EXPERIMENTS IN AUTOMATED ANALYSIS OF CONCURRENT SOFTWARE SYSTEMS

George S. Avrunin[*,1]
Laura K. Dillon[†,2]
Jack C. Wileden[‡,1,3]

COINS Technical Report 89–49
April 1989

[*]Department of Mathematics and Statistics
University of Massachusetts
Amherst, Massachusetts 01003

[†]Computer Science Department
University of California
Santa Barbara, California 93106

[‡]*Software Development Laboratory*
Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

This work was supported in part by the following grants:
[1]National Science Foundation grant CCR-8806970 and ONR grant N00014-89-J-1064
[2]National Science Foundation grant CCR-8702905
[3]National Science Foundation grant CCR-87-04478 with cooperation from the Defense Advanced Research Projects Agency (ARPA Order No.6104).

Experiments in Automated Analysis of Concurrent Software Systems

George S. Avrunin[1]
Department of Mathematics and Statistics
University of Massachusetts, Amherst

Laura K. Dillon[2]
Department of Computer Science
University of California, Santa Barbara

Jack C. Wileden[1,3]
Department of Computer and Information Science
University of Massachusetts, Amherst

Abstract

It is unlikely that any single approach to analysis of concurrent software systems will meet all the needs of software developers throughout the development process. Thus, experimental evaluation of different concurrent analysis techniques is needed to determine their relative strengths and practical limitations. Such evaluation requires automated tools implementing the analysis techniques.

This paper describes a prototype toolset automating the constrained expression approach to the analysis of concurrent software systems. The results of preliminary experiments with the toolset are reported and the implications of these experiments are discussed.

[1]Partially supported by NSF grant CCR-8806970 and ONR grant N00014-89-J-1064
[2]Partially supported by NSF grant CCR-8702905
[3]Partially supported by NSF grant CCR-8704478 with cooperation from DARPA (ARPA order 6104).

1 Introduction

A wide variety of techniques have been proposed for analyzing the behavior of concurrent software systems. These differ in their underlying models of concurrent computation, in the questions about behavior they attempt to answer, and in the stages of the software development process in which they are applied. It is unlikely that any single approach to analysis can possibly meet all the needs of software developers throughout the development process.

The effective use of analysis techniques during software development requires an understanding of their relative strengths and practical limitations. While virtually all existing analysis techniqes are known to have limitations of various kinds, little is known about the practical significance of these limitations. This determination can only be made through experimental application of the techniques to a wide range of concurrent systems. Clearly, experiments must be conducted with systems of realistic size and complexity, and so automated tools implementing the analysis techniques will be required. In this paper, we report on a prototype toolset supporting the *constrained expression* approach to analysis, and the results of some preliminary experiments with that toolset.

The next section of the paper briefly describes the constrained expression approach. The third section describes the toolset, and the fourth reports some of our experience in using the toolset. Finally, we discuss the implications of these experiments for further work on constrained expressions.

2 Constrained Expressions

In the constrained expression approach to analysis of concurrent systems, the system descriptions produced during software development (e.g., designs in some design notation) are translated into formal representations, called *constrained expression representations*, to which a variety of analysis methods are then applied. This approach allows developers to work in the design notations and implementation languages most appropriate to their tasks. Rigorous analysis is based on the constrained expression representations that are mechanically generated from the system descriptions created by software developers.

This section contains a brief overview of the constrained expression formalism. A detailed and rigorous presentation is given in [10], and a less formal treatment presenting the motivation for many of the features of the formalism appears in [5]. The use of constrained expressions with a variety of development notations is illustrated in [5] and [12].

The constrained expression formalism treats the behaviors of a concurrent system as sequences of events. These events can be of arbitrary complexity, depending on the system characteristics of interest and the level of system description under consideration. Associating an *event symbol* to each event, we can regard each possible behavior of the system as a string over the alphabet of event symbols.

We use interleaving to represent concurrency. Thus, a string representing a possible behavior of a system that consists of several concurrently executing components is obtained by interleaving strings representing the behaviors of the components. The events themselves are assumed to be atomic and indivisible. "Events" that are to be explicitly regarded as overlapping in time are represented by treating their initiation and termination as distinct atomic events.

