Petri Net Models of Process Synchronization Mechanisms

W.M. Zuberek
Department of Computer Science
Memorial University of Newfoundland
St.John’s, Canada A1B 3X5

Abstract
Inhibitor Petri net models of several popular process synchronization mechanisms are presented and discussed. Semaphores and extended semaphores, monitors and rendezvous concepts are used in simple examples of process synchronizations. The corresponding Petri net models are used to verify basic properties such as mutual exclusion, presence or absence of deadlocks, or priorities in accessing shared resources.

1. Introduction
Petri nets are formal, mathematical models of systems with asynchronous concurrent activities [1, 18, 14]. Examples of such systems include multiprocessor computer systems, distributed databases and real-time industrial process control systems. As a modeling tool, Petri nets offer a simple and general formalism for representation of concurrent activities and synchronization of events, with a well-developed formal foundation for analysis of such models.

At a higher level of abstraction, Petri nets can model many synchronization and coordination mechanisms developed for concurrent programming [3, 20]; mutual exclusion in accessing shared information and message passing in distributed systems are examples of simple applications of such mechanisms. Due to developments in processor technology, multiprocessor systems constructed from a number of similar self-contained processors, are becoming quite popular. In order to use such systems on a single task, the component processors must be able to communicate and to synchronize with each other. Many methods have been presented for such synchronization: semaphores, critical and conditional critical regions, monitors, path expressions, rendezvous and others. Although they have been demonstrated to be adequate for their purpose, there is no widely recognized criterion for choosing among them.

The main goal of this paper is to show that many diverse synchronization mechanisms can be compared within one uniform framework of (inhibitor) Petri net models. Moreover, simple properties of Petri nets (e.g., boundedness, absence of deadlocks) can be used for verification of typical synchronization problems. This can help to understand and clarify different notations that have been proposed in the literature to deal with parallelism.

2. Inhibitor Petri nets
An inhibitor (place/transition, ordinary) Petri net [18, 14] is a quadruple \( N = (P.T. A.B) \), where \( P \) is a finite, nonempty set of places, \( T \) is a finite, nonempty set of transitions. \( A \) is a set of directed arcs connecting places with transitions and transitions with places, \( A \subseteq P \times T \cup T \times P \), and \( B \) is a finite set of inhibitor arcs connecting places with transitions, which is disjoint with \( A \). \( B \subseteq P \times T \). \( A \cap B = \emptyset \).

A place \( p \) is an input (or an output) place of a transition \( t \) iff there exists an arc \((p,t)\) (or \((t,p)\), respectively) in the set \( A \). The sets of all input and output places of a transition \( t \) are denoted by \( \text{In}(t) \) and \( \text{Out}(t) \), respectively. Similarly, \( \text{In}(p) \) and \( \text{Out}(p) \) denote the sets of input and output transitions of a place \( p \). A place \( p \) is an inhibitor place of a transition \( t \) iff there exists an inhibitor arc \((p,t)\) in the set \( B \).

A marked inhibitor net \( M = (N,m_0) \), where \( N \) is an inhibitor Petri net, \( N = (P.T. A.B) \), and \( m_0 \) is an initial marking function which assigns a nonnegative integer number of so-called tokens to each place of the net, \( m_0 : P \rightarrow \{0,1,\ldots\} \).

Let any function \( m : P \rightarrow \{0,1,\ldots\} \) be called a marking of a net \( N = (P.T. A.B) \). A transition \( t \) is enabled by a marking \( m \) if every input place of this transition contains at least one token and every inhibitor place of \( t \) contains zero tokens. The set of all transitions enabled by a marking \( m \) is denoted by \( E(m) \).

Every transition enabled by a marking \( m \) can fire. When a transition fires, a token is removed from each of its input places (but not inhibitor places) and a token is added to each of its output places. This determines a new marking in a net, a new set of enabled transitions, and so on.

A marking \( m_1 \) is directly reachable from a marking \( m_0 \) if...
\( m_i \) in a net \( N \) iff there exists a transition \( t_k \) enabled by the marking \( m_i, t_k \in E(m_i) \), such that for all \( p \in P \):

\[
m_j(p) = \begin{cases} 
  m_i(p) - 1, & \text{if } p \in \text{Inp}(t_k) - \text{Out}(t_k), \\
  m_i(p) + 1, & \text{if } p \in \text{Out}(t_k) - \text{Inp}(t_k), \\
  m_i(p), & \text{otherwise}.
\end{cases}
\]

Also, a marking \( m_j \) is (generally) reachable from a marking \( m_i \) in a net \( N \) iff there exists a sequence of directly reachable markings \( m_{i_0}, m_{i_1}, m_{i_2}, \ldots m_{i_k} \) such that \( m_{i_0} = m_i \) and \( m_{i_k} = m_j \).

