Modeling of B-ISDN for Performance Simulation of ATM Services

Joaquim E. Neves*, and Mário J. Leitão**

* DEI, Universidade do Minho, Azurém, P-4800 Guimarães
** DEEC-FEUP, Universidade do Porto, Rua dos Bragas, P-4000 Porto

INESC - Instituto de Engenharia de Sistemas e Computadores
Largo Momplih 22, P-4000 Porto
Phone: + 351 2 2094000, Fax: + 351 2 318692
Email: eneves@inescn.pt, mleitao@inescn.pt

Abstract

The Broadband Integrated Services Digital Network (B-ISDN), using the Asynchronous Transfer Mode (ATM) has capabilities to support simultaneously many different traffic sources. This paper presents a model for the simulation of the B-ISDN transmission and switching, as well as the routing and flow control procedures. A model for ATM services characterized by Markovian space states will be used in the simulation of the traffic generation. The routing and flow control procedures incorporate the prediction of the traffic at B-ISDN interfaces by Neural Networks.

1 Introduction

The International Telephone and Telegraph Consulting Committee (CCTTT) has published recently a set of recommendations where the Broadband Integrated Services Digital Networks (B-ISDN) are defined [1], [2]. The protocol reference model (PRM) of the B-ISDN is structured in layers. The lowest layers of the PRM, namely the physical layer and the ATM layer are related to the transmission and are service independent. For the adaptation of the higher layer services to the ATM layer, the ATM Adaptation Layer (AAL) is also introduced and standardized.

The B-ISDN is a connection oriented network, although it may support both connectionless services and connection oriented services. The Asynchronous Transfer Mode (ATM) provides the transport and the switching of B-ISDN services in fixed size data packets called cells. The transport of the ATM services requires a call establishment phase and the adaptation of the information flow in cells [2]. A call is a concatenation of Virtual Path Connections
(VPC) and Virtual Channel Connections (VCC) within which the cells are transmitted [3]. The Asynchronous Transfer Mode (ATM) is independent of the services and the transmission system provided. ATM cells can be transported in the existent Plesiochronous (PDH) or Synchronous (SDH) Digital Hierarchies or in the new cell-based transmission systems [4], [5].

Signaling messages and the associated control functions are supported out of band in separate virtual circuits. At the establishment of a call, the user has to negotiate with B-ISDN control entity the traffic characteristics of the call and the quality of service requested. The network control entity can accept the request and allocate network resources for the support of the service, or propose a lower quality of service, or in the limit reject the call, if not enough network resources are available.

Two main B-ISDN operation scenarios are usually referred. One for residential areas where the audio-visual services are expected to be more required by network users and other for business areas where the data services will generate most of the ATM traffic. Generally the audio-visual services can tolerate more error rate but less delay variation than data services.

B-ISDN has mechanisms for usage parameter control by policing the call traffic at user interfaces and taking appropriate actions if the usage values of the information flow parameters are exceeded in a virtual channel or virtual path. The traffic parameters that can be negotiated and policed by the network are peak and average bit rate, peak duration and burstsiness [3].

The deployment of broadband ATM networks requires major efforts in studies and field trials, with the aid of simulation tools which contribute to provide a better view of the complex network mechanisms. This paper addresses the problem of B-ISDN modeling for the performance simulation of ATM services. In the next section, it will be focused the traffic generation at the user network interface by ATM services, and the call admission control and routing. In Section 3 we present a simulation model for the B-ISDN transmission and switching. Section 4 describes the performance simulation of a broadband network, and presents simulation results with different traffic generation scenarios and call admission control methods. Section 5 summarizes the conclusions.

2 B-ISDN Flow Control

The flow control functions in B-ISDN includes preventive and reactive actions. The connection admission control and the usage parameter control policing are examples of preventive actions, while the cell discharge and some other congestion control mechanisms are included in the reactive actions. The traffic characterization within the ATM layer is a pre-requisite for the control of the connection admission, usage parameter policing and cell discharge. Besides network methods, congestion control can also be made by the control of the cell generation in active calls, by higher layer end to end protocols.

2.1 ATM Traffic Characterization

Most ATM traffic models define the generation of ATM traffic by bursty sources where the parameters that characterize the sources are the same which are policed by the network. Those models only cope with some particular types of sources or the highly multiplexed traffic in the network node interface.
The ATM traffic model proposed in [6] has capabilities to characterize generally any single ATM traffic source or mixed sources. This model can be used in the ATM traffic generation at any network interface and is defined by three functional levels. In the generation level, a traffic source is specified by a Markovian state space and associated timing relations of the information flow events. The synchronization level incorporates the timing characteristics of the environment, which may include the limitations in the implementation of the generation devices. The adaptation level performs the low pass filter functions for the cell stream, to warrantee at the B-ISDN user network interface the peak cell rate and consequently the burstiness negotiated at the establishment of the call. The adaptation level acts as a finite length buffer for the cells provided by the generation level.

