

Algorithm Development for Measurement and Prediction the Traffic in Broad Band Integrated Service Networks

Mohammad Al Rawajbeh, Ibrahim Mansi and Majed Hussien
Department of Management Information Systems,
Faculty of Administrative and Financial Sciences,
Al Isra University 11622, P.O. Box 22, 23, Jordan

Abstract: Problem statement: In this study an effort had been made to develop an algorithm for traffic prediction and measurement for a wide use networks such as B-ISDN. Such algorithm should base on valuable parameters during the various stages of data elements transmission. The offered algorithm was expected to enforce the main sources of congestion problem in network. With this technique the admission decision is made according to the prediction of a few quality of service parameters expected for the new connections. **Approach:** This research aimed to find out the suitable method for improving the performance and the quality of service during the real time work in wide used networks now days. The improvement of quality of services in B-ISDN can be achieved by a significant estimation of the network state and discovering the most critical situation during the work time. The most repeated problems in such networks are loses of connection, delay time and saturation of communication lines. These problems are known and some solution could be sufficient to deal with, but not for a long time. **Results:** The proposed solution was based on the need of traffic prediction method in real time to determine the state of network and decide how to deal with. Such method should base on finding the most significant parameters in network during the real time work and use them as prediction variable to predict the situation or the state of network in the future time and then take the appropriate action. Suffering from permanent problems finding the most significant parameters, which can be estimated at the real time to help for solving a raw problems which encountered during the work in networks such as saturation, bottleneck, disconnecting and time delay. **Conclusion:** The results which achieved by this research was based on the developed algorithm which can be used to predict the traffic in B-ISDN, optimizing the bandwidth and making the bandwidth available to the behaving sources under congestion situation and also when there is no congestion. Meet the QoS demands of the network traffic during congestion situation and also when there is no congestion. Reject/drop all the packets. First guideline for future study in this direction is investigation additional objective functions, such as bandwidth limitation for archive more general results.

Key words: B-ISDN management, traffic prediction, measurement algorithm, quality of services

INTRODUCTION

Many flow control mechanisms require the prediction of traffic statistics, in order to provide faster and more accurate control actions to avoid the effects of excessive load situations. The knowledge of ATM traffic profiles inside the network is a pre-requisite for the connection admission control. Since the acceptance of a connection should guarantee the quality of service negotiated, without affecting already established connections, traffic prediction is an essential step that must be supported by a suitable model for B-ISDN components (Allman *et al.*, 1999).

The prediction of traffic parameters within the switching nodes and transmission links can be made using analytical models for the traffic sources and B-ISDN components, or by processing samples of the traffic parameters observed on these components. Since the prediction by analytical techniques is usually based on convolution operations, which become very complex for realistic traffic and network models, it is preferable to perform the prediction by processing the traffic parameters (Neves *et al.*, 1997).

There are two main characteristics of the traffic in B-ISDN-a unit of data and the method of forging these units. The unit of data can be: bits, byte, octets,

Corresponding Author: Mohammad Al Rawajbeh, Department of Management Information Systems,
Faculty of Administrative and Financial Sciences, Al Isra University 11622, P.O. Box 22, 23, Jordan

transaction and block. They are packed in files, packets, frames and cells. They can also be transmitted in bulk (without packaging). In terms of ATM a packed data is given Protocol Data Unit (PDU) (Stallings, 2002).

Speed is measured in units of data per unit time, for example, packet sec^{-1} , bytes sec^{-1} , transactions min^{-1} . Speed also determines the time required for transferring units of data over the network (Kolbanev and Jakovlev, 2004).

The actual amount of transmitted data over the network consists of real data and necessary frame information which form an overhead to the transmission. Many technologies set constraints on the minimum and maximum size of packet.

As mentioned above, there are four most common characteristics of traffic:

- Throughput
Tolerance to delays
- Response time
- The capacity and bandwidth

These characteristics, taking into account the routing, priorities, connections and other issues, at the same time determine the nature of network functionality.

Throughput: It is the rate at which data flow past some measurement point in the network. It can be measured in bits sec^{-1} , bytes sec^{-1} or packets sec^{-1} . Throughput is measured by counting the traffic over an interval and a care must be taken to choose this interval appropriately. A long interval leads to averaging out transient bursts and lulls in the traffic. A shorter interval will record these temporary effects, Even if they are not important in the context of the measurement.