A set of reachable markings, \( M(M) \), of a marked net \( M = (N, m_0) \) is the set of all markings which are reachable from the initial marking \( m_0 \) in the net \( N \).

A marking graph \( G(M) \) of a marked Petri net \( M \) is a directed graph \( G(M) = (W, D) \) where \( W \) is a set of vertices which is equal to the set of reachable markings of the net \( M, W = M(M) \), and \( D \) is a set of directed arcs, \( D \subset W \times W \), such that \( (m_i, m_j) \) is in \( D \) iff \( m_j \) is directly reachable from \( m_i \) in \( M \). Quite often additional information is attached to vertices or arcs of a marking graph. In particular, the arcs connecting the nodes (i.e., markings) can be labeled by the firing transitions.

Marking graphs provide complete behavioral characterization of marked nets. One of the most important behavioral properties of nets is boundedness; a marked net \( M \) is bounded iff its set of reachable markings \( M(M) \) is finite. For nets without inhibitor arcs many properties can be deduced from the structure of the net [6]; for bounded inhibitor nets, such structural properties are often insufficient, so the set (or graph) of reachable markings is used for further analyses.

Each net \( N = (P, T, A, B) \) can conveniently be represented by a connectivity (or incidence) matrix \( C : P \times T \rightarrow \{-1, 0, +1\} \) in which places correspond to rows, transitions to columns, and for all \( p \in P \) and all \( t \in T \), the entries are defined as:

\[
C[p, t] = \begin{cases} 
  -1, & \text{if } t \in \text{Out}(p) - \text{Inp}(t), \\
  +1, & \text{if } t \in \text{Inp}(p) - \text{Out}(p), \\
  0, & \text{otherwise}.
\end{cases}
\]

If a marking \( m_j \) is obtained from another marking \( m_i \) by firing a transition \( t_k \) then (in vector notation) \( m_j = m_i + C[k] \), where \( C[k] \) denotes the \( k \)-th column of \( C, i.e., the column representing \( t_k \).

Connectivity matrices disregard inhibitor arcs and ‘self-loops’, that is, pairs of arcs \((p, t)\) and \((t, p)\); any firing of a transition \( t \) cannot change the marking of \( p \) in such a self-loop. A pure net is defined as a net without self-loops [18].

A \( P \)-invariant (place invariant) [18, 14, 6] of a net \( N \) is any nonnegative, nonzero integer (column) vector \( I \) which is a solution of the matrix equation

\[
C^T \times I = 0,
\]

where \( C^T \) denotes the transpose of matrix \( C \). It follows immediately from this definition that if \( I_1 \) and \( I_2 \) are \( P \)-invariants of \( N \), then also any linear (positive) combination of \( I_1 \) and \( I_2 \) is a \( P \)-invariant of \( N \).

A basic \( P \)-invariant of a net is defined as a \( P \)-invariant which does not contain simpler invariants. All basic \( P \)-invariants \( I \) of ordinary nets are binary vectors \( I : P \rightarrow \{0, 1\} \).

A net \( N_i = (P_i, T, A_i, B_i) \) is a \( P \)-implied subnet of a net \( N = (P, T, A, B) \), \( P_i \subset P \), iff:

1. \( T_i = \{ t \in T \mid \exists p \in P_i : (p, t) \in A \} \),
2. \( A_i = A \cap (P_i \times T \cup T \times P_i) \),
3. \( B_i = B \cap (P_i \times T) \).

It should be observed that in a pure net \( N \), each \( P \)-invariant \( I \) of \( N \) determines a \( P \)-implied (invariant) subnet of \( N \), where \( P_i = \{ p \in P \mid I(p) > 0 \} \); all nonzero elements of \( I \) select rows of \( C \) and each selected row \( i \) corresponds to a place \( p_i \) with all its input (+1) and all output (−1) arcs associated with it.