The generation level defines a set of timing relations of the information flow events for any traffic source. Since in B-ISDN, the information flow events can be connection requests or generation of ATM cells, each ATM traffic source will be actually defined by two Markovian state spaces. The mentioned ATM traffic model is stochastic, but it can generate deterministic traffic or deterministic components of stochastic traffic. The discipline imposed by the state transition probability matrix determines the time evolution of staying or leaving a traffic state. The quantum duration of each state and the probability distribution functions of the event duration and the time between events are the specific parameters of each traffic state. The quantum of the duration of each state is a deterministic parameter, but in general the time that the traffic source stays in each state is stochastic with a geometrical distribution. Independently of the probability distribution function of each traffic state, the number of expected call connections and the average cell rate can be easily calculated in the generation level of these traffic model.

2.2 Call admission Control

At the call establishment, the user has to negotiate with the control entity the traffic characteristics of the call and the quality of service requested. A call request of an ATM service is accepted if the B-ISDN has available resources for the associated end-to-end cell transfer.

The resource allocation for ATM call connections can be made by taking the peak or the average bit rate of ATM cells sources as reference. With allocation by the peak rate no statistic gain is obtained by multiplexing many sources, and the cell loss rate can be neglected because the buffers may be dimensioned with enough capacity to accommodate easily sporadic traffic fluctuations. If the resource allocation is made by the average bit rate of ATM cell sources the statistic gain obtained by multiplexing many sources is maximum but the simultaneous occurrences of the peak periods of some sources can drastically increase the cell loss rate.

The strategy for call admission control, introduced in [7] establishes a compromise between the maximum number of calls accepted and the satisfaction of the quality of services negotiated for calls established. The resource allocation criterion is based in a function, called Quality of Operation, which incorporates quality of service parameters such as cell loss rate, delay and delay variation.
2.2.1 Quality of Operation

The Quality of Operation [7] is a function associated to the network performance which can be quantified and applied both to the whole network or to some network parts. It incorporates the total amount of traffic allocated, as well as cell loss rate, delay and delay variation. The availability of network resources is quantified in the Quality of Operation function by the inclusion of components related to the free transmission capacity that can be allocated to each service. The equilibrium between the call rejection rate of different ATM services can be desirable, namely in high load situations, and this is also quantified in the proposed Quality of Operation function.

The parameters of the Quality of Operation can be determined during the operation of the B-ISDN in different traffic situations and operation scenarios. These parameters have to be adequate to the requirements of the predominant ATM services, which can be grouped by classes, according to bit rate, error rate, delay and other common characteristics.

When a request for resource allocation to a new call connections arrives to the control entity, the processors associated to every node and link of the path are inquired about the quality of operation expected with and without the new connection; the resources are then allocated to the call if the overall quality of operation is expected to be higher if the new connection is accepted. In case where a call has available alternative routes, every network node and link of every call path is inquired about the network quality of operation expected for the call. The decision can be based on a cost function of the routing algorithm, which includes the quality of operation expected in all nodes and links of the call path.

2.2.2 Traffic Prediction by Neural Networks

Neural networks are proposed in [8], [9], [10] to control the routing in spatial switches and in [11] three hierarchical levels of neural networks are referred to implement the control of ATM networks. Neural networks can also be applied in the policy functions, in the control of a selective cell discard, in alarms, traffic statistics and other OAM signals processing [12], [13] and in service coding, namely in video and audio compression [14], [15], [16].

In the call admission control we also use neural networks for the traffic prediction. Patterns of the traffic load in a node or link are memorized during the operation of the B-ISDN to be used as training patterns of a neural network. The training of the neural network can be performed in real time, or in batch mode with data collected in the actual broadband network or alternatively by network simulation in a computer.

