

April 15, 1998

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 January 1, 1998 to March 31, 1998 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

The group presently consists of Prof. Lynch and graduate students Victor Luchangco, Roberto DePrisco, Mandana Vaziri, Henrik Jensen, Josh Tauber, Roger Khazan, Carl Livadas, and Kate Dolginova. Dr. Stephen Garland and graduate student Anna Chefter of Guttag's group are also working closely with us.

Information about these people can be found at URL <http://theory.lcs.mit.edu/tds/people.html>.

I. Funding issues

A large part of Prof. Lynch's time in March was devoted to attempting to find stop-gap funding for the research group for next year, since our group's coordinated DARPA proposal for 1998-2001 was turned down. Other time in March was taken in beginning the planning for the extensive work of writing a set of three separate proposals this summer, as suggested by DARPA program managers.

II. Modelling and verification tools

We continued our project on the IOA language and toolset, which are designed to support our "abstract distributed programming" formal approach to distributed system design and analysis. The design of the IOA language is substantially complete, and appears in a language manual on the web.

This quarter, Dr. Garland spent two weeks at the Laboratoire Spécification et Vérification of the Ecole Normale Supérieure in Cachan, France, at the invitation of Professor Michel Bidoit, the director of the laboratory. While there he gave a talk about the IOA project, and he worked with Professor Bidoit to explore the feasibility of formalizing the semantics of the IOA language in CASL (a new Common Algebraic Specification Language, which extends Larch and is being developed by a consortium of European universities). In the long term, CASL may be an attractive alternative to Larch because of its treatment of subsorts, partial functions, and parameterized datatypes. However, CASL currently has no tool support comparable to that available for Larch (i.e., to the

Larch Prover and its various front-ends). Hence, in the short term, Larch still provides the most satisfactory underpinning for IOA.

This quarter, work also continued on the development of tools for the IOA language; our toolset will include a parser and static semantic checker, composition routine, support for levels of abstraction, interfaces with theorem provers and model checkers, a simulator, and a code generator for real distributed code.

- Vaziri continued her work on translating IOA to PROMELA, the input language of the SPIN model-checker. She had previously devised a translation scheme for a subclass of IOA programs. During this period, she expanded the subclass by providing PROMELA implementations for data structures such as Set and Sequence. She also worked on examples with the objective of optimizing the performance of programs obtained from translation, by refining the translation scheme and providing the user with a set of guidelines for making efficient use of the model-checker.

Vaziri started with the Peterson's mutual exclusion algorithm as written in Lynch's Distributed Algorithms book. Some non-determinism was taken out of the original IOA program to make the PROMELA program work for 3 processes. Subsequent examples considered were Dijkstra's mutual exclusion algorithm and the Randomized Dining Philosophers, also from Lynch's book. Both algorithms worked for 3 processes as well. The main guidelines derived for IOA users are: (i) unnecessary non-determinism should be avoided as much as possible, (ii) the environment should be as restrictive as possible, (iii) complicated data structures should be avoided. These examples also helped in refining the translation scheme. Vaziri is currently working on a Replicated State Machine example that uses Lamport's logical time.

The implementation of the translator is pending the completion and finalization of an intermediate language being currently designed and implemented for the IOA system.

- We continued work on the ambitious project of translating distributed IOA programs to running Java or C++ code. This quarter, Tauber extended the interface between local IOA programs and the Message Passing Interface (MPI) service to be used in the distributed implementation. He proved that a set of I/O automata that formally describe that interface and the behavior of an implemented MPI system exhibit the desired reliable communication characteristics. Tauber also extended his translation scheme for a subclass of IOA programs so that it can resolve issues of scheduling nondeterminism. He has begun building a library to emit Java code that implements IOA data types and message formats for use in the compiler back-end.
- Chetter wrote her thesis proposal where she documented the design of the IOA simulator. She designed the IOA intermediate language and implemented a parser for it, developed and

implemented a configuration specification language for the simulator, and ran simulations of simple I/O automata.

