A PRACTICAL APPROACH FOR AUTOMATIC GENERATION OF NETWORK SEGMENT TRAFFIC BASELINES[1]

Mario Lemes Proença Jr., Fabio Sakuray, Camiel Coppelmans, Mauricio Bottoli, Antonio M. Alberti, Leonardo de S. Mendes

Resumo - Este artigo apresenta um modelo para geração automática de baseline que visa a caracterização do tráfego em segmentos de rede. A utilização do baseline permite ao administrador: identificar limitações e pontos críticos da rede; determinar a real utilização dos recursos de rede; controlar melhor o uso dos recursos e o estabelecimento de limites para geração de alarmes mais precisos e inteligentes que se adaptam as reais características da rede. É apresentado também um sistema de alarmes que utiliza o baseline e proporciona a automação de uma tarefa realizada de forma manual pelo gerente da rede com base em seus conhecimentos empíricos. Para o desenvolvimento, implementação e testes dessas funções foi utilizada uma ferramenta chamada GBA. Além disso, serão apresentados resultados com a utilização prática do baseline e do sistema de alarmes no gerenciamento de segmentos da rede. Os resultados obtidos validam o experimento e demonstram na prática ganhos expressivos no gerenciamento de redes.

Palavras-chave: Gerenciamento de Rede, baseline, alarmes, caracterização de tráfego, monitoramento de redes.

Abstract - This paper presents a model for automatic generation of a baseline which characterizes the traffic of network segments. The use of the baseline concept allows the manager to: identify limitations and crucial points of the network; learn about the actual status of use of the network resources; be able to gain better control of the use of network resources and to establish thresholds for the generation of more accurate and intelligent alarms, better suited to the actual characteristics of the network. Also presented is an alarm system that relies on the baseline and that provides the automation of a task that can be performed manually by the network manager, based on his empirical knowledge of the network. A tool known as GBA was used for the development, implementation and testing of these functions. Moreover, some results obtained with the practical use of the baseline as well as of the alarm system in the management of network segments, are also presented. The results obtained validate the experiment and show, in practice, significant advantages in their use for network management.

Keywords: Network management, baseline, alarms, traffic characterization, network monitoring.

1. INTRODUCTION

The traffic that flows across the Internet nowadays is characterized mainly by services like data, voice and image. New services and media are being continuously developed in order to enable better interaction and communication among network users, such as Voice over IP, video on demand and on-line games; consequently, there is a growing need for effective traffic management and implementation of service quality (QoS) in connections that are latency-sensitive such as video and voice [1][2].

Extensive work has been done and discussions have been held to improve ways to implement quality of services and traffic management that consist in accomplishing traffic control and traffic engineering along the Internet backbone [3]. Technologies such as architecture of integrated services (Intserv), architecture of differentiated services (Diffserv) and the Multi-Protocol Label Switching (MPLS) were implemented and are being tested with the purpose of offering QoS along the Internet backbone [2][4][5].

Due to the increase of the computer network and of the Internet, the search for the QoS between the network connections and the use of new services, becomes the management of the network resources most important focus with the intention to archive a better optimization and utilization of these resources.

Several existing tools and network management systems (NMS) aim at helping with the management of the five areas defined by ISO as fundamental for the network management [6][7]: Fault, Configuration, Accounting, Security and Performance. However, the construction of baseline suitable for the characteristics of each segment that make up the network backbone is an important task that is not usually found in the network management systems. The Baseline can be defined as a set of basic information that describes the traffic profile in a network segment through network thresholds about volume of traffic, number of errors, types of protocols and services that flow through this segment during the day [8].

The rest of this paper is organized as follows. Section 2 presents works related to the proposal in this paper. Section 3 presents a description about the advantages and motivations that lead to the construction of the baseline which we also refer to as digital signature of network segments (DSNS). Section 4 discusses the model used for the construction of the baseline. Section 5 presents how the model is validated and results that show practical gains for the network management. At last in section 6 we conclude and mention future works to be presented.

Mario L. Proença Jr. and F. Sakuray are professors of computer science department of state university of Londrina, PR, Brazil. C. Coppelmans and M. Bottoli are students of DECOM-FEEC-UNICAMP, SP, Brazil. A. Alberti is professor of INATEL, Santa Rita, Brazil. Leonardo S. Mendes is Professor of Communication department of FEEC-UNICAMP, SP, Brazil. E-mails: proenca@uel.br, sakuray@uel.br, camiel@uel.br, mauricio.bottoli@igniscom.com.br, alberti@inatel.br, lmendes@decom.fee.unicamp.br.

