MAXIMAL PROGRESS STATE EXPLORATION

M.G.Gouda and Y.T.Yu

Dept. of Computer Sciences
University of Texas at Austin
Austin, TX 78712

ABSTRACT

Given two machines which communicate by exchanging messages over two finite-capacity channels, it is possible to generate all reachable states of the system and check whether any of them is a nonprogress state. This technique is called state exploration; and it usually requires large execution time and storage. In this paper, we discuss a more efficient variation of this technique. In particular, we show that the task of generating all reachable states can be divided into two independent subtasks. In each subtask, only the states reachable by allowing maximal progress for one machine are generated. We prove that a given system cannot reach a nonprogress state iff none of the states generated in each subtask is a nonprogress state. Since the two subtasks are completely independent, and since in most cases the time and storage requirements for each subtask are less than those for the origianal task, maximal progress state exploration can save time or storage over conventional state exploration.

I. INTRODUCTION

Many communication protocols can be modeled as two finite state machines that communicate by exchanging messages over two one-directional, FIFO channels [2,7,11,13]. Examples of such protocols are X.75 [8], and the call establishment/clear protocol in X.21 [10], and X.25 [1]. If the two channels in any such protocol are assumed to have finite capacities, then the protocol can be validated by generating all reachable states and checking whether any of them is a nonprogress state. This technique is referred to as state exploration. Some automatic protocol validation systems based on state exploration have been developed, and applied with reported success on actual protocols [3,5,9].

A major problem with state exploration is that it requires large execution time and storage. The problem is caused by the following assumption upon which state exploration is usually based. In order to generate all reachable states of a protocol, one needs to consider all possible progress speeds for the two machines in the protocol's model. This assumption leads to generating many "unimportant" states; it also leads to generating the same state many times since the same state can be reached by many different progress speeds for the two machines.

To counter this problem, Rubin and West [5] have shown that if the two machines progress in equal speeds, then the resulting reachable states can be used to detect deadlocks and unspecified receptions (but not necessarily overflows). This suggests to modify state exploration such that only states reachable by assuming equal progress speeds for the two machines are generated and examined. This modified state exploration requires, in most cases, less execution time and storage than conventional state exploration. (For example, the idea of restricting the two machines to equal progress speeds has led to an efficient algorithm to detect deadlocks in the case where the two machines exchange one type of message [12].) On the other hand, this technique cannot be used to detect nonprogress since nonprogress can be caused by overflows as well as by deadlocks and unspecified receptions.

In this paper, we discuss another variation of state exploration where the state generation task is divided into two independent subtasks. In each subtask, only states reachable by allowing one of the two machines to make maximal progress are generated and examined. We show that in this case the generated states can be used to detect deadlocks, unspecified receptions, and overflows; thus they can be used to detect nonprogress, if any. We also show that this technique can be used to verify properties, other than progress, such as detecting all nonexecutable transitions in each machine.

The paper is organized as follows. The model of communicating machines is discussed in section II. Then in section III, the concept of legal sequences, upon which conventional state exploration is based, is presented. In section IV, a special class of legal sequences called maximal progress sequences are characterized; and we prove

that these sequences can be used to detect deadlocks, unspecified receptions, and overflows. In section V, we discuss how to perform maximal progress state exploration (which is based on maximal progress sequences) to detect nonprogress states, if any. Concluding remarks are in section VI.

II. COMMUNICATING MACHINES

A communicating machine M is a directed labelled graph with two types of nodes called sending and receiving nodes. An output of a sending (or receiving) node is called a sending (or receiving) edge, and is labelled send(g) (or receive(g)) for some message g in a finite set G of messages. Each node in M must have at least one output edge; and outputs of the same node must have distinct labels. One of the nodes in M is identified as its initial node; and each node in M is reachable by a directed path from the initial node.

For simplicity, this model is a special case of the one in [2], where no node is allowed to have both sending and receiving outputs. In [4], we extend the discussion to communicating machines with mixed nodes (i.e., nodes that have both sending and receiving outputs).

Let M and N be two comunicating machines with the same set G of messages. A state of M and N is a four-tuple [v,w,x,y],

where: v and w are two nodes in M and N respectively, and

x and y are two strings of messages from the set G such that $|x| \leq K$ and $|y| \leq K$, where $|x|$ (or $|y|$) is the length of string x (or y, respectively), and K is a positive integer called the channel capacity between M and N.

