security offered by leased lines hence allowing geographically dispersed parties to communicate economically with each other.

With the increasing popularity of one-to-many and many-to-many applications and services being deployed over public networks, many multicast protocols are being developed to solve issues in group communications. One major issue is scalability, which is a resultant of more members joining the broadcast session thus overwhelming the server causing a decline in server performance, which in turn slows server data delivery. Another issue is security, which is required to keep the broadcast session hidden and unmodified by unauthenticated/unauthorized recipients. One approach to solving the issue of scalability is to move the broadcast session closer to the members by caching the broadcast data onto the nearest routers securely, thus improving scalability and reducing the server loads.

In this paper we introduce a novel approach under development called "deployable virtual private networks" (DVPNs) to help build dynamic (i.e. on demand) secure multicast trees between the broadcast sever and the cached routers. The novelty of the approach comes from the use of active network technology to deploy VPNs. Active network is a new networking technology that inserts intelligence inside the network to offer on demand processing and programming capabilities to network nodes and packets traversing the network. This allows routers to perform computations up to the application layer, thus increasing the flexibility of the network to speed up the deployment of new protocols and applications. Active networks offer many benefit in terms of secure code execution, access control, congestion control and data caching.

Scheduled RMP and a Stock Market Application, Nicholas F. Maxemchuk, Columbia University

The reliable multicast protocol,RMP (a.k.a RBP, a.k.a the Chang/Maxemchuk protocol) was implemented in 1983, and was the first reliable multicast protocol.RMP has several characteristics that made it a likely candidate for the stock market application. In particular:

- 1. the receivers in RMP all place the messages from all of the sources in the same sequence,
- 2. we can eventually guarantee that all of the receivers have a message, and,
- 3. at a later time we can guarantee that all of the receivers know that all of the other receiver have the message.

Using RMP, we created three international stock market systems, with useful fairness properties:

- 1. A unified stock ticker of the transactions that are being conducted on the various physical and electronic exchanges. This system simultaneously delivers the same composite ticker to all receivers, anywhere in the world.
- 2. A unified sequence of buy and sell offers that are delivered to a single exchange or a collection of exchanges. This system gives all of the traders the same fair access to an exchange, independent of their relative distances to the exchange or the delay and loss characteristics of the international network.
- 3. A distributed, electronic trading floor. This system has the fairness attributes of the first two systems and also conducts irrefutable, distributed trades. The original version of RMP implemented a distributed database on a local area network. There were a few tens of sources and receivers and a small number of messages per minute.

The stock market applications have to accommodate tens of millions of sources and receivers, transferring millions of messages per second, internationally, in a hostile environment, here protocol participants cannot be trusted to cooperate. Most of the differences in requirements between the original RMP application and the stock market applications were solved architecturally, by good engineering, rather than by modifying the protocol.During this presentation, I will describe the architecture of the stock market systems and show how it provides the fairness and security that are required. The high message throughput and the short delivery delays in the stock market applications required a modification of the protocol. We solved these problems by changing the protocol from an event driven protocol to a scheduled protocol. The original version of RMP is driven by the arrival of source messages.Each source message is only explicitly acknowledged by one receiver, instead of by each receiver, by using a combination of positive acknowledgements, negative acknowledgements, and token passing.When a receiver accepts the token, it implicitly acknowledges all of the previously, explicitly acknowledged messages.The penalty for reducing the number of acknowledgement messages is the delay until we know that all of the receivers have received an acknowledged a message. When there are N receivers, the delay is the time required to acknowledge the next N-1 source message, and is therefore, a function of both the number of receivers and the arrival times of future source messages. The modified version of RMP sends acknowledgements at scheduled times,rather than being driven by the source messages. In this scheduled protocol if there is no negative acknowledgement by a specified time it is the same as a positive acknowledgment. As a result, we can guarantee that all of the receivers have a source message at a predetermined time after it is first acknowledged, independent of the number of other receivers or the message arrival rate. I am currently determining how tight I can make the interval between the time a message is first transmitted until I can guarantee that all of the receivers have the message.The time depends on the network loss rate and the propagation delay. In certain networks the guarantees are adequate to meet the quality of service required by real time applications.