2. RELATED WORK

There is a lot of work done in traffic characterization and traffic measurement that is related to the proposal in this work [9][10][11]. Traffic characterization and traffic measurement are important aspects that have to be considered for network management and control. In [9][10] is presented a survey of the main research done for traffic characterization in telecommunication networks. However these models intend to traffic modeling in a generic way, while the proposal presented in this paper intends to a traffic characterization generated from collected real data of each segment of analyzed network. The aim of this characterization is to create a particular profile for each monitored segment, which we call a baseline or digital signature of the network segments (DSNS).

In [12] traffic modeling of a sub-network using ARIMA is proposed with intend to capture the characteristics of the internet traffic of a sub-network that can be used for analysis of internet performance. In [13] is presented a proposal that is close to ours presented in this work, they proposed a baseline for automatic detection of network anomalies that uses asymptotic distribution of the difference between successive estimates of a model of network traffic. One problem that exists in this model is that it assumes that the training data is pure with no anomalies. In our case we calculate the baseline based on real data gathering from the network segment. Our baseline is generated based on statistical analyses of these data.

Another important area that is related to work presented in this paper is anomaly detection [14][15][16]. Thottan et al [14], presents a review about anomaly detection methods and a statistical signal processing technique based on abrupt change detection that uses analysis of SNMP MIB variables for anomaly detection. In [14] is used a 15s sampling frequency, and it assumes, like an open issue, that there exist some changes in MIB data that don't correspond to network anomalies. The use of an effective and real baseline can help to solve this problem for knowing the real behavior of the traffic.

Papavassiliou et al. [15], presents a tool with intend to facilitate the network management, reducing costs and minimizing the human errors. They use a similar approach to ours for the construction of baselines, when they separate workdays from weekends.

Recently, Krishnamurthy et al. [16] presented a sketch-based change detection for traffic anomalies, They used a variety of time series forecast models. The efficiency of this technique still doesn't prove operations in real time mode and depends on the accuracy of the periodically recomputing of the forecast model parameter. Our objective is the construction of a simple mechanism that works in real time, and helps a network manager in taking decisions efficiently and reliably.

3. BASELINE USES

The baseline or the digital signature of network segment (DSNS) can be defined as the set of basic information that shows the traffic profile in a network segment through minimum and maximum thresholds that indicate which would be the normal behavior of this segment along the day.

The forecast of a determined instant, about the characteristics of the traffic of the segments that make up the network backbone, make the management decisions on anomalies that might be happening, more reliable and safer [14][15].

The use of the DSNS can help the network manager to identify limitations and control the use of resources that are critical for services that are latency-sensitive such as Voice over IP and video transport, because they can't support retransmission or even network congestion. Besides improving the resources control, its use also facilitates the capacity planning on the network, because it clearly identifies the real use of resources and the critical points along the backbone, avoiding problems of performance and fault that might happen.

The use of the DSNS also offers the network manager advantages related to performance management, by means of the previous knowledge of the maximum and minimum quantities of traffic in the segment along the day. This enables the establishment of more effective and functional alarms and controls, because they are using thresholds that suit the DSNS, respecting the variations of traffic along the day instead of using the linear thresholds that are set based on the expertise of the human network manager [13][17]. Deviations in relation to what are being monitored real-time and what the DSNS expresses must be observed and analyzed carefully, and can or can not be considered as problems. In order to do that, the use of an alarm system integrated to the DSNS and to the real-time monitoring will deal with these problems, warning the network manager when it is necessary.

As for security management, the use of the DSNS can offer information related to the analysis of the users behavior, because the previous knowledge of the behavior and the traffic characteristics of a determined segment is directly related to the profile users manipulation, using this as information to prevent intrusion aspects or even network attacks, by means of the intrusion detection software [18][19][20].

Another application for the DSNS is related to the monitoring of a network segment which is normally performed manually by means of visual control, based on empirical knowledge with the network acquired by the manager. An example of this can be seen with the utilization of tools like GBA (Automatic Backbone Management) [21] and MRTG (Multi Router Traffic Grapher) [22] that generate graphs with statistical analysis which consist of averages along a determined period of time about an analyzed segment or object. However, the simple use of these graphs establishes limitations for the network manager concerning discovery and solution of problems. The limitations are caused especially by the non-automation of this task, where the monitoring of these graphs is performed visually, depending exclusively on the empirical knowledge about the functioning of the network acquired by the manager and due to the large quantity of graphs that have to be analyzed continuously. It allows the detection of the problems and unusual situations in a reactive way.

