REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0562

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT ... DATES COVERED
	09/26/2000	Final (March 1996 - March 1998)

4. TITLE AND SUBTITLE
Some Practical Issues in the Design and Implementation of Group Communication Services.

5. FUNDING NUMBERS
F49620-96-1-0103

6. AUTHOR(S)
Shivakant Mishra

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Wyoming
9th & Ivinson Street
Laramie, WY 82071-3355

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/PKC
801 North Randolf Street, Room 732
Arlington, VA 22203-1977.

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

20001030 052

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The main objective of the proposed research was to investigate four important practical issues in the understanding, design, and implementation of group communication services. These issues were (1) **Performance:** maintain good overall performance of group communication services under several different operating conditions, (2) **Semantics:** design a small set of powerful group communication semantics to ensure ease of implementation and good performance of the applications built using these services, (3) **Flow control:** design and evaluate efficient general purpose flow-control techniques that handle temporary buffers in group communication services, and **Real-time responsiveness:** investigate the feasibility of designing and implementing real-time responsive group communication services implemented in distributed systems that are not completely synchronous.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Group Communication, Flow Control, Concurrent Events	10
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

OBJECTIVES

The main objective of the proposed research was to investigate four important practical issues in the understanding, design, and implementation of group communication services. These issues were:

1. **Performance:** maintain good overall performance of group communication services under several different operating conditions.

2. **Semantics:** design a small set of powerful group communication semantics to ensure ease of implementation and good performance of the applications built using these services.

3. **Flow control:** design and evaluate efficient general purpose flow-control techniques that handle temporary buffers in group communication services.

4. **Real-time responsiveness:** investigate the feasibility of designing and implementing real-time responsive group communication services implemented in distributed systems that are not completely synchronous.

STATUS OF EFFORT

All four issues have been investigated as a part of this research and some significant results were obtained. These include

Timewheel Group Communication Service

A new group communication service called the *timewheel group communication service* has been designed and implemented. An extensive performance measurement has shown that this service provides one of the best overall performance under several different operating conditions. This service provides four unique characteristics:

1. this is the first group communication service designed specifically for a *timed* asynchronous distributed system model. This system allows the construction of dependable protocols that specify what outputs and state transitions should occur in response to inputs as well as a real-time interval in which these outputs and state transition will occur.

2. the three protocols that comprise the timewheel group communication service, namely clock synchronization, atomic broadcast, and group membership, are fail aware. This means that they provide mechanisms by which a user is informed of the failures when these protocols may not be able to provide real-time guarantees.

3. the timewheel group communication service supports multiple semantics simultaneously: three order semantics—*unordered*, *total ordered* and *time ordered*, three atomicity semantics—*weak atomicity*, *strong atomicity* and *strict atomicity*, and one termination semantic. These result in a total of nine group communication semantics. A user can dynamically choose one of these nine semantics while broadcasting an update.

4. the strict atomicity together with the time order semantics provided by the timewheel group communication system preserves causal order between updates that may arise due to "hidden" channels. This is a first attempt in preserving causal order arising due to "hidden" channels in an asynchronous system.

Newsmonger

A novel technique called the newsmonger technique has been invented. This technique can be incorporated in most existing atomic broadcast protocols and it results in improving their

performance. It has been shown that this technique improves the group communication stability times by as much as 10 times, particularly when there are communication failures or non-uniform update arrival patterns. In addition, it does not affect any other performance indices of group communication services.

Concurrent Events

An important observation made in this research is that the performance of a group communication service is significantly affected by the implementation technique used for handling concurrent event occurrences. A sequential implementation technique is preferable to a parallel implementation technique for handling concurrent event occurrences in group communication services. A sequential implementation provides better performance, scales better in terms of number of clients or group size, and results in a simpler, more manageable and portable code. The key reason for this result is that although concurrent events occur in group communication services, there is very little parallelism in these services. Servicing of an event interferes extensively with the servicing of other events that occur concurrently. As a result, these events must be serviced in an appropriate sequential order.

Flow Control

Flow-control techniques enable group members to manage their local buffers, which they use to temporarily store multicast updates. Despite buffer overflow being one of the main causes of process failures, flow control had not been studied much in literature. This research investigated different flow-control techniques used in some group communication services and identified two generic flow-control techniques: a conservative and an optimistic technique. All existing flow-control techniques for group communication can be classified as either conservative or optimistic. An extensive performance evaluation under several different operating conditions shows that an optimistic flow-control technique is preferable to a conservative one under non-uniform update arrival patterns. Under the uniform update arrival pattern, a conservative flow-control technique is preferable for smaller group sizes, while an optimistic flow-control technique is preferable for larger group sizes.

ACCOMPLISHMENTS/NEW FINDINGS

The research addressed four important practical issues in the design and implementation of group communication services. This has resulted in five significant findings:

1. It is possible to provide multiple group communication semantics ranging from very weak to very strong in a group communication service and still preserve good overall performance of such a service.

2. Real time guarantees can be provided in a group communication service even when it is implemented in a non-synchronous distributed computing environment.

3. The newsmonger technique is a very useful technique for improving the performance of group communication services.

4. A sequential implementation technique for handling concurrent event occurrences in group communication services is a preferable implementation method compared to a parallel implementation technique.

5. An optimistic flow-control technique is preferable to a conservative flow-control technique for managing temporary buffers in group communication services.

PERSONNEL SUPPORTED

Faculty: Shivakant Mishra

Post-Docs: None

Graduate Students: Dongliang Wang, Ronguang Yang, Lei Wu, Sudha Kuntur, and Guozhao Pang.