![Fig. 1. Net model of mutual exclusion.](image1)

![Fig. 2. \( P \)-invariant–implied subnets for Fig. 1.](image2)

For the Petri net shown in Fig. 1, the connectivity matrix is:

<table>
<thead>
<tr>
<th>( NCS1 )</th>
<th>( t_1 )</th>
<th>( t_2 )</th>
<th>( t_3 )</th>
<th>( t_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C1 )</td>
<td>-1</td>
<td>+1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( CS1 )</td>
<td>+1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( sem )</td>
<td>-1</td>
<td>+1</td>
<td>-1</td>
<td>+1</td>
</tr>
<tr>
<td>( CS2 )</td>
<td>0</td>
<td>0</td>
<td>+1</td>
<td>-1</td>
</tr>
<tr>
<td>( NCS2 )</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>+1</td>
</tr>
</tbody>
</table>

and there are three basic \( P \)-invariants, \( I_1 = [1, 1, 0, 0, 0] \), \( I_2 = [0, 1, 1, 1, 0] \), and \( I_3 = [0, 0, 0, 1, 1] \). It can be observed that the basic invariants correspond to the smallest subsets of rows of the connectivity matrix with the (component-wise) sums equal to (vector) zero.

The \( P \)-implied subnets are simple nets in which all transitions have single input and single output places, as shown in Fig. 2. Consequently, the total number of tokens in each \( P \)-invariant subnet remains the same for all reachable marking. If a net is covered by such \( P \)-invariants, it is bounded for any initial marking \( m_0 \).
Finding basic P-invariants is a ‘classical’ problem of linear algebra [12, 6].

3. Semaphores

Providing mutual exclusion for a set of concurrent processes which access common (or shared) data is one of the basic synchronization problems [3, 17]. Usually such an exclusive access to shared data is localized within critical sections, so a synchronization mechanism is needed to guarantee that, at any time, at most one of concurrent processes is in its critical section.

An elegant solution to the mutual exclusion problem was proposed by Dijkstra in the form of semaphores [7]. Informally, a (counting) semaphore is an integer variable with just two indivisible operations called P (test and decrement) and V (increment). A process executing a P operation must wait until the semaphore is positive before it can decrement the semaphore’s value and continue. A V operation simply increases the semaphore’s value, possibly allowing some other process to execute a delayed P operation and continue. No two P or V operations on the same semaphore can be executed simultaneously.

A simple solution to mutual exclusion of two cyclic processes, Process 1 and Process 2, with their critical sections CS1 and CS2, respectively, uses a global semaphore $sem$, initialized to 1, as shown in Tab.1.

<table>
<thead>
<tr>
<th>var sem : semaphore = 1;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process1: process;</td>
</tr>
<tr>
<td>begin loop</td>
</tr>
<tr>
<td>.....</td>
</tr>
<tr>
<td>P(sem);</td>
</tr>
<tr>
<td>critical section 1;</td>
</tr>
<tr>
<td>V(sem);</td>
</tr>
<tr>
<td>end loop</td>
</tr>
<tr>
<td>end process;</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Process2: process;</td>
</tr>
<tr>
<td>begin loop</td>
</tr>
<tr>
<td>.....</td>
</tr>
<tr>
<td>P(sem);</td>
</tr>
<tr>
<td>critical section 2;</td>
</tr>
<tr>
<td>V(sem);</td>
</tr>
<tr>
<td>end loop</td>
</tr>
<tr>
<td>end process;</td>
</tr>
</tbody>
</table>

Tab.1. Mutual exclusion using semaphores.

A Petri net model of this solution is shown in Fig. 1. The semaphore $sem$ is represented by a place with the initial marking representing its initial value (i.e., 1 in this case). The processes are represented by two cyclic subnets in which P and V operations are modeled by transitions with arcs from (for P operations) or to (for V operations) $sem$; each P operation requires a positive value of the semaphore (otherwise the transition cannot be enabled), and each V operation increases the number of tokens in the semaphore by one.

It should be observed that the semaphore $sem$ and both critical sections belong to one of the P-invariant–implied subnets shown in Fig. 2. Since the initial marking assigns only one token to this subnet (the initial value of $sem$), the places CS1 and CS2 cannot be marked simultaneously, so at most one of critical sections can be ‘active’ at any time. Consequently, the solution guarantees the mutual exclusion of critical sections. Moreover, the model can easily be extended to any number of processes with any number of interactions between processes (controlled by identical or independent semaphores). Then, however, deadlocks can be created.