Networks with a large number of segments turn their management more complex, considering the great quantity of graphs to be analyzed [15]. The graphs usually present information on the volume of input and output of traffic of a certain segment, not aggregating information that could help the manager more efficiently in his decision-making with the purpose of solving problems that might be happening or that might have already happened.

4. BASELINE IMPLEMENTATION

The main purpose to be achieved with the construction of the DSNS is the characterization of the traffic of the segment it refers to. This characterization should reflect initially the profile expected for the traffic along the day as well as other existing characteristics such as: types of protocols, types of applications, types of services. These characteristics are used to create a profile of the users. The DSNS was initially developed to analyze the quantity of input and output of octets stored in the ifInOctets and ifOutOctets objects which belong to the Interface group of the MIB-II [23].

The use of the GBA tool (Automatic Backbone Management) was chosen as a platform for the development of the baseline due to the great quantity of historical information related to monitoring carried out along the last years in the main network segments of UEL. The GBA was initially developed to help with management of ATM backbone and it performed its duty as it became a platform of learning and development, helping with the management as well as with the understanding about the networks functioning. Further information on the GBA can be found at http://proenca.uel.br/gba or in [24].

As for the tests and validation of the model, the data gathered by the GBA have been used since 2002 up to the present. The use of the data from the last two years was considered an important sample, characterized by periods of winter and summer vacations as well as holidays which contributed to the tests and validations of the ideas presented in this work. The analyzed data is related to the network segments with traffic TCP/IP based on Ethernet and ATM with LAN Emulation. The tests of the proposed model were carried out in three segments, which are described below:

1. The first one which is called segment S_1 is responsible for interconnecting the ATM router to the other backbone segments of State University of Londrina (UEL) networks; it gathers a traffic of approximately 2500 computers;
2. The second one which is called S_2 interconnects the office for undergraduate studies of academic affairs in UEL; it gathers a traffic of 50 computers;
3. The third one which is called S_3 interconnects State University of Campinas (UNICAMP) network to academic network at São Paulo (ANSP); it gathers a traffic of all UNICAMP (about 5000 computers) to Internet.

For the generation of the DSNS, a model was developed based on statistical analyses that we call BLGBA. The analyses were carried out for each second of the day, each day of the week. Figure 4.1 illustrates the operational diagram used in the implementation of the DSNS, which is carried out by the GBA generated Baseline/DSNS module. This module reads information from the database, with data gathered daily from GBA collect samples from MIB module, and generates the DSNS based on a period requested by the network manager.

Two types of DSNS were created, one called ***bl-7*** which consists of seven DSNS files, one for each day of the week, and the other one called ***bl-3*** which consists of three DSNS files, one for the workdays from Monday to Friday, one for Saturday and another one for Sunday, as shown in Figure 4.1. The choice for generating the DSNS separating the