Informally, a state [v,w,x,y] means that the execution of machine M has reached node v and the execution of N has reached node w, while the input channels of M and N have the message sequences x and y respectively. The above definition also implies that each of the two channels between M and N has a finite capacity of K messages.

The initial state of M and N is $[v_0,w_0,E,E]$ where v_0 and w_0 are the initial nodes in M and N respectively, and E is the empty string.

Let s=[v,w,x,y] be a state of M and N, and let e be an output edge of node v or w. A state s´ is said to follow s over e, denoted s--e-->s´, iff the following four conditions are satisfied:

i. If e is a sending edge, labelled send(g), from v to v´ in M, then $|y|<K$ and s´=[v´,w,x,y.g], where "." is the concatenation operator.

ii. If e is a sending edge, labelled send(g), from w to w´ in N, then $|x|<K$ and s´=[v,w´,x.g,y].

iii. If e is a receiving edge, labelled receive(g), from v to v´ in M, then x=g.x´ and s´=[v´,w,x´,y].

iv. If e is a receiving edge, labelled receive(g), from w to w´ in N, then y=g.y´ and s´=[v,w´,x,y´].

Let s and s´ be two states of M and N, s´ follows s, denoted s-->s´, iff there is a directed edge e in M or N such that s--e-->s´. If s-->s´ then s´ is called a follower of s.

Let s and s´ be two states of M and N. s´ is reachable from s iff s=s´ or there exist states $s_1,\ldots,s_r$ such that $s=s_1$, $s´=s_r$, and s_i-->s_{i+1} for i=1,...,r-1.

A state s of M and N is said to be reachable iff it is reachable from the initial state of M and N.

Since the two channels between M and N have finite capacities, the set of all reachable states is finite; and so usual state exploration techniques [3,8] can be used to generate all reachable states and check whether any of them is a nonprogress state. There are three types of nonprogress states, namely deadlock states, unspecified reception states, and overflow states. These are defined next:

A state [v,w,x,y] is a deadlock state iff v and w are receiving nodes, and x=y=E (the empty string).

A state [v,w,x,y] is an unspecified reception state iff one of the following two conditions holds.

i. $x=g_1 \cdot g_2 \cdot \ldots \cdot g_r$ and v is a receiving node and none of its outputs is labelled receive(g_1).

ii. $y=g_1 \cdot g_2 \cdot \ldots \cdot g_r$ and w is a receiving node and none of its outputs is labelled receive(g_1).

If condition i holds, then the state is called an unspecified reception state for M; if condition ii holds, it is called an unspecified reception state for N.

A state [v,w,x,y] of M and N is an overflow state iff one of the following two conditions holds.

i. Node v is a sending node; and $|y|=K$.

ii. Node w is a sending node; and $|x|=K$.

If condition i holds, then the state is called an overflow state for M; if condition ii holds, it is called an overflow state for N.

A state is called a nonprogress state iff it is a deadlock state, an unspecified reception state, or an overflow state; otherwise it is a progress state.

In the above discussion, a state s of M and N is reachable if there exist states $s_0,\ldots,s_r$ such that s_0 is the initial state of M and N, $s=s_r$, and s_i--e_{i+1}-->s_{i+1} for some edge e_{i+1} in M or N, i=0,...,r-1. One can view that state s is reachable by the "sequence" $\langle e_1,\ldots,e_r\rangle$ of edges. It is more convenient to replace each edge e_i in this sequence by a "symbol" q_i as follows:

If e_i is labelled send(g) in M, then q_i=´Msend(g)´.

If e_i is labelled send(g) in N, then q_i='Nsend(g)'.

If e_i is labelled receive(g) in M, then q_i='Mreceive(g)'.

If e_i is labelled receive(g) in N, then q_i='Nreceive(g)'.

Therefore, s is reachable by the sequence of symbols $Q=q_1.q_2. \ldots .q_r$, where "." is the concatenation operator. The concept of a sequence of symbols which can reach a state is central to the discussion in this paper; and it is characterized formally in the next section.

