REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		R&D Status Report 10/1/96-12/31/96

4. TITLE AND SUBTITLE
Applications of the Theory of Distributed and Real Time Systems to the Development of Large-Scale Timing Based Systems

5. FUNDING NUMBERS
F19628-95-C-0118

6. AUTHOR(S)
Nancy Lynch

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Airforce
Electronic Systems Center (AFMC)
Hanscom Air Force Base, MA 01731

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
N/A

12a. DISTRIBUTION/AVAILABILITY STATEMENT
No limits on disclosure.

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES

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

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD	- See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE	- See authorities.
NASA	- See Handbook NHB 2200.2.
NTIS	- Leave blank.

Block 12b. Distribution Code.

DOD	- Leave blank.
DOE	- Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA	- Leave blank.
NTIS	- Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

January 14, 1997

Mr. Harry Koch
ESC/ENS
5 Eglin Street, Building 1704
Hanscom Airforce Base, MA 01731-2116

Dear Mr. Koch:

This letter contains our R & D Status Report covering the period from October 1, 1996 to December. 31, 1996 for Contract F19628-95-C-0118, entitled "Applications of the Theory of Distributed and Real-Time Systems to the Development of Large-Scale Timing-Based Systems".

Technical Progress

In the following report, more information about the people mentioned can be found on our group's "people" page, at URL http://theory.lcs.mit.edu/tds/people.html.

I. Modelling and verification tools

- Garland and Lynch now have completed a tentative design of a programming language for I/O automata, which they are now calling "IOA". This language allows simple abstract description of distributed systems, in order to aid in system development, testing and verification. A parser for the language was completed prior to this reporting period. During this quarter, two graduate students, Vaziri and Yang, joined this project. Vaziri has been working on a language manual. Yang has begun working on portions of the static semantic checker related to the I/O automaton model, while Garland has been working on portions involving the Larch theorem prover. Garland and Lynch have begun discussions with others (Vaandrager, Heitmeyer), with the idea of linking our language with various existing verification and simulation tools.

- Jensen and Vaziri have been examining and developing techniques for the integration of model checking and theorem proving methods for verification of concurrent systems. Specifically, they have been studying the feasibility of abstracting from an infinite-state system to a finite-state system. They have developed a property-preserving abstraction theorem for the I/O automaton framework. They are currently examining uses of this theorem in the verification of concurrent read/write and mutual exclusion algorithms.

- Segala, with some help from Lynch, has rewritten the paper "Liveness in Timed and Untimed Systems", by Segala, Gawlick, Sogaard-Andersen and Lynch, for journal submission. This paper presents a compositional treatment of liveness properties for both untimed and timed

distributed systems. During this reporting period, the main proofs were substantially simplified and shortened, using insights derived from Segala and Lynch's recent work on modelling hybrid systems. The paper still needs a small amount of cosmetic work, but is basically done.

- De Prisco developed a new "Clock Timed Automaton" model. It provides a systematic way of describing timing-based systems in which there is a notion of "normal" timing behavior, but that do not necessarily always exhibit this "normal" behavior. This model is intended to be used for stating and proving performance and fault-tolerance properties for practical systems. In particular, it is useful for properties that hold when the system stabilizes to a situation in which timing behavior is normal and no additional failures occur. Examples of this sort of analysis appear in items III.A3 and III.A4 below.

- Hoest and Shavit have developed a novel mathematical model for evaluating the complexity of algorithms in an asynchronous setting. This model, which is a continuation of Shavit's computability modelling work, is based on the use of tools from algebraic topology. They are now writing an article to be submitted to a conference shortly.

II. Algorithms and impossibility results

- In the previous reporting period, Della Libera and Shavit completed their work on reactive diffracting trees, a new version of the diffracting tree synchronization primitive that grows and shrinks according to the load on the data structure. During this quarter, they worked on preparing a conference version of the results.

- Shavit and Zemach continued their work on a highly concurrent priority queue design based on their earlier "combining forest" data structure. They also began work on a linearizable stack structure. They continued to perform empirical evaluations of the designs using the Proteus simulator. Shavit and Zemach also completed their work with Upfal of IBM Almaden on a journal version of their SPAA 96 paper providing a mathematical model for analyzing diffracting tree performance, and submitted it for publication. Finally, Shavit and Zemach completed their design of a new "wait-free" sorting algorithm – one that takes, on average, logarithmic parallel time and runs (though slightly less effectively) even if all processes but one fail.