Semaphores are often used in resource allocation systems providing exclusive use of (shared) resources. Tab. 2 shows two cyclic processes. Process1 and Process2, dynamically requesting (P operations) and releasing (V operations) two resources r1 and r2 controlled by semaphores R1 and R2. It is known [16] that in such single-request systems immediate granting of requests may result in a deadlock.

<table>
<thead>
<tr>
<th>var R1,R2 : semaphore = 1,1;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process1: process;</td>
</tr>
<tr>
<td>begin loop</td>
</tr>
<tr>
<td>.....</td>
</tr>
<tr>
<td>P(R1);</td>
</tr>
<tr>
<td>P(R2);</td>
</tr>
<tr>
<td>use r1,r2;</td>
</tr>
<tr>
<td>V(R2);</td>
</tr>
<tr>
<td>V(R1);</td>
</tr>
<tr>
<td>end loop</td>
</tr>
<tr>
<td>end process;</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Process2: process;</td>
</tr>
<tr>
<td>begin loop</td>
</tr>
<tr>
<td>.....</td>
</tr>
<tr>
<td>P(R2);</td>
</tr>
<tr>
<td>P(R1);</td>
</tr>
<tr>
<td>use r1,r2;</td>
</tr>
<tr>
<td>V(R1);</td>
</tr>
<tr>
<td>V(R2);</td>
</tr>
<tr>
<td>end loop</td>
</tr>
<tr>
<td>end process;</td>
</tr>
</tbody>
</table>

Tab.2. Single-resource allocation using semaphores.

A Petri net model of this process synchronization is shown in Fig. 3, and its reachability graph in Fig.4. The node 9 clearly indicates a deadlock which can be reached by executing P(R1) operation by Process1 and then P(R2) operation by Process2 (or first P(R2) by Process2 and then P(R1) by Process1).
Petri Net Models of Process Synchronization Mechanisms

The second solution uses multiple semaphore operations, i.e., P and V operations which update simultaneously a list of semaphores [2], as shown in Tab.3, with a net model shown in Fig.5.

```
var R1, R2 : semaphore = 1, 1;
Process1: process;
begin loop
  P(R1, R2);
  use r1, r2;
  V(R1, R2);
end loop
end process;
```

Tab.3. Resource allocation with multiple semaphores.

Fig.5. Resource allocation with multiple semaphores.

Modeling of systems with priorities of some operations has resulted in the discovery of some limitations of Petri net models [11]. Inhibitor arcs have been proposed as an extension of basic Petri nets [1, 14], and it has been shown that Petri nets with inhibitor arcs are equivalent, with respect to ‘modeling power’, to Turing machines. Readers and writers problem [16] is a good illustration of limitations of Petri nets without inhibitor arcs.

A classical solution to this problem uses three semaphores [2], two counting semaphores nr and nw, both initialized to zero, and a binary semaphore s initialized to one. Extended semaphore operations P and V can be performed on several semaphores simultaneously, and also P operations can test semaphores for ‘zero conditions’ (if the second list, separated by ‘;’, is nonempty) [2]. The solution shown in Tab.4 assumes two classes of identical Reader and Writer processes (in general case the processes may be different).

```
var nw, nr, s : semaphore = 0, 0, 1;
Reader: process;
begin loop
  P(s; nw);
  V(nw);
  read;
  P(nr);
  V(s);
end loop
end process;

Writer: process;
begin loop
  V(nw);
  P(s; nr);
  write;
  P(nr);
  V(s);
end loop
end process;
```

Tab.4. Readers-writers synchronizations using extended semaphores.

Fig.6. Readers-writers synchronization.

A Petri net model of this solution is shown in Fig.6 (inhibitor arcs have small circles instead of arrows). By analyzing P-invariants and the set of reachable markings, it can be shown that the solution provides priority of Writer processes over Reader ones (i.e., when a Writer process is ‘ready’, no new Reader processes are allowed to enter their “read” section), that Reader processes have concurrent access to “reading”, and that there is mutual exclusion of Reader and Writer processes.

4. Monitors

An approach inspired by the class concept of Simula-67, and called monitor [4], is formed by encapsulating both, the shared data objects and operations that manipulate them:

```
<monitormame> : monitor;
<decls of common variables and conditions>
<definitions of monitor procedures>
<definitions of other (local) procedures>
begin <initialization code> end;
```

A monitor consists of a collection of variables that can be manipulated by all monitor procedures (but
Petri Net Models of Process Synchronization Mechanisms

which are inaccessible from outside the monitor), a set of monitor procedures that are used for manipulations of monitor variables and which are invoked by prefixed invocations:

\[ \text{<monitorname> }, \text{<proc-name> } (\text{<list of args> } ) \]

and a set of local procedures used by monitor procedures only. The \text{<initialization code> } is executed when a monitor is created.