III. Applications

A. Distributed system building blocks

We continued our work on building-blocks for fault-tolerant distributed systems. Much of our progress on this topic this quarter involved dynamic view-oriented group-communication services.

- DePrisco, Fekete, Lynch, and Shvartsman worked intensively on models and proofs for dynamic view-oriented group communication services. Such services are useful for fault-tolerant distributed computing requiring coherent data, in systems where processes can join and leave routinely.

We have provided a formal specification, DVS, of a dynamic quorum service. To show the usefulness of our specification, we have provided both an implementation of the service, based on a design of Lotem, Keidar and Dolev, and a totally-ordered-broadcast application that runs on top of the service. This quarter, we completed correctness proofs for both algorithms, and produced a conference version for the 1998 Principles of Distributed Computing conference. Our proofs uncovered many subtle aspects of the implementation.

The work in the PODC 98 paper deals with safety properties of the service only. This quarter, Lynch and Fekete also began work on formulating performance and fault-tolerance guarantees; we have a tentative sketch for some properties of this kind. Lynch and Fekete also outlined a generalization of dynamic views to dynamic *configurations*, where a configuration allows alternative sets of processors (quorums) to be accessed. This generalization is intended to allow the service to tolerate transient failures within a particular view. We outlined how such a service can be used for implementing coherent replicated data, using a strategy based on one by Attiya, Bar-Noy and Dolev; this usage requires the dynamic configuration service to support a combined broadcast-convergecast communication discipline, similar to one developed by Lynch and Shvartsman for FTCS 97. The dynamic configuration idea also appears capable of supporting a dynamic version of the Liskov-Oki primary copy replicated data strategy. Working all this out carefully remains for our future research.

- Lynch began working with van Renesse and Hickey at Cornell to provide accurate I/O automaton models for actual layers in Birman and van Renesse's Ensemble system (for view-oriented group communication). These models include global specifications for the services provided as well as detailed specifications for the distributed programs. This work complements (and connects formally to) current work in Constable's NuPRL project on verifying local properties of the low-level ML code that implements the system.

- Khazan continued his work on modeling a load-balancing replicated data server. His implementation relies on the underlying group-communication service to achieve fault-tolerance and efficiency. During this reporting period, Khazan made a preliminary analysis of the performance and fault-tolerance properties of the replicated service, and in particular, the load-balancing part of it. Based on this analysis, Khazan outlined a number of factors that make multicast group communication services potentially suitable for load-balancing, and started preliminary work on modeling a generalized load-balancer on top of a group communication service. In addition, Khazan has been writing up his Masters thesis, and has prepared a conference paper to be submitted to the 12th International Symposium on Distributed Computing (DISC 98).
- Luchangco and Lynch completed an extensive revision of the journal version of their paper with Fekete, Gupta and Shvartsman on eventually-serializable data services (ESDS), and submitted it for publication. ESDS defines a specification for replicated data services that trades off immediate consistency guarantees for improved system performance and availability, while ensuring the long-term consistency of the data. This paper also gives an algorithm that implements ESDS based on a lazy replication strategy. This version included several new results including:
 - Explicit guarantees on the behaviors allowed by ESDS that should be helpful when ESDS is used as a building block in the design of applications.
 - Guarantees on the performance of the algorithm presented in the paper to implement ESDS.
 - A consideration of the fault-tolerance of the algorithm, under timing failures, message loss or duplication, and process crashes and restarts.
 - A more formal treatment of the optimizations of the algorithm.

The specification and some notation was also changed to be more understandable, and many of the proofs were reworked considerably to be simpler and more complete than before.

B. Multiprocessor shared memory models

- Frigo and Luchangco's paper on computation-centric memory models was submitted and accepted to the ACM Symposium on Parallel Algorithms and Architectures to be held in June/July 1998. Computation-centric models characterize multi-processor shared memories from the point of view of the programmer, and are particularly well-suited to programming in a more general framework which allows the programmer to ignore details such as which processor may actually run a particular piece of code. A computation is a generalization of

parallel instruction streams. Memory models are expressed in terms of these computations, allowing the programmer to reason about what a program specifies rather than low-level system details. The paper defines sequential consistency in this framework, along with several weak consistency models, and show some characteristics of these models, as well as relationships among them. They also define properties that characterize "reasonable" memory models, i.e., they argue that memory models not having these properties are undesirable from the programmer's point of view. On this basis, they suggest a candidate for the weakest "reasonable" memory model.