The set of strings representing behaviors of a particular concurrent system is obtained by a two-step process. First, a regular expression, called the *system expression*, is derived from a description of the system in some notation such as a design or programming language. The language of the system expression includes strings representing all possible behaviors of the system. It may, however, also include strings that do not represent possible behaviors, as the system expression does not encode the full semantics of the system description. This language is then "filtered" to remove such strings, using other expressions, called *constraints*, which are also derived from the original system description. A string survives this filtering process if its projections on the alphabets of the constraints lie in the languages of the constraints. The constraints (which need not be regular) enforce those aspects of the semantics of the design or programming language, such as the appropriate synchronization of rendezvous between different tasks or the consistent use of data, that are not captured in the system expression. The reasons for this two-step process, which might not seem as straightforward as generating behaviors directly from a single expression, are discussed in [12].

Our main constrained expression analysis techniques require that questions about the behavior of a concurrent system be formulated in terms of whether a particular event symbol, or pattern of event symbols, occurs in a string representing a possible behavior of the system. For example, questions about whether the system can deadlock might be phrased in terms of the occurrence of symbols representing the starvation of component processes of the system.

Starting from the assumption that the specified symbol, or pattern of symbols, does occur in such a string, we use the form of the system expression and the constraints to generate inequalities involving the numbers of occurrences of various event symbols in segments of the string. If the system of inequalities thus generated is inconsistent, the original assumption is incorrect and the specified symbol or pattern of symbols does not occur in a string corresponding to a behavior of the system. If the inequalities are consistent, we use them in attempting to construct a string containing the specified pattern.

Constrained expression analysis, then, is a static, event-based approach (though the construction of a behavior from a solution of a system of inequalities has similarities to dynamic analysis). The constrained expression formalism is closely related to path expressions [7], event expressions [19], and COSY [18]. More detailed discussion of the relation between constrained expressions and a variety of methods for describing and analyzing concurrent software systems can be found in [5] and [22]. The constrained expression analysis techniques can be regarded as rigorous formulations of methods based on arguments about the order and number of occurrences of events. Such methods have been widely used in conjunction with concurrent software systems (e.g., [15]).

In summary, the constrained expression approach is applicable to systems expressed in a variety of notations and languages. It offers a focused approach to analysis, which, by keeping the amount of uninteresting information produced to a minimum, can be very efficient. One potential difficulty in applying the approach is that it requires that analysts correctly formulate questions about the behavior of a system in terms of patterns of event symbols in strings representing system behaviors. Other potential drawbacks include the difficulty of automating some aspects of generating and reasoning about the systems of inequalities.

After manually applying the constrained expression analysis techniques to a number of small examples with encouraging results (e.g., [2], [5], [6], [22]), we began to construct prototype tools automating various aspects of the analysis. An important goal of this automation effort is to support experimentation directed at determining the practical significance of the potential problems cited above. This paper describes the first complete version of the prototype toolset and reports the results of some experiments with it.

3 The Constrained Expression Tools

The prototype toolset (see Figure 1) consists of five major components: a ***deriver*** that produces constrained expression representations from concurrent system designs in a particular design language; a ***constraint eliminator*** that replaces a constrained expression with an equivalent one involving fewer constraints; an ***inequality generator*** that generates a system of inequalities from the constrained expression representation of a concurrent system; an ***integer programming package*** for determining whether this system of inequalities is consistent or inconsistent, and, if the system is consistent, for finding a solution with appropriate properties; and a ***behavior generator*** that uses the solution found by the integer programming package (when the inequalities are consistent) to produce a string of event symbols corresponding to a system behavior with the desired properties. The organization of the toolset is illustrated in the figure.

The current toolset is intended for use with designs written in the Ada-based design language CEDL (Constrained Expression Design Language) [11]. CEDL focuses on the expression of communication and synchronization among the tasks in a distributed system, and language features not related to concurrency are kept to a minimum. Thus, for example, data types are limited, but most of the Ada control-flow constructs have correspondents in CEDL. We have chosen to work with a design notation based on Ada because Ada is one of the few programming languages in relatively widespread use that explicitly provides