- Shavit and Shvartsman started working on the journal version of their paper with Lynch and Touitou. The paper deals with linearizability of counting networks and includes an extended empirical analysis. A preliminary version of these results appeared in PODC96.

- Shvartsman revised the manuscript, *A Theory of Fault-Tolerant Parallel Computation.* This monograph synthesizes the latest results for parallel computation in the presence of failures,

restarts and delays. The publication date by Kluwer Academic has been set for February, 1997.

III. Applications

A. Distributed system building blocks

- Shvartsman and Oleg Cheiner, an M.Eng. student, continued experimentation using a prototype distributed algorithm based on the eventually serializable data service of Fekete, Gupta, Luchangco, Lynch and Shvartsman, presented in PODC 96. Early results using a LAN-based message-passing implementation suggest that even when operations are not required to be serialized, the proportion of operations that yield results inconsistent with the eventual order is very small. Cheiner completed and submitted his M.Eng. proposal.

- Lynch and Shvartsman completed their paper defining a new reconfigurable quorum-based broadcast-convergecast communication primitive. They used the primitive to obtain new fault-tolerant distributed implementations of serializable shared read/write memory. The paper was submitted to FTCS 97 under the title "Robust Emulation of Shared Memory Using Dynamic Quorum-Acknowledged Broadcasts".

- Fekete, Lynch and Shvartsman are completing a conference version of their paper on group communication services. They have developed automaton-based specifications for group communication primitives such as those used in the Isis, Transis, Horus and Psynch systems. In particular, they have developed specifications for a virtually synchronous group communication (VSGC) service and for a totally ordered broadcast service. They have modelled an algorithm, derived from one of Dolev and his students, that uses VSGC to implement totally ordered broadcast. They have a good outline of an assertional correctness proof for this algorithm, plus specifications and proofs giving performance and fault-tolerance properties. The performance and fault-tolerance analysis is a "stabilized" analysis of the sort discussed in item I4 above. This work is being prepared for a conference submission.

  Also, working with Khazan, Lynch and Shvartsman have begun modelling a load balancing algorithm that also uses VSGC.

- De Prisco has continued his work on modelling, improving and verifying the Paxos algorithm. Work done this quarter was devoted to finishing the timing model (see item I4 above), and completing the performance analysis. The result of the work will be presented in De Prisco's M.S. thesis which is scheduled to be completed at the end of January.

- Luchangco is continuing research on modelling distributed memory, trying to understand models that have been proposed by both theoreticians (e.g., causal memory, pipelined RAM)

and architects (e.g., Alpha, PowerPC), demonstrate the relationships amongst them, and distill the key ideas. He is also developing a general framework for analyzing precedence-based memory models, which are generalizations of standard multi-processor memory models, and demonstrating how this framework can be used to prove theorems about the various models and their implementations.

B. Transit

- Livadas continued his work on the use of Hybrid I/O Automata to model vehicle protection subsystems, as used in the Raytheon Personal Rapid Transit project. His model allows simple composition of protectors that require correct operation of each other. The correctness proofs are nearly complete for protectors preventing overspeed and collisions both for the case of a single straight track and the case of a track comprised of Y shaped merges and diverges. Livadas hopes to finish his M.S. thesis by the end of January.

- Dolginova and Lynch continued their work on modeling and analyzing safety of the platoon join maneuver for the California PATH intelligent highway project, using Hybrid I/O Automata. Two papers have been written: one, written with Mike Branicky, presented at the Hybrid Systems workshop in Ithaca, Oct. '96 and currently undergoing revision for submission to the symposium proceedings; and one accepted for the HART'97 workshop in Grenoble, France in Mar. '97.

- John Lygeros, who joined the group in November, has begun working with Lynch and Livadas on ways of combining techniques from control theory and discrete system theory to reason about hybrid vehicle control systems.

C. Communication

- Smith and Lynch worked out carefully an impossibility result for the "at-most-once fast delivery problem". This problem is the one that the TCP/IP transport level protocol T/TCP is designed to solve. The impossibility result says that if the client and server do not have accurate clocks, then no protocol can solve this problem. In addition to the proof of impossibility, this work presents an interesting formal model for systems with local clocks that must have liveness requirements. This model is based on the general model for timed systems with liveness requirements of Segala, Gawlick, Sogaard-Andersen and Lynch, described in item I3 above.