- The paper by Vaziri, Lynch "Proving Correctness of a Controller Algorithm for the RAID Level 5 System," was accepted at FTCS-28.

C. Automated Transportation Systems

We continued our analysis work on case studies for intelligent highway system maneuvers and aircraft control maneuvers. We also moved very near to finishing our comprehensive journal paper on the basic Hybrid Input/Output Automaton model for hybrid systems.

- Dolginova continued her thesis work on modeling and analyzing safety criteria of the platoon maneuvers for the California PATH intelligent highway project, using Hybrid I/O Automata. Prior to the reporting period, the ideal case, and some of the more complicated cases involving delays and some forms of uncertainty were modeled and verified. This quarter, the outbound uncertainty, and variation of vehicle parameters were modeled and verified. Additionally, the original paper is being revised and extended to be Dolginova's Master's Thesis.
- Lygeros, Lynch and Livadas have continued their work on the modelling, proof, and analysis of an abstract version of the Traffic Alert and Collision Avoidance System II (TCASII). This quarter, we worked on a technical report version of the paper, with detailed proofs. This is still in progress.
- Lynch, Segala and Vaandrager made another pass over our journal paper on the basic HIOA (hybrid input/output automaton) model. This quarter's work involved general improvements to the mathematics and presentation, especially in the most difficult proofs. This should be ready for journal submission within a few weeks.
- Livadas began studying the literature on several of the automated tools being developed in academia for analysis of hybrid systems. The goal of this work is, first, to study the approach used by other people to analyze hybrid systems, and, second, to determine the usefulness of the functionality of the currently available tools. This literature research is the preliminary stage of the design of automated analysis tools for hybrid systems specified as

hybrid I/O automata. At this point in time, the HIOA analysis tools are envisioned as being the extension of the IOA language and toolset described earlier in this report, extended to allow the specification and analysis of the behavior of continuous variables.

IV. Algorithms and impossibility results

- Jensen has continued his work on using property preserving abstractions to reduce large, possible infinite state, verification problems to smaller, finite state, problems. He has most recently used his abstraction method to prove a fundamental property of the Bounded Concurrent Timestamp Algorithm of Dolev and Shavit. A paper with work title "An Abstract Interpretation of the Bounded Concurrent Timestamp Algorithm" is in progress.

An earlier case-study utilizing the abstraction method has been reported on in paper: "A Proof of Burns N-Process Mutual Exclusion Algorithm using Abstraction", by Jensen and Lynch. As reported in the last quarterly report, this paper has been accepted for the TACAS'98 (Tools and Algorithms for the Construction and Analysis of Systems) conference, held at the Gulbenkian Foundation, Lisbon, Portugal, March 31 - April 3. Jensen attended the TACAS'98 conference and presented the accepted paper.

Special Programs and Major Items of Equipment

None.

Changes in Key Personnel

Plans have been made for Dr. Idit Keidar of the Hebrew University to join the group in summer, 1998, as a postdoctoral research associate.

Prof. Lynch is on sabbatical this Spring, 1998, travelling sporadically. She is continuing to supervise the project closely.

We have initiated new collaborations with several systems groups: Birman and Constable's groups at Cornell, the Isis group at Stratus computer corp., and Malki's group at AT&T.

Trips, Talks and Conferences

1. Nancy Lynch. Sabbatical trip to University of Sydney January 20, 1998-February 10, 1998. Lynch worked with Fekete and his graduate student on several projects involving dynamic group communication services.
2. Nancy Lynch. "The IOA Language and Toolset: Support for Mathematics-Based Distributed Programming." Dartmouth College, Hanover, NH, February 23, 1998.