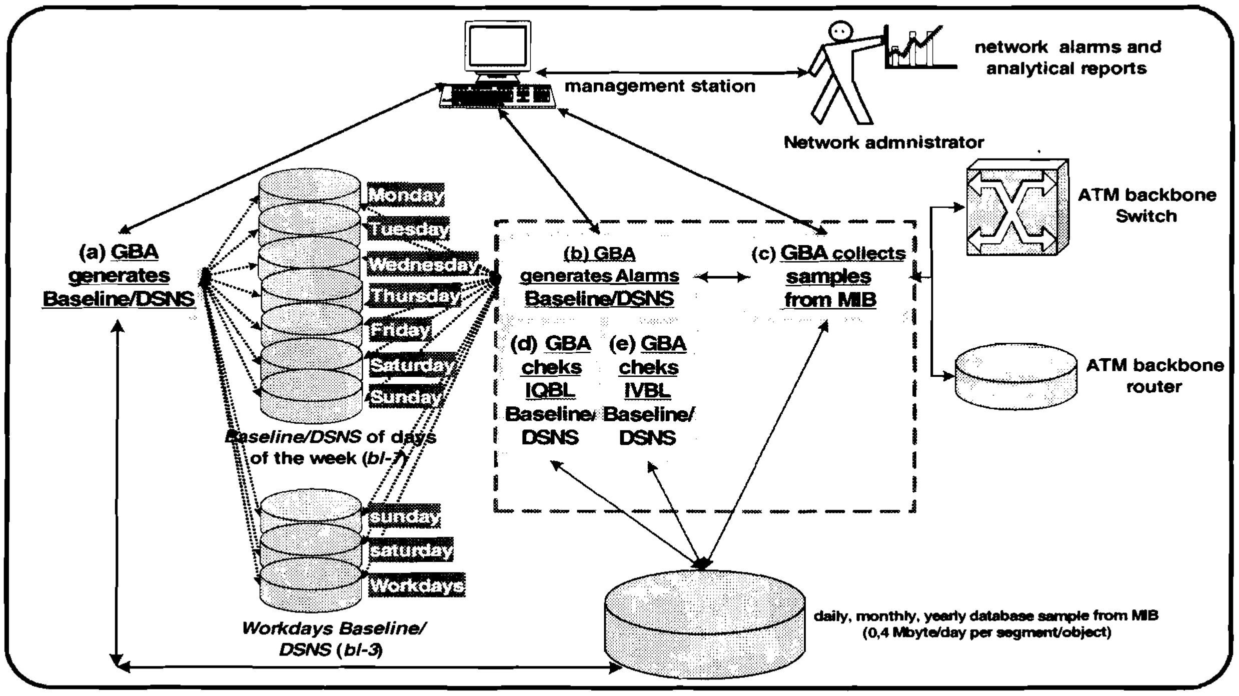
management station
network alarms and analytical reports
Network admnistrator
ATM backbone Switch
ATM backbone router
(a) GBA generates Baseline/DSNS
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Baseline/DSNS of days of the week (bl-7)
sunday
saturday
Workdays
Workdays Baseline/ DSNS (bl-3)
(b) GBA generates Alarms Baseline/DSNS
(c) GBA collects samples from MIB
(d) GBA cheks IQBL Baseline/ DSNS
(e) GBA cheks IVBL Baseline/ DSNS
daily, monthly, yearly database sample from MIB (0,4 Mbyte/day per segment/object)

Figure 4.1 – Operational functioning diagram for the generation of DSNS and alarms

workdays of the week from Saturday and Sunday, was in order to minimize the margin of error in the final result, concerning the alterations in the volume of traffic that occur between the workdays and the other days [25]. The results showed that it was the right choice, because the variation that was found in the volume of traffic between the workdays was of 10% and over 200% comparing workdays and weekends, as can be seen in figure 4.2 and 4.3.

The model for DSNS generation proposed and presented in this work, performs statistical analysis of the collected values, respecting the exact moment of the collection, second by second for twenty-four hours, preserving the characteristics of the traffic based on the time variations along the day. For the generation of the DSNS, the holidays were also excluded due to the non-use of the network on these days. Moreover, the process of DSNS generation also considered faults in the collected samples which occur along the day, eliminating these faults from the calculations for the DSNS generation.

The GBA makes measurements at each second at the MIBs of the network equipments. Along each day, 86,400 samples are expected. Problems usually occur and may affect some of these samples due to the loss of package or congesting in the network. In this case, for the generation of the DSNS, the exclusion of these samples was chosen in the calculation of the DSNS related to that second. This problem occurs in less than 0.05% a day, for the analyzed samples.

The processing for the DSNS generation is done initially in batch aiming at its creation through data related to a pre-established period. The DSNS is generated second by second for a period of days represented by N which makes up the set nj (j = 1, 2, 3, 4, ..., N); with the daily gathering there is a set of samples of the day represented by a_i (i = 0, 1, 2, ..., 86,399). Then the bi-dimensional matrix is built with 86,400 lines and N columns which must be previously sorted and that will be represented by ***Mij***.

The algorithm used for the calculation of the DSNS (BLGBA) is based on a variation in the calculation of mode, which takes the frequencies of the underlying classes as well as the frequency of the modal class into consideration. The calculation takes the distribution of the elements in frequencies, based on the difference between the greatest G_{aj} and the smallest S_{aj} element of the sample, using only 5 classes. This difference divided by five, forms the amplitude h between the classes, $h = (G_{aj} - S_{aj}) / 5$. Then the limits of each LCk class are obtained. They are calculated by $L_{Ck} = S_{aj} + h*k$, where Ck represents the k class ($k = 1...5$).