Hierarchical Path QoS on the QoS-based Multicast Protocol SRSVP, Yasuo Okabe, Takaaki Sekiguchi, Kenji Fujikawa, Kazuo Iwama, Kyoto University

We argue on a method to collect information of each existing multicast flow on hierarchical networks, and describe an algorithm for Qos multicast routing based on it.

In computing a multicast path with per-flow QoS guarantees precisely, it is an important problem how routers collect information of each existing multicast flow. SRSVP, a QoS-based multicast routing protocol, is designed as it collects flow-specific detailed information, called PathQoS, by putting it into signaling messages, so that the derived QoS path becomes more efficient.

In this paper, we have designed an algorithm to calculate PathQoS corresponding to an aggregated link state among areas on a hierarchical network for SRSVP to compute more efficient QoS paths. We have presented that the methods to calculate PathQoS among areas are classified to four cases, have considered dependence upon information needed by the calculations, and have designed an algorithm for routers to be able to process the methods efficiently. We have implemented the mechanism to collect PathQoS based on the algorithm. Now SRSVP is fully adapted to a hierarchical network model.

Bid-based Pricing Approaches for Multicast Sessions in Heterogeneous Networking Environments, Dan Rubenstein, Columbia University, and Micah Adler, University of Massachusetts

One reason that multicast has seen limited deployment is that ISPs are unsure of how they should charge for such a service. A justifiable pricing mechanism must satisfy three requirements. First, it must guarantee the network a profit. Second, the protocols that implement the pricing mechanism should must themselves be simple and efficient. Last, the mechanism must permit admission of sessions that allow for session heterogeneity: receivers within a session should be allowed to receive at differing rates, and it should not be assumed that all nodes within the network will provide multicast (i.e., some form of multicast overlay may be required). In this talk, we demonstrate that it is indeed possible to efficiently implement a bid-based pricing mechanism in a multi-rate overlay multicast that optimizes network profits when a session's transmission is restricted to a single tree. We also explore the complexities that arise when we attempt to extending the session's transmission beyond a single tree. Last, we demonstrate how these bid-based protocols can be used in heterogeneous environments to implement pricing mechanisms that are strategyproof, such that receivers have no incentive to bid less than what receiving the transmission is worth to them.

Distance Learning using IP Multicasting for the Ateneo Educational System, Rafael P. Saldana, Ateneo de Manila University

The Ateneo Educational System consists of schools located in various parts of the Philippine archipelago. Lead by the Ateneo de Manila University, a distance education program is currently being implemented in the educational system. Due to IP multicasting technology, the delivery of high performance computing applications over the the Internet is now a reality. To enhance the delivery of distance learning in the Ateneo educational system we are proposing the utilization of IP multicasting.

How to Spread a Rumor when Not Everyone is Talking to each other OR Multicasting in Heterogeneous Networks, Baruch M. Schieber, IBM Watson Labs

In heterogeneous networks, sending messages may incur different delays on different links, and each node may have a different switching time between messages. The well-studied telephone model is obtained when all link delays and switching times are equal to one unit. We investigate the problem of finding the minimum time required to multicast a message from one source to a subset of the nodes of size k. The problem is NP-hard even in the basic telephone model. We present a polynomial-time algorithm that approximates the minimum multicast time within a factor of O(log k). Our algorithm improves on the best known approximation factor for the telephone model by a factor of O(logn/loglogk), where n is the total number of nodes. No approximation algorithms were known for the general model considered by us. This is joint work with Amotz Bar-Noy, Sudipto Guha and Seffi Naor.