Tolerance: Tolerance is the delay is the reaction of applications in all kinds of delays in the network. For example, the application which execute the financial transaction in real time, hasn't any delay. Large delays can cause a malfunction of such applications.

Applications vary greatly in their allowable delay time. There is an application working in real time (video conference)-the delay time in this case must be very small. On the other hand, there are applications with high tolerant to delays of several minutes or even hours (e-mail and downloading files).

The capacity of the network: It is the actual amount of resources available for the user in a certain way of data transmission. Network bandwidth determined by the total amount of data that can be transmitted per unit of

time. The capacity of the network differs from the network bandwidth due to the presence of overhead costs, which depend on how to use the network.

Network capacity: Capacity depends on the size of transfer window, delay and the speed of data transferring. We use the following designations: W-the size of transfer window in bytes; R-Speed of data transfer (bps) for a specific connection by sender; D-delay (sec) when sending data between the sender and the recipient through a connection (Wei, 2004).

Problem statements: Suppose that sender starts transmission of sequence bytes to receiver through a connection. In order for the first byte has reached the recipient, will take time D.

The same time-D seconds will be required to obtain confirmation. During this time the sender can transmit $2RD$ bit, or $RD/4$ bytes. In fact, sender is restricted the size of the window in W bytes and can not move the window until receive confirmation. Only when $W \geq RD/4$ in this connection, the maximum possible throughput can be achieved.

If $W < RD/4$, the proximity to the maximum capacity is determined by the ratio of W to $RD/4$. Consequently, the normalized throughput S can be expressed as:

$$S = \begin{cases} 1, & W \geq RD/4 \\ 4W / RD, & W < RD/4 \end{cases} \quad (1)$$

The maximum window size can be $2^{16}-1 = 65\,535$ bytes. This window size should be sufficient for most applications. Enough to enlarge the window scaling option up to 4-this will lead to a significant increase in the size of the window.

The above parameters have a major impact on the efficiency of transmission protocol TCP. However, there are many critical factors that should also be taken into account.

First, in most cases, TCP connections multiplexed in one channel, so that each connection receives a portion of its available bandwidth. This leads to a reduction in the speed of transmission R and, consequently, to reduce the effectiveness of protocol. The reason TCP enjoys this importance is that it constantly evolves to keep up with the changing network demands (Bhandarkar *et al.*, 2006).

Secondly, most TCP connections are passing through routers. In this case, time D will be equal to the amount of delay in each network and the delays at each router in the path. Thus the total delay at routers often makes the main contribution to the delay time D,

especially in overload cases (Katabi *et al.*, 2002; Kelly, 2003).

In third, the value of speed R , used in the above formula, determines the speed of data transmission has which available for connections only on the side of sender. If one of the transitions in the path from the sender to the recipient's transfer speed is less than this speed, then attempt to transfer at the maximum speed will lead to bottleneck formation, which will inevitably increase the time of D . Finally, if the a segment is lost, it must be referred again, which leads to reduction of throughput. The relative impact of segments loss on the effectiveness of transmission efficiency depends on the repeating policy (Griffiths, 1998; Kylgin, 2000).

In today's distributed networks, several segments of TCP protocol may be loss due to errors on lines.

Most segments are lost when using mechanisms of packets reset in routers or switches (e.g., frame relay) in case of network congestion (Mahdavi and Floyd, 1997). The task of overload control can be assigned to the protocol RSVP (a resource reservation setup protocol that can be used by a host to request specific QoS for multicast multimedia flows on the Internet (Guerin and Peris, 1999)).

Solution:

Defining the parameters of the system in bursts:

Data elements in the network transmitted with some average rate of transfer λ (measured in elements per sec). At a certain moment a certain number of data elements will be in queue. Denote the average number of data elements in the queue ω and the average time that the data elements must wait in the queue-a symbol of $T\omega$. This parameter is defined for all incoming data elements, including those who were not wait at all. Server processes the incoming data elements, spending on this a time in average T_s . This time interval is counted from receiving moment of data element to the server until the processing of the server (i.e., sending). Server utilization (load level) ρ -is the proportion of the total time during which the server was busy (Jin and Tierney, 2003).

In addition, there are two parameters that characterize the system as a whole. These include: The average number of data elements that are in the entire system, including data elements, which have begun to treat and the elements which waiting for processing q and the average time that these data elements are in the system, waiting for their turn or are already in the processing- T_q .