Execution of all monitor procedures is `automatically' guaranteed to be mutually exclusive. This ensures that the monitor common variables are never accessed by more than one process. Moreover, special \text{condition} variables are used to delay processes executing monitor procedures. Two operations, \text{signal} and \text{wait}, are defined for condition variables. If \( x \) is a condition variable, then execution of \( x/\text{wait} \) causes the invoking process to be blocked on \( x \) and to relinquish its mutually exclusive control of the monitor. However, there are several 'interpretations' of the operation \text{signal} [16]. In one [4], the process invoking the operation immediately leaves the monitor making it available for the reactivated process (the invocation of \text{signal} operation is required to be the last statement of the corresponding monitor procedure). In another interpretation [9], the execution of \( x/\text{signal} \) depends upon the condition of the variable \( x \); if there is no process blocked on \( x \), the invoking process continues, otherwise the invoking process is temporarily suspended and one process blocked on \( x \) is reactivated and continued; the process suspended due to a \text{signal} operation continues when there is no other process executing in the monitor. Also, such processes are given priority over processes trying to begin execution of monitor procedures.

The monitor implementation of the bounded-buffer producer-consumer scheme uses a monitor Buffer with (buffer) operations \text{store} and \text{fetch} (the actual buffer is represented by an array \( B \) of \( K \) data elements), as shown in Tab.5. The buffer is accessed by cyclic

\text{Producer} and \text{Consumer} processes shown in Tab.6.

\begin{tabular}{|l|}
\hline
\textbf{Buffer : monitor:} \\
\textbf{var} \( B \) : array [1..\( K \)] of data; \\
\text{first}, \text{last}, \text{count} : integer; \\
\text{empty}, \text{full} : \text{condition}; \\
\text{procedure} \text{store} (x : data); \\
\text{begin} if \text{count} = \text{\( K \)} then \text{full} : \text{wait}; \\
\( B[\text{last}] := x; \) \\
\text{last} := (\text{last mod} \( K \)) + 1; \\
\text{count} := \text{count} + 1; \\
\text{empty} : \text{signal} \\
\text{end}; \\
\text{procedure} \text{fetch} (\text{var} x : \text{data}); \\
\text{begin} if \text{count} = 0 then \text{empty} : \text{wait}; \\
x := B[\text{first}]; \\
\text{first} := (\text{first mod} \( K \)) + 1; \\
\text{count} := \text{count} - 1; \\
\text{full} : \text{signal} \\
\text{end}; \\
\text{begin} \text{count} := 0; \text{first} := 1; \text{last} := \text{1 end}; \\
\hline
\end{tabular}

\begin{tabular}{|l|}
\hline
\textbf{Producer:} \text{process}; \\
\textbf{var} \text{item} : \text{data}; \\
\textbf{begin loop} \\
\text{produce(item)}; \\
\text{Buffer.fetch(item)}; \\
\text{Buffer.store(item)} \\
\text{end loop} \\
\text{end process}; \\
\hline
\end{tabular}

\begin{tabular}{|l|}
\hline
\textbf{Consumer:} \text{process}; \\
\textbf{var} \text{item} : \text{data}; \\
\textbf{begin loop} \\
\text{consume(item)} \\
\text{end loop} \\
\text{end process}; \\
\hline
\end{tabular}

Fig.7 shows a Petri net model of the producer-consumer bounded-buffer monitor; inhibitor arcs in
this model are used to indicate priorities of simultaneous events. It can be verified that the model is covered by simple P-inv.

5. Rendezvous

Rendezvous is an intertask communication and synchronization mechanism introduced in Ada. The primary mechanisms is composed of \texttt{accept} statements:

\begin{verbatim}
accept <entryname> ( <parameters> ) do <body> end
and entry calls:
<taskname> . <entryname> ( <parameters> )
\end{verbatim}

For two cooperating tasks (or processes), \(T_1\) and \(T_2\), let task \(T_1\) issue a call of an entry of task \(T_2\). There are two possible executions:

1. The entry call is issued before \(T_1\) reaches the corresponding \texttt{accept} statement; in this case \(T_1\) is suspended until \(T_2\) reaches its \texttt{accept} statement and completes the execution of its body.