The proposal for the calculation of the DSNS of each Bli second has the purpose of obtaining the element that represents 80% of the analyzed samples. The ***Bli*** will be defined as the greatest element inserted in class with accumulated frequency equal or greater than 80 %. The

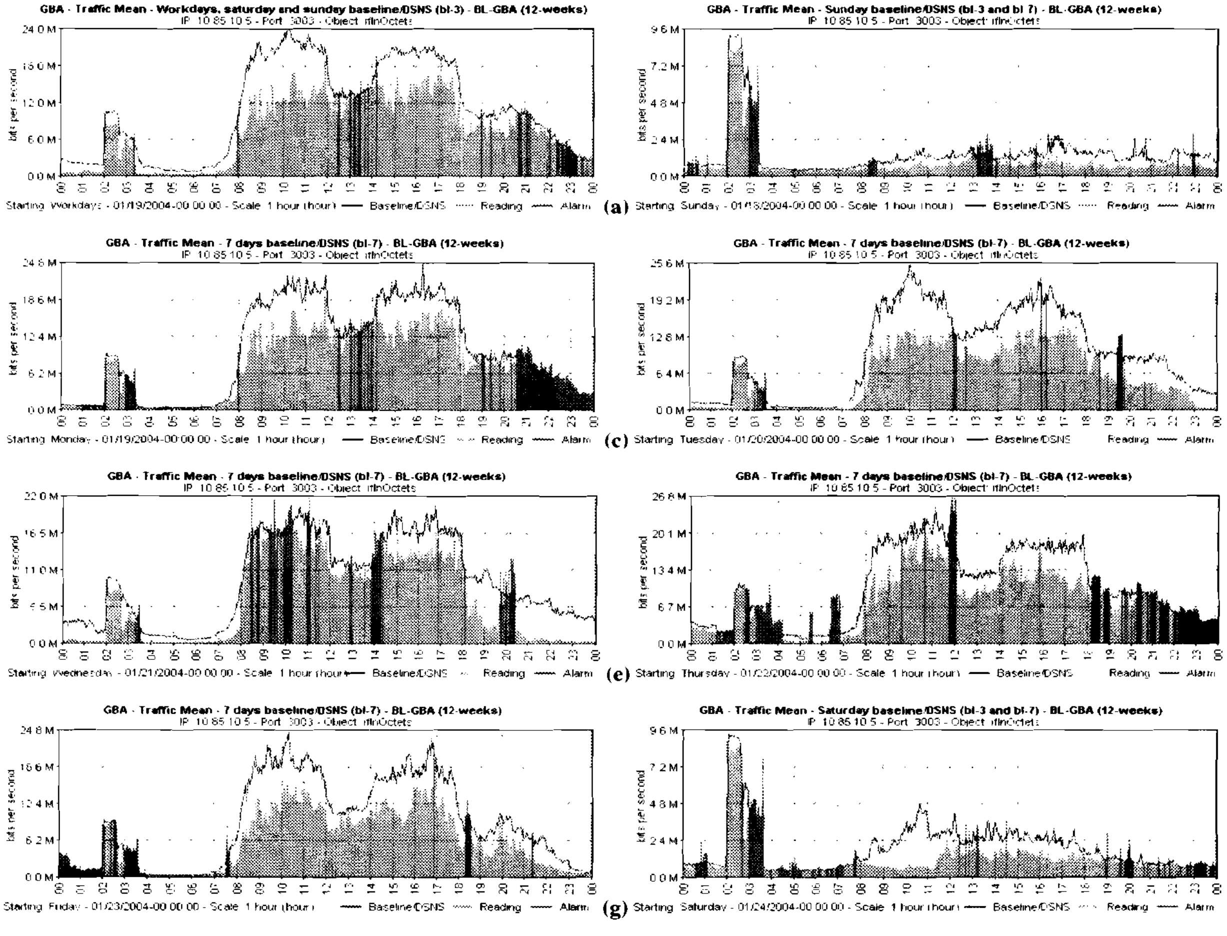

Figure 4.2 – DSNS and the daily movement for S_1 segment analyzed from 01/18/2004 to 01/24/2004.

purpose is to obtain the element that would be above most samples, respecting the limit of 80%. This process is used for the generation of DSNS models ***bl-7*** and ***bl-3***.