- Fekete, Lynch, and Shvartsman have produced formal specifications for several multicast communication service, including a virtually synchronous group communication (VSGC) service

similar to the ones used in systems like Isis, Horus, Transis, Totem, etc. These specifications include performance and fault-tolerance properties as well as "ordinary" correctness properties. An algorithm using VSGC to implement a totally ordered broadcast service has been developed, verified and analyzed. This is discussed in more detail in item III.A3 above.

D. Probabilistic Systems

- Segala and Lynch worked intensively on trying to finish the work of Pogosyants and Segala on modelling, verifying and analyzing the Aspnes-Herlihy randomized consensus protocol. (Pogosyants died in a car crash in Dec. 95, and this was her thesis project at the time.) Besides a proof of this particular algorithm, many useful techniques for analyzing randomized distributed systems have been produced in the course of this work. The full paper is substantially done – it needs only minor cosmetic changes. Segala and Lynch are currently preparing a paper based on this for submission to PODC97.

Special Programs and Major Items of Equipment

None.

Changes in Key Personnel

1. Prof. Alan Fekete returned to Sydney University at the end of December.
2. Dr. Vicente Cholvi-Juan returned to the University of Jaume I in Spain at the end of December.
3. New graduate students Roger Khazan, Henrik Jensen and Gunnar Hoest joined the group.
4. Dr. John Lygeros joined the group as a postdoc, after finishing his PhD in Prof. Sastry's group at Berkeley. He will work on modelling hybrid systems, in particular, transportation control systems.

Trips, Talks and Conferences

1. Nancy Lynch spoke on "Correctness of Vehicle Control Systems: A Case Study", at the 17th IEEE Real-Time Systems Symposium, Washington, D. C. in December, 1996.
2. Steve Garland spoke on "Computer-Assisted Verification of an Algorithm for Concurrent Timestamps" at the FORTE/PSTV'96 Conference, Kaiserslautern, Germany, in October. 1996.

3. Mark Smith spoke on "Formal Verification of Communication Protocols" at the FORTE/PSTV'96 Conference, Kaiserslautern, Germany, in October. 1996.

4. Alan Fekete spoke on "Reasoning about Transaction Management" at Northeastern University, in October.

5. Alan Fekete spoke on "Distributed Commit" at the University of Massachusetts, Boston, in November.

6. Alan Fekete spoke on "Reasoning about System Infrastructure" at Carnegie-Mellon University, in November.

7. Alan Fekete spoke on "Reasoning about System Infrastructure" at Imperial College in London, in November, 1996.

8. Mandana Vaziri attended the autumn school on verification: "Combining model checking and theorem proving", in Aahrus, Denmark, in October-November, 1996.

Areas of Concern

None.

Statement of Sufficiency

The contractually prescribed effort appears to be sufficient to achieve the objectives of this contract.

Degrees awarded

None.

Related Accomplishments

During this reporting period the following papers have been submitted for publication, accepted for publication, or published:

[1] Nancy Lynch and Alex Shvartsman. Robust emulation of shared memory using dynamic quorum-acknowledged broadcasts, 1996. Submitted for publication.

[2] N. Shavit and E. Upfal and A. Zemach. A steady state analysis of diffracting trees. Submitted for journal publication.

[3] Paris C. Kanellakis and Alex A. Shvartsman. *A theory of fault-tolerant parallel computation.* Kluwer Academic Publishers. To appear in February 1997.

[4] Ekaterina Dolginova and Nancy Lynch. Safety verification for automated platoon maneuvers: A case study. *International Workshop on Hybrid and Real-Time Systems (HART'97)*, Grenoble, France, March 1997. To appear.

[5] Michael S. Branicky, Ekaterina Dolginova, and Nancy Lynch. A toolbox for proving and maintaining hybrid specifications. Presented at *HS'96: Hybrid Systems*, October 12-16, 1996, Cornell University, Ithacs, NY. To be published in proceedings.