3. Nancy Lynch. "The IOA Language and Toolset: Support for Mathematics-Based Distributed Programming." Cornell University, Ithaca, NY, Mar. 2, 1998.
4. Nancy Lynch. Sabbatical trip to Cornell University February 24, 1998-March 6, 1998. Lynch worked with Birman, van Renesse, Constable, Kleinberg, and graduate students on issues mainly related to group communication services and mathematical and tool support for abstract programming.
5. Nancy Lynch. "View-Oriented Group Communication Services" Cornell University, Ithaca, NY, Mar. 3, 1998.
6. Nancy Lynch. "Mathematical Models and Proof/Analysis Methods for Timing-Based Systems and Their Application to Communication, Fault-Tolerant Distributed Computing, and Hybrid Systems." DARPA Networking PI Meeting, Tucson, Arizona, March 11-13, 1998.-
7. Henrik Ejersbo Jensen presented paper "A Proof of Burns N-Process Mutual Exclusion Algorithm using Abstraction" by H.E. Jensen and Nancy Lynch, at the 4th International Conference, TACAS'98 (Tools and Algorithms for the Construction and Analysis of Systems). Held as part of the Joint European Conferences on Theory and Practice of Software, ETAPS'98, Lisbon, Portugal, March/April 1998.
8. Ekaterina Dolginova. SIGCSE (Special Interest Group in Computer Science Education) Conference in Atlanta, GA, 2/26-2/28. Receive CRA's Outstanding Undergraduate Award (see below under awards).
9. Stephen Garland, "A language and set of tools for the design, analysis, and construction of distributed systems," Laboratoire Spécification et Vérification, Ecole Normale Supérieure de Cachan, France, March 5, 1998.
10. Chefter gave a talk in the SDS (Professor Guttag's group) group meeting.

Areas of Concern

Since our DARPA proposal for 1998-2001 was turned down, we are quite concerned about how to continue the project. We are attempting to secure stop-gap funding for 1998-1999, while writing three new proposals for the subsequent three years.

Statement of Sufficiency

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

Degrees Awarded

Related Accomplishments

During this reporting period the following papers were submitted for publication, accepted for publication, or published. Additionally, a list of papers that are in progress follows.

Submitted for publication:

- [1] Elizabeth Borowsky, Eli Gafni, Nancy Lynch, and Sergio Rajsbaum. The BG Distributed Simulation Algorithm. Submitted for journal publication, December, 1997.
- [2] Stephen J. Garland and Nancy A. Lynch. The IOA Language and Toolset: Support for Mathematics-Based Distributed Programming. Submitted to FORTE/PSTV '98.
- [3] John Lygeros, Carl Livadas and Nancy Lynch. A New Method for Modeling, Verifying and Analyzing Safety-Critical Air Traffic Management Software. Submitted to the Second international Air Traffic Management R&D seminar, ATM-98.
- [4] Alan Fekete, David Gupta, Victor Luchangco, Nancy Lynch, and Alex Shvartsman. Eventually-Serializable Data Services. Journal version. Submitted for publication.

Accepted:

- [5] Alex Shvartsman and Oleg Cheiner. Implementing and Evaluating an Eventually-Serializable Data Service. A short version of the paper was accepted by at PODC'98 in the brief presentations track. A longer version of the paper was invited and accepted for publication in an upcoming DIMACS volume.
- [6] Mandana Vaziri and Nancy Lynch. Proving Correctness of a Controller Algorithm for the RAID Level 5 System. *28th International Symposium on Fault-Tolerant Computing Systems*, Munich, Germany, June 1998. To appear.
- [7] Roberto De Prisco, Alan Fekete, Nancy Lynch and Alex Shvartsman. A Dynamic View-Oriented Group Communication Service. *Proceedings of the 17th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing*, Puerto Vallarta, Mexico, June-July, 1998. To appear.
- [8] Henrik Ejersbo Jensen and Nancy Lynch. A Proof of Burns N-Process Mutual Exclusion Algorithm using Abstraction. *4th International Conference, TACAS'98 (Tools and Algorithms for the Construction and Analysis of Systems)*. Part of the Joint European Conferences on Theory and Practice of Software, ETAPS'98, Lisbon, Portugal, March/April 1998. To appear.
- [9] Matteo Frigo and Victor Luchangco. Computation-Centric Memory Models. *Proceedings of the 17th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing*, Puerto Vallarta, Mexico, June-July, 1998. To appear.