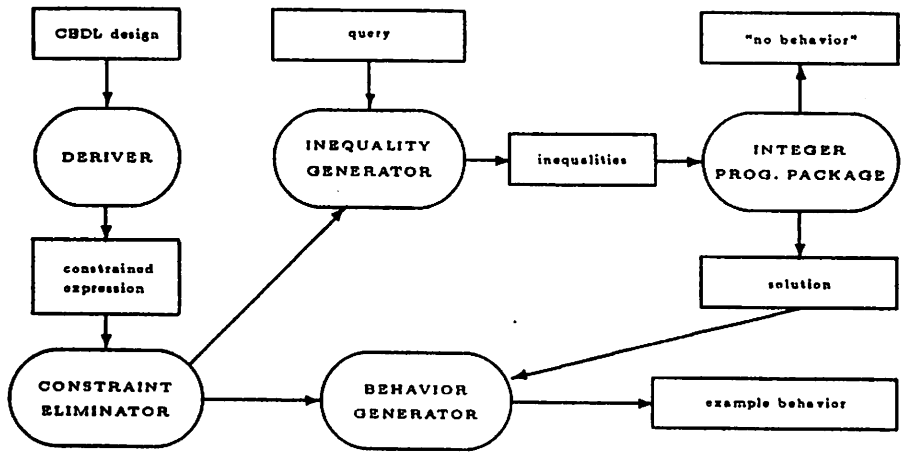

Figure 1: Diagram of Constrained Expression Toolset

for concurrency, and because we expect our work on analysis of designs to contribute to and benefit from the Arcadia Consortium's work on Ada software development environments [21]. Those aspects of the toolset that depend on CEDL are noted below.

The deriver [1] produces constrained expression representations from CEDL system designs. It is written in Ada, and was developed using Arcadia-produced versions of standard compiler construction tools and the Graph Definition Language and GRAPHITE processor [8]. The deriver generates a graph representing the constrained expression. Eventually, this will be the standard internal representation for constrained expressions. Currently, however, the prototypes of the other tools expect input in other formats, and small utility programs convert between the formats. For a CEDL design, the system expression of the constrained expression representation produced by the deriver consists of the interleave of ***task expressions*** representing the behavior of the tasks in the system. The deriver also generates all required constraints.

The constraint eliminator [13] is written in Common LISP, and shares a common front end with the behavior generator. It takes a "generator expression", which is a subexpression of the system expression, and constraints involving symbols from the generator expression, and produces a new expression whose language is the set of strings in the language of the generator expression that satisfy the constraints. The constraint eliminator converts the generator expression and constraints into finite state automata (the constraints in a constrained expression representation of a CEDL system are regular), from which it produces a new automaton accepting the intersection of the appropriate languages. It then returns an expression corresponding to this automaton. In principle, the generator expression need only be regular, and so could be the full system expression. However, the process of intersecting the finite state automata quickly becomes intractable if the generator expression involves the interleave operator. For this reason the current constraint eliminator will only accept generator expressions that use the standard regular expression operators. When analyzing constrained expressions derived from CEDL designs, the constraint eliminator is typically used with a task expression and the constraints that enforce correct dataflow within tasks. The input task expression is then replaced with the expression returned by the constraint eliminator and the intra-task data flow constraints are eliminated. This process facilitates certain aspects of the analysis of the constrained expression, as indicated below.

The inequality generator [3], which is also written in Common LISP, takes a constrained expression representation in essentially the same format as that accepted by the constraint eliminator, and generates a system of linear inequalities representing part of the semantics of the constrained expression. The inequality generator builds an abstract syntax tree representing each task expression, and generates inequalities reflecting the semantics of regular expressions. It then generates additional inequalities derived from some of the constraints. The full system of inequalities thus involves both the total numbers of occurrences of various event symbols and the numbers of times various branches in the abstract syntax trees are traversed. However, these inequalities do not reflect the *complete* semantics of the constrained expression. For example, not all the information about the relative order of event symbols is represented, so that the constraints that enforce correct dataflow are not reflected in the generated system of inequalities. (The significance of this problem for intra-task dataflow is reduced by application of the constraint eliminator.) In addition, the generated system of inequalities does not completely reflect the semantics of the alternation operator when one of its operands is the Kleene star of an expression. The full semantics would require quadratic inequalities, and the integer programming package we are currently using only handles linear systems. As mentioned below, we are currently investigating another integer programming package that may eliminate this problem.