The BLGBA model used for the calculation of the DSNS was chosen after the performance of tests with other statistical models based on the mean, octile, decile average and on the mode. The choice for the BLGBA model was based on:

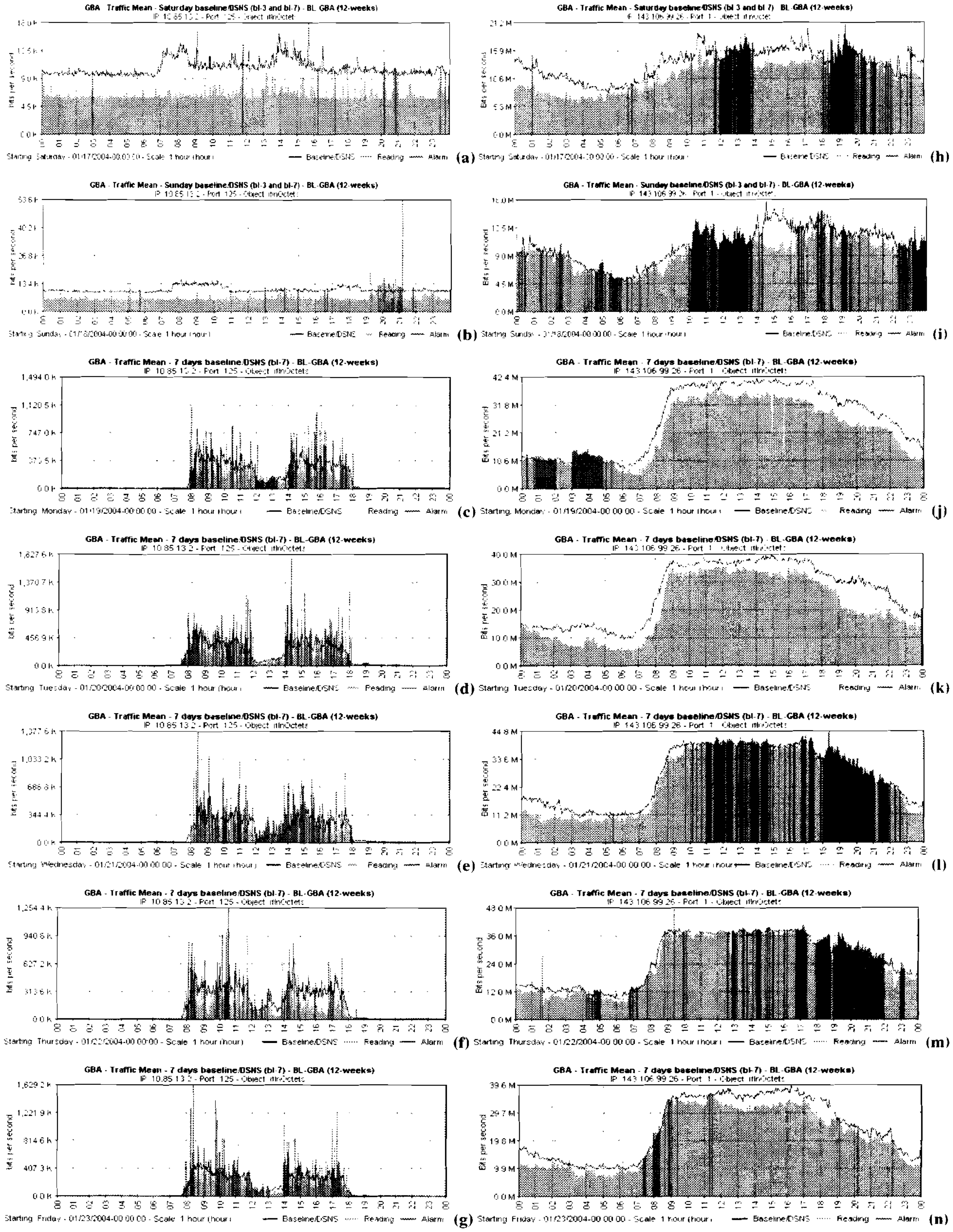

Figure 4.3 – DSNS and the daily movement for S2 and S3 segments from 01/17/2004 to 01/23/2004.

1. Visual analysis of graphics containing the DSNS and its respective daily movement, as illustrated in figures 4.2 and 4.3. Through the visual analysis of the real movement and baseline, it was observed that BLGBA presented more satisfactory results in relation to the other analyzed models. The mean and mode models presented a forecast below the real movement while the octile and the decile average models have the characteristic to present traffic trends that have occurred few times, this we call generation vice which was eliminated by the BLGBA;
2. Deviation analysis proposed by Bland and Altman [27][28], takes into consideration the differences between the predicted and observed movements. Such differences must lie between an interval defined by $\bar{d} \pm 2 * s$, where $\bar{d}$ is the differences mean and s is the standard deviation of these differences. With this an upper and lower limit are set where the deviation must be contained. The model that presented better adjustment was the BLGBA, with 95% of the differences in these limits;
3. Residual analysis – the model which showed less residual index between the predicted and the occurred movements was the BLGBA;
4. Linear regression [29][31] between the models aimed at evaluating which one showed a better correlation coefficient between the DSNS and the daily movement. Figure 4.4 shows the result of the correlation tests for the segment S_1 related to the months of September to November 2003. In this figure it is possible to notice that the BLGBA shows a better correlation coefficient between the daily movement and the DSNS.