[6] Tsvetomir P. Petrov, Anna Pogosyants, Stephen J. Garland, and Nancy A. Lynch. Computer-assisted verification of an algorithm for concurrent timestamps. In Reinhard Gotzhein and Jan Bredereke, editors, *Formal Description Techniques IX: Theory, Applications, and Tools* FORTE/PSTV'96: Joint International Conference on Formal Description Techniques for Distributed Systems and Communication Protocols, and Protocol Specification, Testing, and Verification, Kaiserslautern, Germany, October 1996, pages 29–44. Chapman & Hall, 1996.

[7] Nir Shavit and Asaph Zemach. Diffracting trees. *ACM Transactions on Computer Systems.* 14(4):385-428, November 1996.

[8] Mark Smith. Formal verification of communication protocols. In Reinhard Gotzhein and Jan Bredereke, editors, *Formal Description Techniques IX: Theory, Applications, and Tools* FORTE/PSTV'96: Joint International Conference on Formal Description Techniques for Distributed Systems and Communication Protocols, and Protocol Specification, Testing, and Verification, Kaiserslautern, Germany, October 1996, pages 129–144. Chapman & Hall, 1996.

[9] H. B. Weinberg and Nancy Lynch. Correctness of vehicle control systems: A case study. In *17th IEEE Real-Time Systems Symposium*, Washington, D. C., pages 62-72, December 1996.

Papers in progress

Roberto Segala, Rainer Gawlick, Jorgen Sogaard-Andersen, and Nancy Lynch. "Liveness in Timed and Untimed Systems." In progress, for journal submission.

Stephen Garland and Mandana Vaziri. "IOA: A Formal Language for I/O Automata." Draft. In progress.

Giovanni Della Libera and Nir Shavit. "Reactive Diffracting Trees." In progress.

Victor Luchangco. "Precedence Based Memory Models." In progress.

Gunnar Hoest and Nir Shavit. "Towards a Topological Characterization of Asynchronous Complexity." In progress.

Shavit and Shvartsman. Journal version of "Counting Networks are Practically Linearizable" coauthored with Dan Touitou and Nancy Lynch. In progress.

Alan Fekete, Nancy Lynch and Alex Shvartsman. "Specifying and Using a Partitionable Group Communication Service. In progress.

Roberto Segala and Nancy Lynch. (Formerly by Roberto Segala and Anna Pogosyants.) "Verification of the Randomized Consensus Algorithm of Aspnes and Herlihy: a Case Study." In progress.

Roberto DePrisco. "Revisiting the Paxos algorithm." Masters thesis in progress.

Carolos Livadas. "Verification of Automated Vehicle Protection Systems." Masters thesis in progress.

Victor Luchangco. "Building Blocks for Distributed Computing Applications." PhD thesis in progress.

Mark Smith. "Formal Verification of Communication Protocols for Data Streaming and Transactions." PhD thesis in progress.

Oleg Cheiner. "Implementation and Evaluation of an Eventually-Serializable Data Service." Masters thesis in progress.

Awards:

- Nancy Lynch was elected as a Fellow of the Association for Computing Machinery.
- Nancy Lynch was awarded the NEC Chair of Software Science and Engineering
- Mandana Vaziri was selected to attend the BRICS Autumn School on Verification in Denmark.
- Oleg Cheiner received an NSF fellowship.

Sincerely,

Nancy Lynch (jr)

Nancy Lynch
NEC Professor of Software Science and Engineering

Electrical Engineering and Computer Science
(617)253-7225
lynch@theory.lcs.mit.edu

MIT Laboratory for Computer Science

Applications of the Theory of Distributed Real-Time Systems
To the Development of Large-Scale Timing-Based Systems

Prof. Nancy Lynch, Principal Investigator

R & D Status Report
Program Financial Status
ARPA Contract # F19628-95-C-0118
CLIN # 0002
1996 Fourth Quarter (10/96 - 12/96)

	Planned Expenditures	Actual Expenditures at Report Date	% Completion	Budget At Completion	Latest Revised Estimate	Remarks
Total Base Contract	858,443	257,252	29.97%	858,443	858,443	
Current Funding Profile	363,787	257,252	70.72%		257,252	*
Equipment	35,308	0	0.00%			**

* Data reflects all received funding. Current funding is sufficient for this fiscal year.
Next fiscal year's anticipated funding requirements are $190,154

** Equipment funding is for 3 budgeted workstations. None have been purchased as yet.