The inequality generator also provides an interactive facility allowing the analyst to add inequalities representing assumptions or queries about the behavior of the system. The inequality generator produces an output file giving the system of inequalities in the format required by the integer programming package, as well as a human-readable report giving the correspondence between variables in the system of inequalities and event symbols. If the integer programming package finds a solution to the system of inequalities, the inequality generator uses this correspondence to report the solution to the analyst in terms of traversal of the abstract syntax trees and the numbers of occurrences of event symbols. Certain aspects of the inequality generator, including the representation of various constraints, depend on features of CEDL. Its basic structure, however, is compatible with all constrained expressions.

The integer programming package that we are currently using is a branch-and-bound integer linear programming system [17] written in FORTRAN; it was chosen because it had already been installed as part of a previous project at the University of Massachusetts. We have encountered some problems with its branch-and-bound strategy, as described in the next section, and with the limitation to linear systems. We are currently implementing integer programming on top of the MINOS optimization package [20].

The behavior generator [14] is a Common LISP program for producing system behaviors with certain properties. Input to the program consists of a constrained expression and counts for certain event symbols (counts produced by the integer programming package). The behavior generator builds finite state automata corresponding to the task expressions and constraints of the constrained expression. It then uses heuristic search techniques to find a string of event symbols representing a system behavior with the given numbers of symbol occurrences. It can be used with any constrained expression having regular constraints.

4 Using the Toolset

We have begun to use the prototype toolset in the analysis of concurrent systems. The table in Figure 2 gives CPU times for the application of the components of the toolset to several versions of the dining philosophers problem. All times were obtained on a Sun 3/60 workstation.

The first six rows of the table give the times for versions of the standard dining philosophers problem with three, four, five, six, eight, and ten philosophers, respectively. In the analysis reported here, we seek to determine whether a particular philosopher task can wait indefinitely for a rendezvous with a second fork task, and thus starve (in both the concurrent systems and metaphorical senses). These systems do not have a doorkeeper or host to prevent all the philosophers from trying to pick up forks at the same time, and are therefore subject to deadlock in which each philosopher task starves waiting to rendezvous with a second fork task. The dining philosophers systems without host use rendezvous simply for synchronization purposes, and involve no intratask dataflow. We therefore do not use the constraint eliminator in these cases.

The size of the constrained expression representations of these systems goes up linearly with the number of philosophers, as does the execution time of the inequality generator and the size of the system of inequalities generated. We have successfully applied the deriver and inequality generator with systems containing up to twenty philosophers (i.e., forty concurrent tasks), and expect no difficulties with even larger systems. However, the integer programming package we are currently using is unable to solve the systems of inequalities generated in the cases with more than eight philosophers, due to failure of an accuracy test in the course of solving a linear programming relaxation of the

system	deriver	constraint eliminator	inequality generator	int. prog. package	behavior generator	total CPU time
DP-3	74	—	11	4	19	108
DP-4	82	—	13	5	26	126
DP-5	94	—	15	6	32	147
DP-6	109	—	18	9	38	174
DP-8	142	—	24	14	54	234
DP-10	177	—	30	—	—	—
DPH-3	123	77	23	7	—	230
DPH-4	133	194	62	41	—	430
DPH-5	152	330	102	103	—	687

Figure 2: Sun 3/60 CPU times, in seconds, for the constrained expression tools in analysis of several dining philosophers systems

integer linear programming problem. We discuss the implications of this failure below. In the cases where a solution to the system of inequalities is found, the behavior generator produces a behavior exhibiting the deadlock.

The final three rows of the table give times for analysis of three-, four- and five-philosopher versions that have a host task to prevent all the philosophers from entering the dining room and trying to pick up forks at the same time. Again, the analysis seeks to determine whether a particular philosopher can starve. In these cases, the constraint eliminator is applied to the task expression for the host, along with the constraints enforcing consistent use of the variable that counts the number of philosophers in the dining room. The resulting task expression is used in the input to the inequality generator. Because the task expression for the host must represent the effects of all possible execution paths on the variable that counts philosophers in the dining room, the size of the system of inequalities goes up rapidly with the number of philosophers, and is significantly greater for the five-philosopher system with a host than for the eight-philosopher system without a host. Due to the detailed structure of the particular system of inequalities, however, the integer programming package does not encounter accuracy problems here, and, in each of the three cases, reports that no philosopher starves. Thus, it is not necessary to use the behavior generator in these cases.