III. SEQUENCES AND LEGAL SEQUENCES

Let M and N be two communicating machines; and assume that each of the two channels between M and N has a finite capacity of K. A sequence $Q=q_1.q_2. \ldots .q_r$ of M and N is a finite string of symbols which satisfies the following two conditions:

i. Each symbol q_i in Q has one of the following four forms: 'Msend(g)', 'Mreceive(g)', 'Nsend(g)', 'Nreceive(g)', where g is some message in the set of messages of M and N. A symbol of the form 'Msend(g)' or 'Mreceive(g)' is called an M symbol; similarly, a symbol of the form 'Nsend(g)' or 'Nreceive(g)' is called an N symbol.

ii. There are two directed paths D_1 and D_2 which start from the initial nodes in M and N respectively such that the following two conditions are satisfied for each i=1,...,r.

   a. The ith edge in D_1 is labelled send(g) (or receive(g)) iff the ith M symbol in Q is Msend(g) (or Mreceive(g), respectively).

   b. The ith edge in D_2 is labelled send(g) (or receive(g)) iff the ith N symbol in Q is Nsend(g) (or Nreceive(g), respectively).

   Path D_1 (or D_2) is called the projection of Q onto M (or N), and can be referred to as Q_M (or Q_N, respectively).

Informally, a sequence Q defines a relative progress for the execution of M and N with respect to one another. So, if a symbol Msend(g_1) occurs before a symbol Nsend(g_2) in a sequence Q, then Q defines a relative progress where machine M sends a message g_1 before N sends a message g_2. Because of the dependency between the execution of M and that of N, not every relative progress is possible. In other words, not every sequence that can be defined can be executed; for instance, a sequence where Mreceive(g) occurs before Nsend(g) cannot be executed. The class of sequences which can be executed is called legal sequences and is defined next.

A sequence $Q=q_1. \ldots .q_r$ of M and N is called legal if it satisfies the following two conditions.

i. For any i=1,...,r, if the ith Mreceive symbol (or Nreceive symbol) in Q is Mreceive(g) (or Nreceive(g)), then the ith Nsend symbol (or Msend symbol) in Q must be Nsend(g) (or Msend(g)), and it must occur before the ith Mreceive symbol (or Nreceive symbol) in Q.

ii. For every prefix P of Q (i.e., $P=q_1. \ldots .q_s$ where $s \leq r$), $n_{MN}(P) \leq K$ and $n_{NM}(P) \leq K$, where

   $n_{MN}(P)$=
   The number of Msend symbols in P
   - The number of Nreceive symbols in P,

   $n_{NM}(P)$=
   The number of Nsend symbols in P
   - The number of Mreceive symbols in P.

Condition i means that every message should be received after it is sent and in the same order it is sent. Thus, condition i implies that no deadlock state can be reached before completing Q. Condition ii means that as the two machines M and N execute according to the relative progress defined by Q, the two channels between M and N do not overflow. The concept of a state which is reached at the end of a legal sequence is defined next.

Let Q be a legal sequence of M and N. Q is said to reach a state [v,w,x,y] of M and N iff the following two conditions are satisfied

i. Nodes v and w are the last nodes in pathes Q_M and Q_N, respectively. Recall that Q_M and Q_N are the projections of Q onto M and N, respectively.

ii. String x (or y) consists of all the messages sent along Q_N (or Q_M) minus those received along Q_M (or Q_N, respectively).

If Q reaches a state s, then s is said to be reachable by Q.

The next three lemmas follow immediately from the definitions of "reachable state," "sequence," "legal sequence," and "reachable by."

Lemma 1: A state of M and N is reachable iff it is reachable by a legal sequence of M and N. []

Lemma 2: A prefix of a sequence (or a legal sequence) of M and N is a sequence (or a legal sequence, respectively) of M and N. []

Lemma 3: Let P be a proper prefix of a legal sequence of M and N. Then, the state reachable by P is not a deadlock state. []

Based on the concept of legal sequences of M and N, usual state exploration for M and N can be explained as the following three-step procedure:

i. Generate the legal sequences of M and N, one by one, in the order of their lengths. Start by the shortest sequence; then generate longer and longer sequences. (The shortest legal sequence is the empty string; it reaches the initial state of M and N.)

ii. For each generated legal sequence Q, construct the state s reachable by Q, and check whether or not s is a nonprogress state.

iii. Repeat steps i and ii until one of the following two conditions holds.

   a. A nonprogress state is detected, in which case machines M and N can reach a nonprogress state.

   b. No nonprogress state has been detected and no new states can be constructed any more, in which case M and N cannot reach a nonprogress state.