The average rate of departure data elements is equal to the average rate of the proceeds. If the rate of proceeds data elements (which is determined by traffic

entering in system) increases, it increases the load on server and therefore the utilization of server. Increase the queue size increases (or at least, may increase, not reduce) the waiting time of the data elements in it. When $\rho = 1$ the server is loaded to the limit, working 100% of his time. Consequently, the theoretical maximum speed of receipt of data elements in which they can be processed by the server using the following formula:

$$\lambda_{\max} = 1/T_s \quad (2)$$

However, the queue size increases dramatically with the entry system in the saturation regime, tends to infinity, when $\rho = 1$. In practice, therefore, tend to restrict the speed of data on server to 70-90% of the theoretical maximum.

Before any computation for the parameters of system with a queue, it is required to define the conditions and identify the change interval of this system. Below are the conditions under which we would be considered a system with queues.

Determination of the number of data elements: It is anticipated that to the system may entered an infinite number of data elements. This can be interpreted as a requirement that the speed of the proceeds data elements in system does not depend on the number of elements, located in the system.

A queue can grow indefinitely: When considering a system, assumed an infinite size of the queue. In practice, any particular queue has a limit, but in most cases, the theoretical assumption of infinite capacity does not lead to significant errors, as well as real devices use different mechanisms to prevent situations where the data will be discarding.

Determined order of data elements service: When the server becomes idle and in queue there is a few data elements, it is required to define the rules under which the particular data element, chosen by the server for processing. The simplest rule is First In, First Out-first come, first left (FIFO). Another possible rule of service queue can be Last In, First Out-first come, last went (LIFO). Moreover, in practice have to deal with the order of service, which based on service time. Unfortunately, the service queue is based on time parameters, which are difficult for modeling. A more general case of a service queue based on priorities, which will be considered further. The flow of packets with the same priority is considered (Kolbanev and Jakovlev, 2004; Kylgin, 2000; Braden *et al.*, 1998). In Table 1 lists all the parameters that are intended to measure and predict traffic.

Table 1: Traffic measurement and prediction parameters

Symbol	Description
Λ	The average speed of the proceeds data elements in system (number of elements per sec)
T_s	The average time of service received by the elements (sec)
σT_s	Standard deviation of element service time element (sec)
P	Server utilization in service (the percentage of time when server is busy)
U	Traffic intensity
Q	Total number of data elements in system
q	The average number of data elements in system
T_Q	The time which data elements spent in system (sec)
T_q	The average time that data elements spend in system (sec)
σ_q	The standard deviation of q
σT_q	Standard deviation T_q (sec)
Ω	The average number of data elements in the queue awaiting service (queue size)
T_o	The average time that data elements are awaiting services (sec)
T_d	The average waiting time for service data elements that were in the queue (i.e., not including items for which the waiting time is 0)
σ_o	The standard deviation of ω
N	Number of servers
$m_x(r)$	X less than or equal to mx (r) at r per cent of cases

In case of a system has many servers that share a common queue, if there is in time at least one server is idle, data elements immediately sent to this server. It is assumed that in the system all servers are identical. This means that if multiple servers are available, it does not draw any distinction between the servers to select the one that will process the next data element. We can say that the likelihood of data elements for the service on different servers are the same. If all servers are busy, the queue begins to form. One queue for all servers. With the release of one of servers, queue provides data element, selected in accordance with established order.

In case of multiple identical servers, selection a specific server to service a particular data element does not affect the time of service.

The main task of queues analysis is obtaining the information about: the speed of data element reception in the queue, service time of these elements on the server on system entrance, the total number of waiting elements; timeout element in the system. It is important to know the average values of these parameters and their range of changes. Therefore, great importance in the analysis of queues has knowing of the standard (mean) deviation of each of these parameters (Shorten and Leith, 2007).

For the analysis of systems or individual modules of network devices may be useful using other indicators: for example, calculation the size of buffer memory for the bridge or multiplexer.