The choice for the element that represents 80% of the samples for the calculation of the DSNS Bli was done empirically. Analytical tests were carried out through linear regression [29][31] using DSNS with this value ranging between 0 and 100%, with the purpose of verifying if 80% would be the best value to be used by the BLGBA, in the calculation of the Bl_i. Figure 4.5 shows the correlation coefficient R between the DSNS and the samples for values of choice between 0 and 100 %. It is noticed that the DSNS that uses 80%, shows a better correlation coefficient for BLGBA. These tests along with the visual analysis of the graphics with DSNS and their respective daily movements showed that the value of 80% for the calculation of the *Bl* was the most satisfactory one.

The obtained results show the validity of the model for the generation of the DSNS, bearing in mind the performed analyses and the comparison with the real movement that occurred. An example of that can be seen in figure 4.2 that illustrates, in the form of a histogram, the daily movement of the segment S_1, and their respective DSNS for a whole week in January 2004. Figures 4.3 (a), (b), (c), (d), (e), (f) and (g) show a whole week for S_2 and figures 4.3 (h), (i), (j), (k), (l), (m) and (n) for S_3 concerning the third week of January 2004. In these figures some graphs are shown, with the DSNS in blue and the real movement that occurred on the day in green. We came to the following conclusions

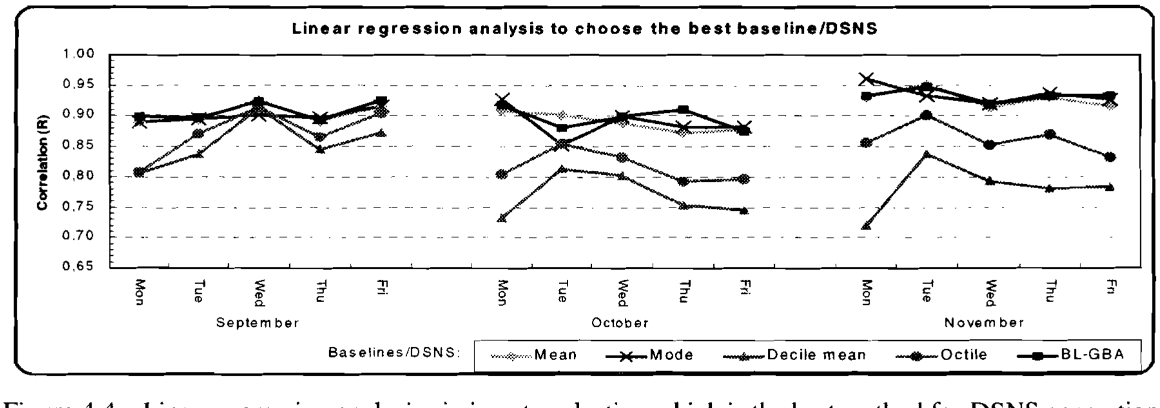

Figure 4.4 – Linear regression analysis aiming at evaluating which is the best method for DSNS generation.

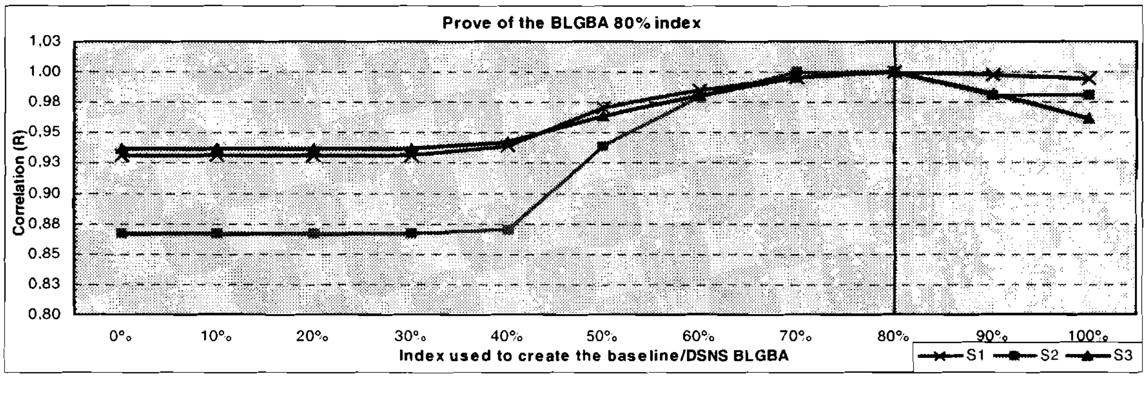

Figure 4.5 – Linear regression analysis aiming at validating the choice for the index of 80% for the BLGBA.