In the next section, we characterize a special class of legal sequences called maximal progress sequences; then in section V, we discuss how to perform state exploration based on this special class of sequences.

IV. MAXIMAL PROGRESS SEQUENCES

Let M and N be two machines which communicate over finite-capacity channels; and let Q be a legal sequence of M and N. Q is called a maximal progress sequence for M iff it satisfies the following three conditions:

i. $Q = A_1.B_1.A_2.B_2. \ \ldots \ .A_n.B_n$, where

   $n \geq 1$,
   A_i $(i=1,\ldots,n)$ is a string of M symbols,
   B_i $(i=1,\ldots,n)$ is a string of N symbols, and
   A_1 and B_n can be empty strings.

ii. For $i=1,\ldots,n$, the sequence $A_1.B_1. \ \ldots \ .A_i$ must reach a state $[v,w,x,y]$ where v is a receiving node and $x=E$ (the empty string).

iii. For $i=1,\ldots,n$, the sequence $A_1.B_1. \ \ldots \ .B_i$ must reach a state $[v,w,x,y]$ where $x \neq E$, and the largest proper prefix of $A_1.B_1. \ \ldots \ .B_i$ must reach a state $[v',w',x',y']$ where $x'=E$.

Informally, a maximal progress sequence for M defines the following relative order of execution for M and N: First, M executes as much as possible; i.e., until it reaches a receiving node while its input channel is empty. Then, N executes until it sends exactly one message to M; i.e., M can now resume execution. Then, M executes as much as possible, and so on.

The next four theorems, whose proofs are in [4], state roughly that if a nonprogress state is reachable, then a nonprogress state (not necessarily the same one) is reachable by a maximal progress sequence. Notice that The converse is also true by lemma 1.

Theorem 1: If a deadlock state is reachable (by a legal sequence), then either a deadlock or an overflow state is reachable by a maximal progress sequence for M. []

Theorem 2: If an unspecified reception state for M is reachable (by a legal sequence), then either an unspecified reception state for M or an overflow state is reachable by a maximal progress sequence for M. []

Theorem 3: If an unspecified reception state for N is reachable (by a legal sequence), then either an unspecified reception state or an overflow state is reachable by a maximal progress sequence for M. []

Theorem 4: If an overflow state for M is reachable (by a legal sequence), then an overflow state is reachable by a maximal progress sequence for M. []

From these four theorems, if every state reachable by a maximal progress sequence for M is a progress state, then every reachable state is neither a deadlock state, an unspecified reception state, nor an overflow state for M. This does not guarantee that every reachable state is a progress state since some reachable states can still be overflow states for N. Therefore, to ensure that every reachable state is a progress state, it is sufficient (and necessary) to ensure that every state reachable by a maximal progress sequence for M or N is a progress state. The next theorem follows immediately.

Theorem 5: A nonprogress state is reachable (by a legal sequence) iff a nonprogress state is reachable by a maximal progress sequence for M or N. []

V. EFFICIENT STATE EXPLORATION

Based on Theorem 5, the following algorithm uses maximal progress state exploration to decide whether any two communicating machines with finite-capacity channels can reach a nonprogress state.

Algorithm 1:

Input: Two communicating machines M and N which communicate over two finite-capacity channels.

Output: A decision of whether M and N can reach a nonprogress state.

Steps:

i. Use Algorithm 2 (discussed below) to generate each state reachable by a maximal progress sequence for M. If any of these states is a nonprogress state, then stop: M and N can reach a non-

progress state.

ii. Use an algorithm similar to Algorithm 2 to generate each state reachable by a maximal progress sequence for N. If any of these states is a nonprogress state, then stop: M and N can reach a nonprogress state.

iii. If all generated states in steps i and ii are progress states, then stop: M and N cannot reach a nonprogress state. []

Now it remains to describe an algorithm to generate and examine all states reachable by maximal progress sequences for M. (An algorithm to generate and examine all states reachable by maximal progress sequences for N is similar.)

Algorithm 2:

Input: Two communicating machines M and N which communicate over two finite-capacity channels.

Output: A decision of whether any state reachable by a maximal progress sequence for M is a nonprogress state.