We have also applied the toolset to the gas station system examples of [2, 22]. In these cases, we use the constraint eliminator with the task expression representing the operator of the automated gas station and the constraints enforcing consistent use of the variable that counts the number of the customers pumping gas. Even with only two customer tasks in the system, the behavior of the operator is considerably more complex than that of the host in the dining philosophers systems, and the system of inequalities produced by the inequality generator is too large to be expressed in the input format of the Land-Powell integer programming package. We expect that improvements in the constraint eliminator and the conversion to an integer programming package based on MINOS will very soon allow us to apply the complete toolset to these systems as well.

5 Conclusions

These initial experiments with the prototype constrained expression toolset are encouraging. The toolset provides complete analysis of both versions of the dining philosophers problem, with and without a doorkeeper. Even the prototype versions of the tools are efficient enough to be useful to software developers on examples of moderate size. Furthermore, earlier experiments show that the constrained expression approach can detect a variety of errors and can be used with a broad range of design notations and programming languages.

However, some weaknesses of the prototype toolset are evident. The most significant of these involve the branch-and-bound integer programming package we are currently using [17], and include the limitations on the size of system of inequalities that the package can handle, the accuracy problems noted in the previous section, and the restriction to linear inequalities. This package is an implementation in FORTRAN 66 of the first branch-and-bound algorithm for general integer programs [16]. Its division scheme has been replaced, in virtually all commercial integer programming codes, by the variable dichotomy scheme first proposed by Dakin [9], and we believe that some of its strategies for selecting a branching variable and for exploring the tree may be poorly suited to our systems of inequalities. For these reasons and others, including the ability to handle quadratic inequalities, we expect a considerable improvement in performance from the integer programming package we are currently implementing using the MINOS optimization package [20].

Other drawbacks of the prototype toolset were not as significant in the experiments described here, but may become more important when the toolset is applied to a wider range of concurrent systems. These include the facts that the system of inequalities produced by the inequality generator does not reflect the full semantics of the constrained expression representation (though the use of quadratic inequalities with the MINOS system addresses part of this issue) and that the task expressions returned by the constraint eliminator may lead to larger systems of inequalities than other, equivalent expressions.

We are now beginning to address these issues, and a number of improvements to the toolset are planned. In addition to replacing the Land-Powell integer linear programming package with one based on MINOS that will allow quadratic inequalities, we intend to modify the behavior generator to use all the information contained in a solution to the system of inequalities, rather than just the total numbers of occurrences of the various event symbols. We will be modifying the inequality generator to produce the quadratic inequalities needed to express the semantics of the alternation operator when one of its operands is the Kleene star of an expression, and are investigating other ways to improve the generation of inequalities so that they reflect more of the full semantics of constrained expressions.

In addition, improvements to the interfaces between the human analyst and the toolset and between the tools are underway. Clearly, analysts should be able to formulate behavioral queries in terms of elements from the original system description and at a higher level of abstraction than is currently possible, and a common internal representation would help in integrating the various tools. Eventually, the tools now written in LISP will be reimplemented in Ada to facilitate integration of the tools with each other and with the Arcadia software development environment.

While starting to improve the prototype toolset, we have also begun to explore additional applications for constrained expression analysis, some of which may lead to enhancements to the underlying formalism and further modifications to the tools. In particular, we have begun to study the application of the constrained expression approach to various scheduling and real-time problems [4]. Because expressing some of these scheduling and timing problems, as well as the semantics of certain programming languages for concurrent systems, involves constraints that are not regular expressions, we hope to be able to eliminate the regularity restrictions in some of the tools.

For a more complete understanding of the strengths and weaknesses of the constrained expression approach and the prototype toolset, we need to evaluate the performance of the toolset on a wider range of examples. The problem of designing an appropriate suite of benchmark problems for concurrent software analysis tools has not been carefully studied; we hope to develop some criteria for such a suite in the course of collecting additional examples for experiments with the constrained expression toolset. It is unlikely that a single approach to analysis will meet the needs of developers of concurrent software, and such a test suite would be of significant value in comparing various approaches and determining the types of problems for which each approach has the greatest value.

Based on the prototype toolset and the initial experiments described in this paper, we are very encouraged about the prospective value of the constrained expression approach to automated analysis of concurrent software systems. We therefore plan to pursue the toolset improvements, enhancements to the formalism, and more extensive experimental evaluation outlined above. We expect that these activities, in conjunction with similar experimental evaluations by other researchers developing other analysis techniques, preferably all based on a common test suite, will result in improved understanding of the relative strengths and weaknesses of the constrained expression approach and alternative concurrent system analysis techniques.