For that we have to know the law of reception change of data elements in the system and the law of service time distribution of data elements by the server. It should be noted that, it is very difficult to get the name of data by elementary methods, as well as the original formula for the calculation are quite complex. In order to simplify the calculations, it is necessary to make some natural approval. The most important of these assumptions is that the change in income data elements speed obeys the law of poison. For the applicability of the poison law requires the following hypotheses:

- Reception of a data element is not dependent on another element, events occur independently
- Never received two or more data elements at the same time
- The average amount of income does not change with time (statically distribution)

Under these conditions, the likelihood of the data element reception is a subject to the Poisson law, which is described by the following formula:

$$P_n(t) = e^{-\lambda t} (\lambda t)^n / n! \tag{3}$$

Where:

- λ = The speed of income data elements (Yuan and Liu, 2001; Kylgin, 2000)
- n = Number of entries received elements during the time t

The elements service continuous on the server usually describes by other several laws. Most often used in this case is the law of intervals or exponential law. To this law used a large number of continuous service time. If the intervals of time separating the events, put on a straight flush, we get a number of events, satisfying the Poisson law. For that, a three listed hypothesis remains true to the time of service. However, intervals of the events may not be close, as possible instances of server downtime. In this case, exponential law of service time distribution is used. In addition, to describe the law of income elements distribution and their service we can use a normal distribution law.

To estimate the system, the parameter which characterizes the intensity of the system must be determined (Kylgin, 2000).

Thus, an algorithm of measurement and predict the traffic in B-ISDN, can be represented as shown in Fig. 1.

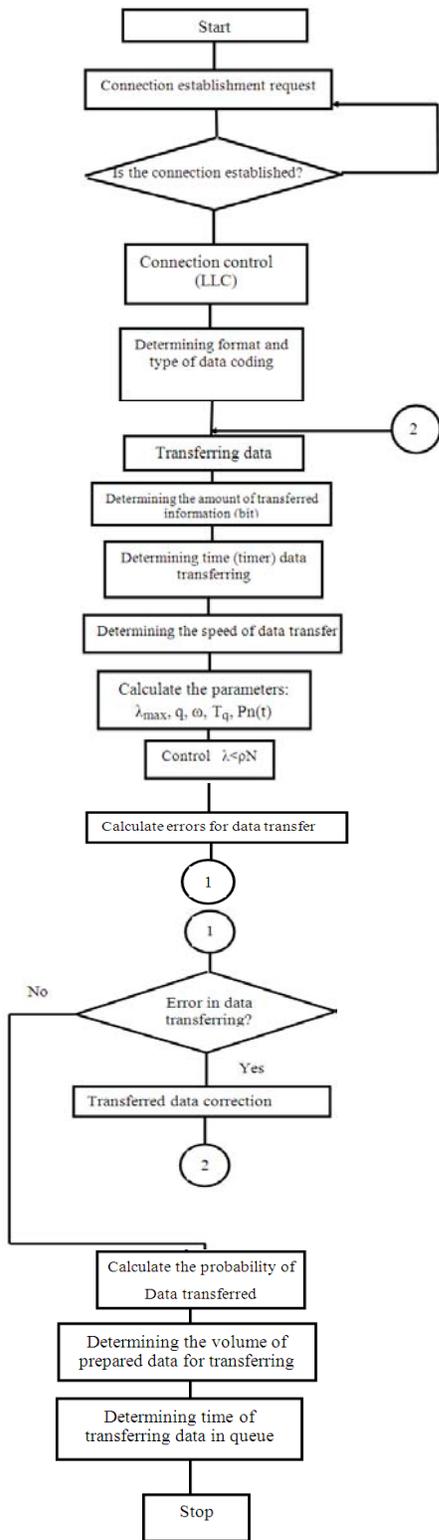


Fig. 1: Algorithm of measurement and predict the traffic in B-ISDN

MATERIALS AND METHODS

Prediction of traffic parameters within the switching nodes and transmission links can be made using analytical models for the traffic sources and B-ISDN components, or by processing samples of the traffic parameters observed on these components. Since the prediction by analytical techniques is usually based on convolution operations, which become very complex for realistic traffic and network models, it is preferable to perform the prediction by processing the traffic parameters. Also, the main task of queues analysis is considered for obtaining the information about: The speed of data element reception in the queue, service time of these elements on the server on system entrance, the total number of waiting elements; timeout element in the system.

RESULTS AND DISCUSSION

The main object of this research was met by determining the most important parameters which can be considered for evaluating the various system condition and prediction the traffic in network. On the base of these parameters the most significant problem is resolved by development the needed algorithm which will be applied to optimize the bandwidth.

The analytical method takes on consideration the queuing theory which considered as an important factor for estimation the condition of traffic, also the last reviews showed that the most suitable way to make prediction for the traffic is based on the processing the parameters as done in this research.