Variables: Two sets of states called OPEN and CLOSE. OPEN contains all the generated states which have not been examined yet; and CLOSE contains the states which have been generated and examined earlier. Initially, OPEN has the initial state of M and N and CLOSE is empty. The algorithm terminates when OPEN becomes empty or when one of the generated states is recognized to be a nonprogress state.

Steps: while OPEN is not empty do

i. remove one state s, selected at random, from OPEN.

ii. if s is a nonprogress state then stop: There is a nonprogress state reachable by a maximal progress sequence for M.

iii. if s is in CLOSE

then skip
else add s to CLOSE;
if s=[v,w,x,y],
where v is a receiving node
and x=E

then
a. generate all states s´ such that s--e-->s´, and e is an edge in N
b. add all the generated states s´ to OPEN.

else
a. generate all states s´ such that s--e-->s´, and e is an edge in M.
b. add all the generated states s´ to OPEN.

endwhile;
if OPEN is empty
(at the end of the while-statement)
then stop:
Each state reachable by a maximal progress sequence for M is a progress state. []

From step iii, not all followers of a reachable state s are generated. In particular, from any state s=[v,w,x,y], either machine M or N, but not both, can progress. If v is a receiving node and x=E (i.e., machine M cannot progress), then only machine N can progress; otherwise, only machine M can progress. Since from each state, at most one machine can progress, the number of states generated by Algorithm 2 are, in most instances, less than the number of states generated by conventional state exploration. (It is possible, however, to construct an M and N such that from each reachable state of M and N, at most one machine can make a progress. In such cases, the number of states generated by Algorithm 2 equals the number of states generated by conventional state exploration.)

Example 1: Figures 1a and 1b show two communicating machines M and N respectively. Assume that M and N communicate over two channels, each with a capacity of two. Figure 1c shows the states of M and N reachable by maximal progress sequences for M; i.e., those generated by Algorithm 2. The states reachable by maximal progress sequences for N are shown in Figure 1d. Since each of the states in Figures 1c and 1d is a progress state, M and N cannot reach a nonprogress state by Theorem 5.

Let us compare this maximal progress state exploration with conventional state exploration. Figure 1e shows all the reachable states of M and N generated by conventional state exploration. There is a total of 16 states in Figures 1c and 1d, compared with 17 states in figure 1e. (Notice that we have counted repeated states since it takes the same amount of effort to generate and check a repeated state as it is to generate a new state.) From this example, maximal progress state exploration can reduce the total number of generated states. Nevertheless, its real advantage lies in dividing the space of reachable states into two independent (but not necessarily disjoint) subspaces, namely the states reachable by maximal progress sequences for M, and those reachable by maximal progress sequences for N. As discussed next, this can be exploited to reduce the execution time or storage, required to perform state exploration.

Reduction in execution time can be achieved by generating the states reachable by maximal progress sequences for M in parallel with those reachable by maximal progress sequences for N. For instance, if the states in Figure 1c are generated in parallel with those in Figure 1d, then the execution time will be reduced by a factor of two. Obviously, two processors, instead of one, are required to execute this parallel state exploration.

During state exploration, the distinct states generated earlier are stored in a set (e.g.,set CLOSE in Algorithm 2) so that they can be compared with each state generated later. Since the number of distinct states reachable by maximal progress sequences for M (or N) is usually less than the number of distinct states reachable by legal sequences, the storage requirement of maximal

progress state exploration is usually less than that of conventional state exploration. For exemple, a set of seven distinct states is needed to perform the maximal progress state exploration in Figures 1c and 1d provided that they are performed in sequence, one after another. This is better than the set of eleven distinct states needed to perform the conventional state exploration of Figure 1e. []

VI. CONCLUDING REMARKS

We have presented a variation of state exploration called maximal progress state exploration. It has the advantage of dividing the state exploration task into two completely independent subtasks. In most cases, each subtask requires less execution time and storage than the original task. Thus, by executing these two subtasks in parallel, on the expense of using two processors instead of one, the required execution time is reduced. By executing these two subtasks in sequence on the same processor, the required storage is reduced.

The actual amount of reduction in execution time and/or storage depends on the two communicating machines under consideration. In all the examples we have tried (including those presented in [4]), the reduction in execution time, measured by the total number of generated states, is about 50%; and the reduction in storage, measured by the number of distincted states, is about 30%. More accurate results concerning the actual gain require considering a large population of examples and long experience with the technique.