When a request for resource allocation to a call arrives to a B-ISDN node, the bit rate requested and the maximum delay and error rate are declared. The node processor asks the neural network the expected traffic load for the actual load with and without the inclusion of the new connection. The neural network answers with the expected traffic patterns. The quality of operation is subsequently quantified in both cases to decide if the new connection is accepted as explained before.
3  B-ISDN Simulation Model

At the border between ATM and ATM adaptation layers, the effects in the cell stream of the physical and ATM layer are expressed by two parameters, the delay and the cell loss. The physical layer introduces cell loss due to error rate in the transmission links and its effects in the cell delineation mechanism, while the delay is proportional to the distance between the transmission end-points. The ATM layer introduces delay variation and cell loss, which are caused, respectively, by accumulation of cells in buffers and overflow in the switching nodes. Errors in the header which are not corrected by the protection mechanism may also increase the error rate (e.g., invalid virtual path or virtual circuit identifier).

The proposed simulation model attempts to incorporate those effects in its network components. These are:

Switching Node
Transmission Link

ATM traffic in the network components is expressed by two entities:

Cell
Call

3.1 Network Components

Switching nodes and transmission links are modeled by a buffer in which the cells are always read by the first in first out (FIFO) discipline. Each node or link (see figure 1) is characterized by:

Buffer Length
Throughput Capacity
Cell Loss Rate
Fixed Delay (Minimum Delay)

Within the switching nodes there is a route table which addresses, for each call, the outgoing link of the call path, while in each link, the route table addresses the destination node. All nodes and links can monitor the actual traffic load and its relevant statistics, namely the number of active calls of each service, the buffer occupation, the delay and the cell error rate. Specific nodes generating traffic incorporate a set of parameters to control the generation of the calls of each ATM service, according to the traffic model referred in the previous section.

The throughput capacity is the maximum number of cells that can be moved, by unit of time, between the input and output of any switching node or transmission link. The cell loss rate parameter incorporates the cell loss rate introduced by the physical layer and ATM layers, except the cell loss introduced by buffer overflow; it also includes the cell discharge resultant from policing of the usage parameter control. The cell loss is stochastic with uniform
distribution and average loss probability equal to the cell loss rate parameter. The transmission delay introduced by a link and the service time introduced by a switching node are specified by the fixed delay parameter. The fixed delay is also the minimum delay, and can be imposed in a network component by forcing the insertion of empty cells, when the node or link buffer reaches a minimum occupation.

Switching nodes can be connected, without restrictions, by transmission links, according to the desired network topology. Figure 2 shows the model of an example of one network with switching nodes connected by transmission links, in a combined tree and ring network topology.
3.2 Traffic Entities

The cell is an empty container generated within a node and destined to one or more nodes. The label attached to each cell includes:

- Call Identifier
- Sequence number (within the call)
- Generation Time
- Time of entry in the last node or link.

A negative value of the cell identifier is used to identify empty cells, introduced in the buffer to maintain the fixed delay in the switching node or transmission link. These cells have to be discharged before they arrive to the next network component.

For the call entity, the following parameters are required:

- ATM service identifier
- Call origin node
- Call destination node
- Call duration
- Sequence number of the last cell generated
- Sequence number of the last cell received

Each call invokes a service characterized by a set of parameters which define the cell generation according to the Markovian ATM traffic model [6]. Although our model has been established for connection oriented services, it can also accommodate connectionless services by associating each call to a given message identifier.

3.3 Simulation Procedures

The main network simulation procedures are call and cell generation, cell transmission and flow control (which include the call admission control and the call routing). The call and cell generation procedures are presented in condensed form in the flow charts of figure 3.

For the call generation, each node originating traffic has a Markovian process associated to each service in the node. The procedure scans sequentially each node n and each service s to determine if the next call has been generated \( t = T_n \) so that it can be processed by the call control and routing mechanisms to allocate resources. \( T_{in} \) denotes the time of the next call from service s at node n, and is calculated by the Markovian process.

In the cell generation procedure, all active calls are scanned every simulation cycle to determine if a cell is due for generation \( t = T_c \). When this occurs, the cell is inserted in the node which has generated the call c. If no cell is generated a check is made if the call is terminated, so that resources can be liberated. Timing and statistical reports can be obtained at the exit of the call generation procedure.

After the buffer in each node receives the cells generated by the calls originated in the node, the cell transmission phase takes place. Figure 4 shows a simplified flow chart of the
Figure 3. Call and Cell Generation.
Figure 4: Cell Transmission in B-ISDN.
cell transmission procedure. Cells that arrived from incoming links are first inserted in the nodes. Then, node by node, each cell at the head of the FIFO is inserted in the outgoing link pointed by the route table. Every time unit each node n or link l can receive and transmit at the most a number of cells equal to its throughput capacity. If a cell arrives and that maximum number is exceeded, or the buffer is full, the cell can optionally be lost or can be retained in the previous buffer. This allows to simulate the network in loss and delay modes, or in a combination of both.