Acknowledgements

Our students Susan Avery, Ugo Buy, Michael Greenberg, RenHung Hwang, and George Walden have played important roles in the development of the toolset and the experimental evaluation reported in this paper.

References

[1] S. Avery. A tool for producing constrained expression representations of CEDL designs. In preparation.

[2] G. S. Avrunin. Experiments in constrained expression analysis. Technical Report 87-125, Department of Computer and Information Science, University of Massachusetts, Amherst, November 1987.

[3] G. S. Avrunin and U. Buy. An inequality generator for constrained expression analysis. In preparation.

[4] G. S. Avrunin, L. K. Dillon, and J. C. Wileden. Constrained expression analysis of real-time systems. Submitted.

[5] G. S. Avrunin, L. K. Dillon, J. C. Wileden, and W. E. Riddle. Constrained expressions: Adding analysis capabilities to design methods for concurrent software systems. *IEEE Trans. Softw. Eng.*, SE-12(2):278–292, 1986.

[6] G. S. Avrunin and J. C. Wileden. Describing and analyzing distributed software system designs. *ACM Trans. Prog. Lang. Syst.*, 7(3):380–403, July 1985.

[7] R. H. Campbell and A. N. Habermann. The specification of process synchronization by path expressions. In E. Gelenbe and C. Kaiser, editors, *Operating Systems*, volume 16 of *Lecture Notes in Computer Science*, pages 89–102. Springer-Verlag, Heidelberg, 1974.

[8] L. A. Clarke, J. C. Wileden, and A. L. Wolf. GRAPHITE: A meta-tool for Ada environment development. In *Proceedings of 2nd International Conference on Ada Applications and Environments*, pages 81–90, April 1986.

[9] R. J. Dakin. A tree search algorithm for mixed integer programming problems. *Computer Journal*, 8:250–255, 1965.

[10] L. K. Dillon. *Analysis of Distributed Systems Using Constrained Expressions*. PhD thesis, University of Massachusetts, Amherst, 1984.

[11] L. K. Dillon. Overview of the constrained expression design language. Technical Report TRCS86-21, Department of Computer Science, University of California, Santa Barbara, October 1986.

[12] L. K. Dillon, G. S. Avrunin, and J. C. Wileden. Constrained expressions: Toward broad applicability of analysis methods for distributed software systems. *ACM Trans. Prog. Lang. Syst.*, 10(3):374–402, July 1988.

[13] L. K. Dillon and G. Walden. A prototype constraint eliminator for constrained expression representations. In preparation.

[14] M. Greenberg and S. Avery. A behavior generator. Software Development Laboratory Memo 89-1, Department of Computer and Information Science, University of Massachusetts, 1988.

[15] A. N. Habermann. Synchronization of communicating processes. *Commun. ACM*, 15(3):171–176, 1972.

[16] A. H. Land and A. G. Doig. An automatic method for solving discrete programming problems. *Econometrica*, 28:497–520, 1960.

[17] A. H. Land and S. Powell. *Fortran Codes for Mathematical Programming: Linear, Quadratic and Discrete.* John Wiley & Sons, Ltd., London, 1973.

[18] P. Lauer, P. Torrigiani, and M. Shields. COSY: A system specification language based on paths and processes. *Acta Informatica*, 12(2):451–503, 1979.

[19] W. E. Riddle. An approach to software system behavior modeling. *Computer Languages*, 4:29–47, 1979.

[20] M. A. Saunders. MINOS system manual. Technical Report SOL 77-31, Stanford University, Department of Operations Research, 1977.

[21] R. N. Taylor, F. C. Belz, L. A. Clarke, L. J. Osterweil, R. W. Selby, J. C. Wileden, A. L. Wolf, and M. Young. Foundations for the Arcadia environment architecture. In *Proceedings SIGSOFT '88: Third Symposium on Software Development Environments*, pages 1–13, December 1988.

[22] J. C. Wileden and G. S. Avrunin. Toward automating analysis support for developers of distributed software. In *Proceedings of the Eighth International Conference on Distributed Computing Systems*, pages 350–357. IEEE Computer Society Press, June 1988.