The discussion in this paper has been limited to a model of communicating machines where no node is allowed to have both sending and receiving output edges. In [4], we discuss how to perform maximal progress state exploration for an extended model where nodes can have both sending and receiving outputs. We also show that using maximal progress state exploration (instead of conventional state exploration) to analyze an X.25 protocol can lead to a noticeable reduction in execution time and storage. Also in [4], we show that maximal progress state exploration can be used to check other properties (i.e., other than progress) for the communicating machines.

In this paper and in [4], we have only addressed the case of two communicating machines. Extending maximal progress state exploration to more than two machines is still an open problem.

REFERENCES

[1] G. V. Bochmann, "Finite state description of communication protocols," Computer Networks, Vol. 2, 1978, pp. 361-371.

[2] D. Brand and P. Zafiropulo, "On communicating finite-state machines," IBM Research Report, RZ1053(#37725), Jan. 1981. To appear in JACM.

[3] J. Hajek, "Automatically verified data transfer protocols," Proc. Int. Comp. Conf., 1978, pp. 749-756.

[4] M. G. Gouda, and Y. T. Yu, "Protocol validation by maximal progress state exploration," Technical Report 211, Dept. of Computer Sciences, University of Texas at Austin, July 1982.

[5] J. Rubin, and C. H. West, "An improved protocol validation technique," Computer Network, April 1982.

[6] H. Rudin, and C.H. West, "A validation technique for tightly coupled protocols," IEEE Trans. Computers, Vol. C-31, No.7, July 1982.

[7] C. A. Sunshine, "Formal modeling of communication protocols," USC/Inform. Sc. Institute, Res. Rep. 81-89, March 1981.

[8] S.T. Vuong, and D.D. Cowan, "Automated protocol validation via resynthesis: The CCITT X.75 packet level recommendation as an example," Tech. Report CS-80-39, Dept. of Computer Science, Univ. of Waterloo, revised May 1981.

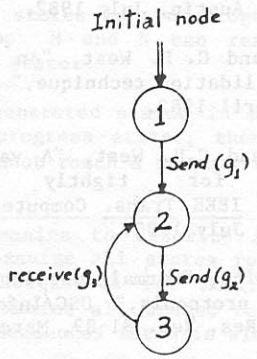
[9] C. H. West, "An automated technique of communication protocol validation," IEEE Trans. Comm., Vol. COM-26, pp.1271-1275, Aug. 1978.

[10] C. H. West and P. Zafiropulo, "Automated validation of a communications protocol: The CCITT.21 recommendation," IBM J. Res. Develop., vol. 22, pp.60-71, Jan. 1978.

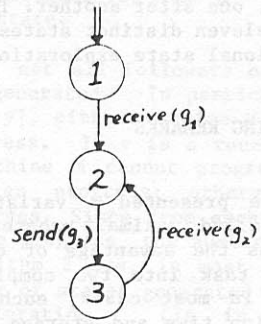
[11] Y. T. Yu "Communicating finite state machines: analysis and synthesis," PH.D Thesis, Dept. of Computer Sciences, Univ. of Texas at Austin, in progress.

[12] Y.T. Yu, and M.G. Gouda, "Deadlock detection for a class of communicating finite state machines," To appear in IEEE Trans. on Comm., Dec. 1982.

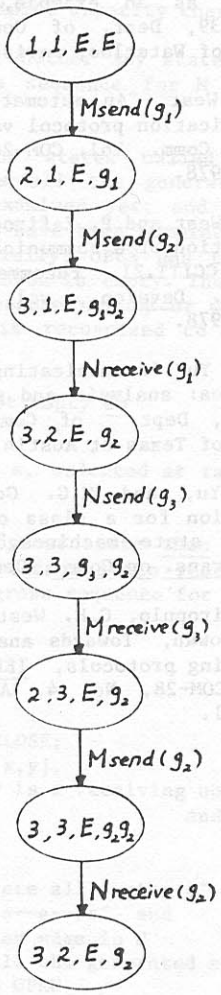
[13] P. Zafiropulo, C.H. West, H. Rudin, and D.D. Cowan, "Towards analyzing and synthesizing protocols," IEEE Trans. Comm., Vol. COM-28, No. 4, April 1980, pp. 651-661.