A Distributed QoS Multicast Routing Protocol, Yuval Shavitt, Bell Labs

Many Internet multicast applications such as teleconferencing and remote diagnosis have Quality-of-Service (QoS) requirements. Such requirements can be additive (end-to-end delay), multiplicative (loss rate) or with a bottleneck nature (bandwidth). For these applications, QoS multicast routing protocols are important in enabling new receivers to join a multicast group. However, current routing protocols are either too restrictive in their search for a feasible path between a new receiver and the multicast tree, or burden the network with excessive overhead. In this talk I'll present S-QMRP, a new Stateless QoS Multicast Routing Protocol that supports all three QoS requirement types. S-QMRP is scalable because it has very small communication overhead and requires no state outside the multicast tree; yet, it retains a high success probability. S-QMRP achieves the favorable tradeoff between routing performance and overhead by carefully selecting the network sub-graph in which it conducts the search for a path that can support the QoS requirement, and by auto-tuning the selection according to the current network conditions. S-QMRP does not require any global network state to be maintained and can operate on top of any unicast routing protocol. Our extensive simulation shows that S-QMRP performs better than the previously suggested protocols. Joint work with Shigang Chen

Some Techniques for Multicasting in Hypercubes, Hal Sudborough, University of Texas, Dallas

We consider techniques to minimize message congestion when doing multicasting in a hypercube architecture. That is, for any integer k, we consider one-to-k maps which specify, for each node x in the hyper cube, to which nodes packets are to be sent from x. Our results give techniques to route such packets that satisfy the requests in specified sets of one-to-k maps and that guarantee minimal message congestion in any of the hypercube connections.

The Challenge of Multicasting over Ad Hoc Mobile Wireless Networks, C-K. Toh, Georgia Tech

Multicasting over wireless networks has traditionally been investigated in the view of mobile users moving from the radio coverage of one cell to another. Current multicasting technqiues used in the Internet are robust but not necessarily efficient when applied to wireless networks. Most multicast routing schemes rely on building some forms of multicast trees, be it source- or core-based. In addition, assumptions were made that intermediate routers in the tree paths do not move. However, an emerging new form of totally wireless and pervasive networks known as ad hoc wireless networks have evolved. Such networks has a completely wireless backbone, and use multiple wireless links for end-to-end communications. Multicasting of audio/video/data information in such networks shall present great challenges to researchers due to the presence of mobility, bandwidth, and power constraints. One should question the suitability of multicast trees in such networks and also suggest mechanisms to deal with group membership dynamics. One should also examine scalability issues associated with multicasting in an ad hoc based wireless intranet. This talk shall reveal these problems and encourage thoughts for further investigations.

Deadlock-Free Prefix Multicasting in Irregular Networks, Jie Wu, Fei Dai, Florida Atlantic University, Li Sheng, Drexel University

A deadlock-free multicast scheme called prefix multicasting in irregular networks is studied. Prefix routing is a special type of routing where a compact routing table is associated with each node (processor). Basically, each outgoing channel of a node is assigned a special label and an outgoing channel is selected if its label is a prefix of the label of the destination node. Node and channel labeling in an irregular network are done by using a pre-defined spanning tree which may or may not be minimum. The routing process follows a two-phase process of going up and then down along the spanning tree, with a possible cross channel between two branches of the tree between two phases. It is shown that the proposed routing scheme is deadlock- and livelock-free. The approach is extended to multicasting in which the multicast packet is first forwarded to the longest common prefix (LCP) of destinations in the multicast. The packet is then treated as a multi-head worm that can split at branches of the spanning tree as the packet is forwarded down the tree. Comparison between prefix multicasting and Libeskind-Hadas et al's approach for multicasting is given. Possible extensions are also discussed including reliable prefix routing and multicasting.

A Scalable Topology for High-Speed LAN and WAN Networks, Sotirios G. Ziavras, NJ Institute of Technology

We propose a topology to implement backbone networks in the LAN or WAN environment. This network topology assumes the employment of fiber optic devices. Wavelengths used in this scalable network are reusable and a high degree of reliability is supported. We present efficient algorithms for randomized many-to-all broadcasting and multicasting in the proposed network. These implementations are tested under realistic communication traffic patterns and message generations. A modification of the multicast algorithm applies congestion control to improve the performance.