2. The \texttt{accept} statement is reached by \(T_2\) before a call is received on that entry; in this case \(T_2\) is suspended until \(T_1\) issues its call.

As soon as \(T_2\) reaches its \texttt{accept} statement and a call of the corresponding entry is issued by \(T_1\), \(T_1\) is suspended while \(T_2\) executes the body of its \texttt{accept} statement, after which both tasks can continue their executions. This interaction is called \textit{rendezvous}. It should be noted that the mechanism is `asymmetric' since the same \texttt{accept} entry can be called by many tasks (there is a queue of tasks associated with each \texttt{accept} entry), while a call of an \texttt{accept} entry always uniquely identifies the entry.

An \texttt{accept} statement can be embedded within a \texttt{select} statement which provides a form of a (non-deterministic) multiple choice statement:

\begin{verbatim}
select when <condition-1> => <statements-1>
or when <condition-2> => <statements-2>
or ... 
else <statements>
end select
\end{verbatim}

Execution of a \texttt{select} statement is composed of three consecutive steps [21]:

1. all when conditions are evaluated to determine which alternatives are "open";
2. an open alternative \texttt{<statements-1>} (i.e., an alternative for which the corresponding \texttt{condition-1} is satisfied) is selected; an open alternative starting with an \texttt{accept} statement may be selected only if the corresponding rendezvous is possible;
3. \texttt{<statements-1>} or, if all alternatives are "closed", the else body \texttt{<statements>} is executed.

Bounded-buffer producer-consumer scheme is again used as an example, and the "Buffer" task is shown in Tab.7, while cyclic \texttt{Producer} and \texttt{Consumer} tasks are shown in Tab.8.

\begin{verbatim}
begin loop
   select when count < K =>
      accept store (item : in data) do B(last) := item end;
      last := (last mod K)+1;
      count := count+1
   or when count > 0 =>
      accept fetch (item : out data) do item := B(first) end;
      first := (first mod K)+1;
      count := count-1
   end select
end loop
end Buffer;
\end{verbatim}

Tab.7. Bounded-buffer task.

\begin{verbatim}
begin loop
   produce item;
   Consumer.fetch(item);
end loop
end Producer;
\end{verbatim}

Tab.8. Producer and consumer tasks.

A Petri net model of bounded-buffer task synchronization is shown in Fig.8; it looks rather simple when compared with Fig.7. The model is covered by P-inv., so it is bounded and guarantees mutual exclusion of its operations.

A more detailed description of intertask communication and synchronization, including aborts and exceptions during rendezvous, is given in [10].

6. Concluding remarks

It has been shown that inhibitor Petri nets can represent many different process synchronization concepts. Moreover, simple properties of nets (e.g., boundedness, absence of deadlocks) are very useful in verification of concurrent programs.

Similar net models can be derived for other concurrent programming constructs, such as path expressions [8], or CSP’s input and output commands [9].

Semaphores (basic as well as extended) provide a "low-level" synchronization mechanism, which is very flexible but also error prone, so it must be used in a rigorous, consistent way in complex applications. More structured constructs restrict all accesses to shared objects to clearly identified sections of code; they also provide mutual exclusion of shared data within such sections. Monitors protect access to shared data by integrating the data and operations performed on them within one structure; in fact, the shared data are
Petri Net Models of Process Synchronization Mechanisms

Fig. 8. Model of a producer–consumer bounded-buffer synchronization.

accessible only “through” corresponding operations. This may appear too inflexible in practical applications since nested monitors as well as mutual monitor operations (i.e., monitor procedures invoked from other monitors) must be used very cautiously because of possibility of deadlocks. Rendezvous concept which provides a synchronous one-way naming mechanism, is quite simple as a basic idea, however, it is considerably obscured by many “special cases”, exceptions, etc. [3, 10]. Therefore successors to existing concurrent programming concepts should probably return to attractive simplicity of semaphores, but also provide a finely grained selective access control and dynamic control of access rights that could be passed from one process to another, independently of the static structure of their definitions. The development of a comprehensive set of concurrent programming primitives must be based on a better understanding of the nature of concurrency, and this remains an important topic for further research. This further research, however, cannot simply ignore several decades of work in this area, work that contributed many important results [5].

Acknowledgements

The Natural Sciences and Engineering Research Council of Canada partially supported this research through Research Grant OGP-8222.

References