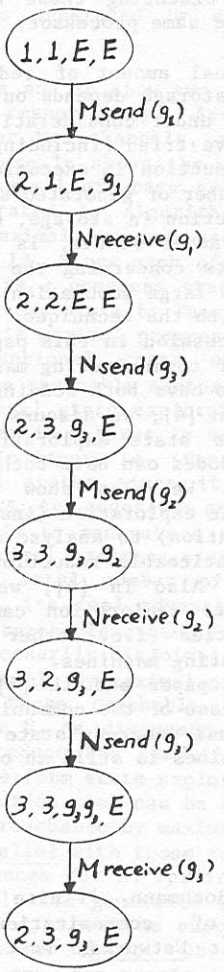
(a) M



(b) N

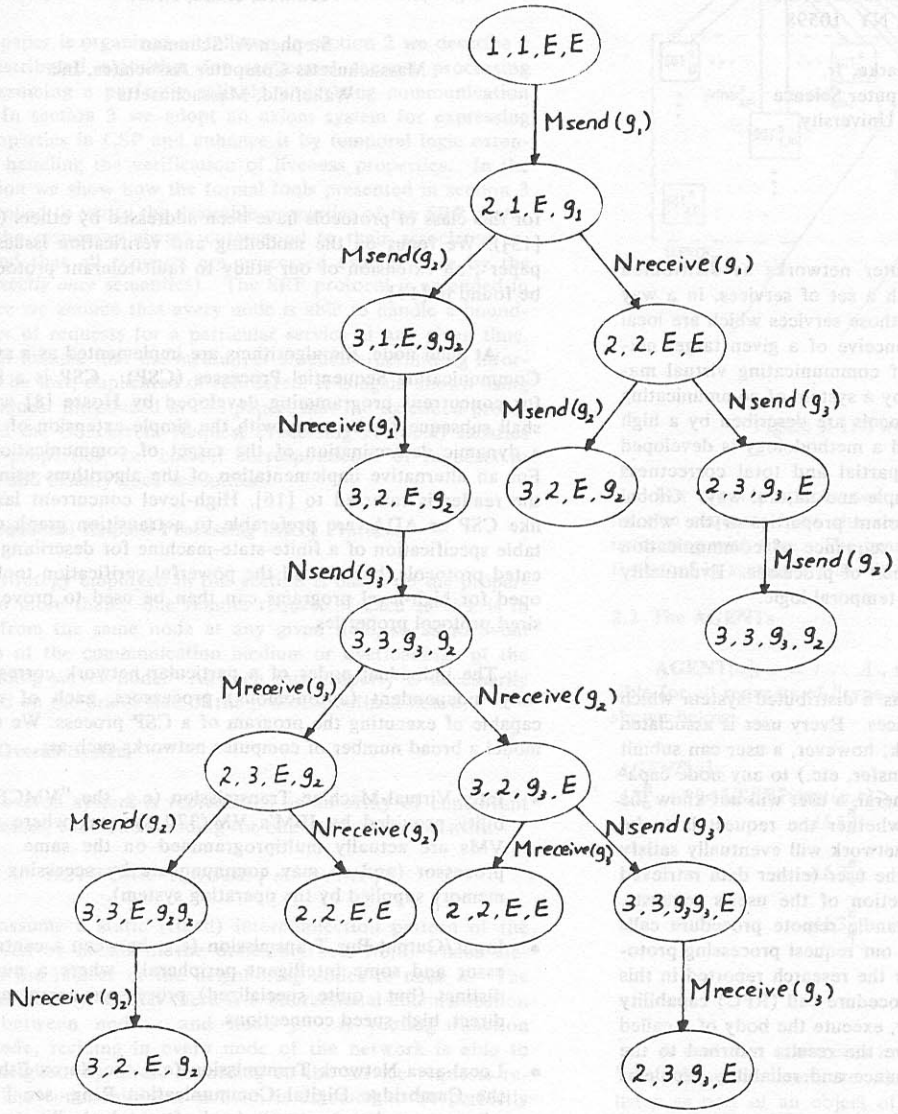


(c) State exploration by Maximal progress for M



(d) State exploration by maximal progress for N

Figure 1. First Example



(e) Conventional state exploration

Figure 1 Continues