- [10] John Lygeros and Nancy Lynch. Conditions for Safe Deceleration of Strings of Vehicles. *Hybrid Systems: Computation and Control*, Berkeley, California, April, 1998. To appear.
- [11] Carolos Livadas and Nancy A. Lynch. Formal Framework for Modeling and Verifying Safety-Critical Hybrid Systems. *Hybrid Systems Computation and Control*, Berkeley, California, April, 1998. To appear.
- [12] Anna Pogoyants and Roberto Segala and Nancy Lynch. Verification of the Randomized Consensus Algorithm of Aspnes and Herlihy: a Case Study. *Distributed Computing*. To appear.
- [13] Mandana Vaziri and Nancy Lynch. Proving Correctness of a Controller Algorithm for the RAID Level 5 System. Technical Memo MIT/LCS/TM-, Laboratory for Computer Science, Massachusetts Institution of Technology, Cambridge, MA, December, 1997. To appear.

Published

- [14] Alan Fekete, M. Frans Kaashoek, and Nancy Lynch. Implementing sequentially consistent shared objects using broadcast and point-to-point communication. *Journal of the ACM*, 45(1):35-69, January 1998.

Papers in progress

- [15] Nancy Lynch, Roberto Segala, Frits Vaandrager, and H. B. Weinberg. Hybrid I/O Automata." Journal version. In progress.
- [16] Nancy Lynch and Alex Shvartsman. Robust Emulation of Shared Memory Using Dynamic Quorum-acknowledged broadcasts, Journal version. In progress.
- [17] Alan Fekete, David Gupta, Victor Luchangco, Nancy Lynch, and Alex Shvartsman. Eventually-Serializable Data Services. Journal version. Submitted for publication.
- [18] John Lygeros and Nancy Lynch. On the Formal Verification of the TCAS Conflict Resolution Algorithms. Technical Report in progress.
- [19] Henrik Jensen. An Abstract Interpretation of the Bounded Concurrent Timestamp Algorithm. In progress.
- [20] Mandana Vaziri and Nancy Lynch. Translating IOA to Promela. In progress.
- [21] , Roberto De Prisco, Alan Fekete, Nancy Lynch and Alex Shvartsman. A Dynamic View-Oriented Group Communication Service. Technical report in progress.

Theses in progress

Roberto Deprisco. PhD thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Kate Dolginova. MEng thesis. "Safety Verification of Automated Car Maneuvers." Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Henrik Jensen. PhD Thesis. "Integration of Deductive and Algorithmic Methods for Verification of Reactive Systems." Aalborg University, Denmark. Visiting MIT.

Roger Khazan. Masters Thesis. "Group Communication as a Base for a Load-Balancing, Replicated Data Service." Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Carl Livadas. PhD thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Victor Luchangco. PhD Thesis. "Consistency Models for Distributed Memories." Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

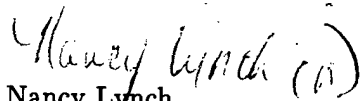
Josh Tauber. PhD Thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Mandana Vaziri. PhD thesis (Untitled). Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

Awards:

Ekaterina Dolginova. Winner of CRA's (Computer Research Association) Outstanding Undergraduate Award. The award included a plaque and a \$1,000 scholarship. NOTE: this might have been mentioned in a previous report, but she actually received the award only now.

Sincerely,



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

Quarterly Report (1/98 - 3/98)

Total Base Contract

Current Funding Profile

Equipment

Planned Expenditures	Actual Expenditures at Report Date	% Completion	Budget At Completion	Latest Revised Estimate	Remarks
858,443	573,094	66.76%	858,443	858,443	
858,443	573,094	66.76%		858,443	*
35,308	23,554	66.71%			

* Data reflects all received funding through 3/98. Current funding is sufficient through 8/98.