Reports related with cell loss and delay are obtained in two occasions: after cells have been inserted in the nodes and after cells have been sent to the outgoing links.

## 4 Traffic Simulations

The proposed network and traffic model was used to produce real time simulations, with different call admission control criteria. Figure 5 sketches the main tasks of a computer program developed in C programming in an UNIX operating system machine.

![Diagram](image)

Figure 5: Performance Simulation of B-ISDN.

The reports dealing with variables are evaluated over short, medium and long time windows. Short window reports are evaluated every simulation cycle while medium window reports are evaluated over a fixed period, specified in the simulation parameters. Long window reports are obtained at the end of the simulation period and express the overall statistics of the simulation processes. The number of active calls from each service, the cell generation time and delay within each call, and the buffer occupation of each network component are examples of short window reports. The average cell rate and the average delay for each call are medium window reports, while the histograms of the delay and buffer occupation are included in the statistical type of reports. Figure 6 shows examples of the three types of reports obtained with a service characterized by two Markovian states with duration geometrically distributed, with probability of staying and leaving each state of 20% and 80%, respectively. The cell rate in the activity state is 33.33 Kcell/s and the duration is an integer multiple of 0.002 seconds. The cell generation is interrupted during a time multiple of 0.002 seconds (silence state) when the service leaves the activity state.

Table 1 presents the main characteristics (average and the peak cell rate and the burstiness factor) of three services ($SC_0$, $SC_1$ and $SC_2$) used in the next simulations.
Figure 6: Example of short window (cell generation time), medium widow (Cell rate), and Statistical (Histogram of the time between cells) Reports.

Table 1: Main characteristics of the services $SC_0$, $SC_1$ and $SC_2$.

<table>
<thead>
<tr>
<th>Service Class (SC)</th>
<th>Cell Rate (KCell/s)</th>
<th>Burst (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Peak</td>
</tr>
<tr>
<td>SC_0</td>
<td>1.604</td>
<td>10.000</td>
</tr>
<tr>
<td>SC_1</td>
<td>3.750</td>
<td>5.000</td>
</tr>
<tr>
<td>SC_2</td>
<td>20.000</td>
<td>20.000</td>
</tr>
</tbody>
</table>

Figure 7 presents simulation results of a node loaded by calls from services $SC_0$, $SC_1$ and $SC_2$. The figure shows the allocated bandwidth with allocation by the average cell rate, by the peak cell rate and by the new call control technique based on the quality of operation. It may be seen that, with the allocation by the average and the peak cell rate, only the narrowest band service class (narrowest average and narrowest peak, respectively) can access the network resources, namely during the significantly loaded periods. With the proposed technique, the figure shows that all the service classes can share the available resources even when demand is higher.

The following simulation was carried out with a network which topology is depicted in figure 8. The figure also shows the services generated in each node: all nodes generate traffic towards
node 4, node 0 generates traffic from the three service classes, node 1 generates only service classes 1 and 2, while node 2 generates classes 0 and 2, and node 3 generates only classes 0 and 1.

<table>
<thead>
<tr>
<th>Simulated Network Topology</th>
<th>Traffic Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Class</td>
<td>Network Node</td>
</tr>
<tr>
<td>SC-0</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>SC-1</td>
<td>x x x</td>
</tr>
<tr>
<td>SC-2</td>
<td>x x x</td>
</tr>
</tbody>
</table>

Figure 8: Topology of the simulated network

Figure 9 presents results of a simulation in this network, of the call control and routing technique based on the quality of operation, showing the average allocated bandwidth during 25 seconds of simulation.

Figure 9: Allocated bit rates in different links

The results exhibit a good balance between the usage of links confirming that the quality of operation, with suitable values of the control parameters for routing purposes, has capabilities to find a suitable route for the calls.
5 Summary

A model for the simulation of the main B-ISDN components (transmission links, switching nodes), traffic entities and procedures has been presented. Transmission links and switching nodes are modeled by a FIFO buffer characterized by delay, error rate, throughput and buffer length. Traffic entities are cells and calls and generation of traffic is modeled by Markovian processes. The call generation associated to different services and users is a Markovian process with variable state duration. Within each state, call duration and time between call births are defined by suitable distribution functions. The cells of each call are also generated by a similar traffic model with the appropriate parameters for each service. The simulation procedures for call generation, cell generation and cell transmission have also been described. Results have been presented for the simulation of traffic in a network using a technique for call control and routing based in a quality of service function.

References