The main findings in this research are: optimizing the bandwidth, meeting the QoS demands of the network traffic during congestion situation and also when there is no congestion. Maximize the throughput for the behaving sources under congestion situation and also when there is no congestion, development prediction algorithm for network traffic.

Recommendations: The authors recommend motivating the future studies in the same direction for achieving more significant results, implementing this method in more realistic simulation methods and making a stress on the finding additional objective functions such as bandwidth limitations.

CONCLUSION

This study has discussed the selection and assigning important parameters for measurement and prediction of traffic in B-ISDN. An algorithm on base

these parameters has been developed to measure the traffic and peek situation in network during data transmission in various conditions. The expected result from this algorithm when it applied for traffic prediction are: optimizing the bandwidth and making the bandwidth available to the behaving sources under congestion situation and also when there is no congestion. Maximize the throughput for the behaving sources under congestion situation and also when there is no congestion. Meet the QoS demands of the network traffic during congestion situation and also when there is no congestion. Reject/drop all the packets.

First guideline for future study in this direction is investigation additional objective functions, e.g. bandwidth limitation for archive more general results. Moreover implementing this method in a more realistic simulation environment, motivate the next studies.

REFERENCES

- Allman, M., V. Paxson and W. Stevens, 1999. TCP congestion control. IETF. <http://tools.ietf.org/html/rfc2581>
- Bhandarkar, S., A.L.N. Reddy, M. Allman and E. Blanton, 2006. RFC4653: Improving the Robustness of TCP to non-congestion eve. FAQs. <http://www.faqs.org/rfcs/rfc4653.html>
- Braden, B., D. Clark, J. Crowcroft, B. Davie and S. Deering *et al.*, 1998. RFC2309: Recommendations on queue management and congestion avoidance in the internet. ACM Press. <http://portal.acm.org/citation.cfm?id=RFC2309>
- Griffiths, J.M., 1998. ISDN Explained Worldwide Network and Applications Technology. 3rd Edn., Wiley, ISBN: 13: 978-0471979050, pp: 306.
- Guerin, R. and V. Peris, 1999. Quality-of-service in packet networks: Basic mechanisms and directions. *Comput. Networks*, 31: 169-189.
- Jin, G. and B. Tierney, 2003. Netest: An tool to measure maximum burst size, available bandwidth and achievable throughput. Proceedings of the ITRE, Aug. 10-13, Lawrence Berkeley National Laboratory, Newark, NJ., pp: 1-5.
- Katabi, D. and M. Handley and C. Rohrs, 2002. Congestion control for high bandwidth-delay product networks. *ACM SIGCOMM Comput. Commun. Rev.*, 32: 89-102.
- Kelly, T., 2003. Scalable TCP: Improving performance in highspeed wide area networks. *ACM SIGCOMM Comput. Commun. Rev.*, 33: 83-91.
- Kolbanev, A.M. and S.A. Jakovlev, 2004. Evolution of services of communication networks. *J. Inform. Oper. Syst.*, 4: 36-41.
- Kylgin, M., 2000. Technology of Corporate Networks. 1st Edn., USA., pp: 609-614.
- Mahdavi, J. and S. Floyd, 1997. TCP-friendly unicast rate-based flow control. PSC. http://www.psc.edu/networking/papers/tcp_friendly.html
- Neves, J.E., M.J. Leita0 and L.B. Almeida, 1997. Neural networks in B-ISDN flow control: ATM traffic prediction or network modeling. *IEEE Commun. Mag.*, 33: 50-56. DOI: 10.1109/ISCC.1997.615971
- Shorten, R.N. and D.J. Leith, 2007. On queue provisioning, network efficiency and the transmission control protocol. *IEEE/ACM Trans. Network.*, 15: 866-877.
- Stallings, W., 2002. High-Speed Networks and Internets Performance and Quality of Service. 2nd Edn., Prentice-Hall, USA., ISBN: 10: 0130322210, pp: 715.
- Wei, D.X., 2004. Congestion control algorithms for high speed long distance TCP connections. Master Thesis, California Institute of Technology.
- Yuan, X. and X. Liu, 2001. Heuristic algorithms for multi-constrained quality of service routing. *Proc. Ann. Joint Conf. IEEE Comput. Commun. Soc.*, 2: 844-854.