Other (Undergraduate Students): Young S. Hyun and Chris Schock

PUBLICATIONS

Journal Publications

Newsmonger: A Technique to Improve the Performance of Atomic Broadcast Protocols (with S. Kuntur). *The Journal of Systems and Software.* To appear in March 2001.

On Group Communication Support in CORBA (with L. Fei, X. Lin and G. Xing). *IEEE Transactions on Parallel and Distributed Systems.* To appear in January 2001.

Low Level Support for Implementing Group Communication Services (with F. Cristian). *ISCA International Journal of Computers and Their Applications.* Vol. 5, No. 4 (Dec 1998).

An Evaluation of Flow Control in Group Communication (with L. Wu). *IEEE/ACM Transactions on Networking.* Vol. 6, No. 5 (Oct 1998).

High Performance Asynchronous Atomic Broadcast (with F. Cristian and G. Alvarez). *Distributed Systems Engineering Journal.* Vol. 4, No. 2 (June 1997), 109–128.

Units of Computation in Fault-Tolerant Distributed Systems (with M. Ahuja). *Journal of Parallel and Distributed Computing,* Vol. 40, No. 2 (February 1997), 194–209.

Refereed Conference Publications

Design and Implementation of an Availability Management Service (with G. Pang). *Proceedings of the ICDCS Workshop on Middleware,* Austin, TX (June 1999).

Improving Performance of Atomic Broadcast Protocols Using the Newsmonger Technique (with S. M. Kuntur). *Proceedings of the Seventh IFIP International Working Conference on Dependable Computing for Critical Applications,* San Jose, CA (January 1999).

Constructing Applications Using the Timewheel Group Communication Service. *Proceedings of the 1998 International Conference on Parallel and Distributed Processing Techniques and Applications,* Las Vegas, NV (July 1998).

The Timewheel Group Membership Protocol (with C. Fetzer and F. Cristian). *Proceedings of the 3rd IEEE Workshop on Fault-tolerant Parallel and Distributed Systems,* Orlando, FL (April 1998).

Thread-based vs Event-based Implementation of a Group Communication Service (with R. Yang). *Proceedings of the 12th IEEE International Parallel Processing Symposium & 9th IEEE Symposium on Parallel and Distributed Processing,* Orlando, FL (April 1998).

Flow Control in High Performance Atomic Multicast Services (with Lei Wu). *Proceedings of the 11th Annual International Symposium on High Performance Computing*, Winnipeg, Canada (July 1997), 295–306.

The Timewheel Asynchronous Atomic Broadcast Protocol (with C. Fetzer and F. Cristian). *Proceedings of the 1997 International Conference on Parallel and Distributed Processing Techniques and Applications*, Las Vegas, NV (June 1997), 1239–1248.

An Experimental Study of Distributed System Behavior. *Proceedings of the 12th ISCA International Conference on Computers and their Applications*, Phoenix, AZ (March 1997), 171–174.

Implementation and Performance of a Stable Storage Service in Unix (with F. Cristian and Y. Hyun). *Proceedings of the 15th IEEE Symposium on Reliable Distributed Systems*, Niagara, Canada (October 1996), 86–95.

Constructing Dependable Distributed Software for Space Applications. *Proceedings of the 1996 Wyoming Space Grant Symposium*, Laramie, WY (October 1996), 9–10.

A Toolkit of Services for Implementing Fault-Tolerant Distributed Protocols. *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications*, Sunnyvale, CA (August 1996) 789–800.

Distributed System Behavior in the Presence of Asynchrony and Failures. *Proceedings of the 11th International Conference on Systems Engineering*, Las Vegas, NV (July 1996) 420–425.

INTERACTIONS/TRANSITIONS

Participation/Presentations At Meetings, Conferences, Seminars, Etc.

1. The 29th IEEE International Symposium on Fault-tolerant Computing, Madison, WI (June 1999).
2. The IEEE International Conference on Distributed Computing Systems, Austin, TX (May 1999).
3. The Seventh IFIP International Working Conference on Dependable Computing for Critical Applications, San Jose, CA (January 1999).
4. The 1998 International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, NV (July 1998), (Session Chair).
5. The 3rd IEEE Workshop on Fault-tolerant Parallel and Distributed Systems, Orlando, FL (April 1998).
6. The 11th Annual International Symposium on High Performance Computing, Winnipeg, Canada (July 1997), (Session Chair).
7. The 1997 International Conference on Parallel and Distributed Processing Techniques and Applications, Las Vegas, NV (June 1997), (Session Chair).
8. The 12th ISCA International Conference on Computers and their Applications, Phoenix, AZ (March 1997), (Program Committee Member).
9. The Wyoming Mathematics, Statistics, and Computer Science Articulation Conference, Sheridan, WY (February 1997), (Panel Chair).
10. The 1996 Wyoming Space Grant Symposium, Laramie, WY (October 1996), (Panel speaker).
11. The 11th International Conference on Systems Engineering, Las Vegas, NV (July 1996), (Session Chair).

Consultative And Advisory Functions To Other Laboratories And Agencies None.

Transitions None.

NEW DISCOVERIES, INVENTIONS, OR PATENT DISCLOSURES

1. It is possible to provide some real-time guarantees in an asynchronous distributed computing environment.

2. An event-based implementation technique is preferable to a thread-based implementation technique for implementing a group communication service.

3. An optimistic flow-control technique is preferable to a conservative one under most operating conditions.

4. The newsmonger technique improves the performance of a group communication service under light load and when communication failures occur.

HONORS/AWARDS

Researcher of Extraordinary Merit Award. University of Wyoming, 1999.