with the results shown in figures 4.2 and 4.3:

1. Clear peaks of traffic in the DSNS, everyday between 2:00 and 3:30 o'clock in the segment S_1, that are related to the backup performed in this period in the network server;
2. The DSNS is influenced by time factors which, in this case, are related to the working day that starts at 8:00 a.m. and finishes at 10:00 p.m.
3. Periods in which the traffic of the day becomes higher than the DSNS. In this case, its color is changed from green to red, which means a peak of traffic above the DSNS, and this could or could not be interpreted as an alarm;
4. The profile of traffic for the workdays, figures 4.2 (a), generated by the ***bl-3*** model and 4.2 (c), (d), (e), (f) and (g), generated by the ***bl-7*** model, is quite similar with a strong time dependence along the day which, in this case, is related to the working day hours of the university where the tests were performed. In the case of Saturdays and Sundays, the DSNS generated for these days are exactly the same as it can be observed in figures 4.2 (b), (h), for ***bl-3*** and ***bl-7*** models;
5. Not only the DSNS generated for the workdays ***bl-3*** but also the one generated for all the days of the week ***bl-7***, showed to be suitable for the characterization of the traffic. The ***bl-7*** is a model of DSNS to be used in cases in which there is the need to respect individual particularities which occur in each day of the week, such as backup days, whereas the ***bl-3*** is the most suitable for the cases where this is not necessary, that is, all the workdays can be dealt with in a single DSNS, leaving the decision on what model to be used to the network manager;
6. The generated DSNS fulfill their main objective which is the characterization of the traffic in the analyzed segments;

Unfortunately, due to the limited quantity of information that is presented in this article, it is not possible to show other figures which corroborate what was presented in this work. Nevertheless, at the address http://gba.uel.br/blgba more information and results obtained through this work can be found.

4.1 DSNS EVALUATION

We created an index with the purpose of evaluating the coefficient of variation of the DSNS of one month in relation to the other. This index is called Index of Variation of the Baseline (IVBL). The IVBL is calculated based on the difference between one DSNS and the other, as shown in equation (1). With the IVBL it was possible to conclude that there is usually a increasing in the volume of traffic from one month to the other, showing that despite being small, there is a tendency of growth in the volume of traffic

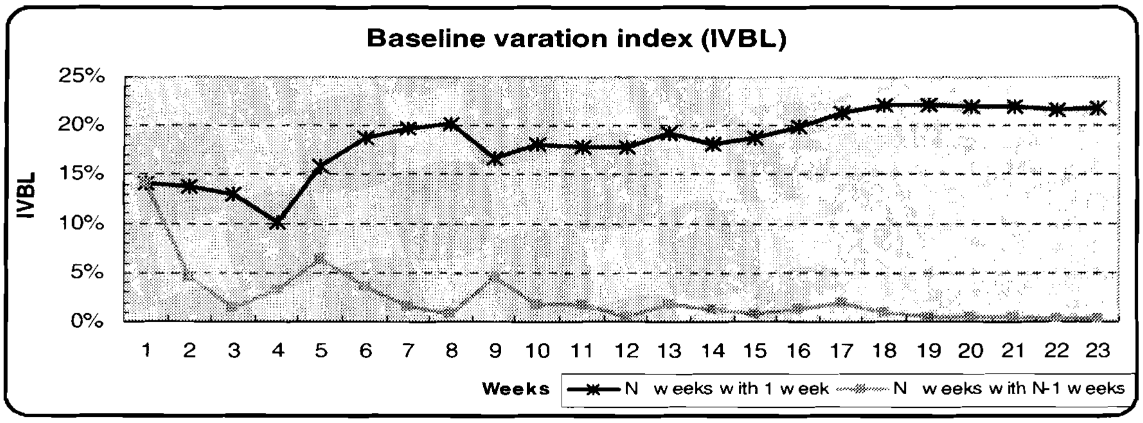

Figure 4.6 - % of variation of the baseline of ***n*** weeks compared to the (***n*** – 1) weeks and 1 week.

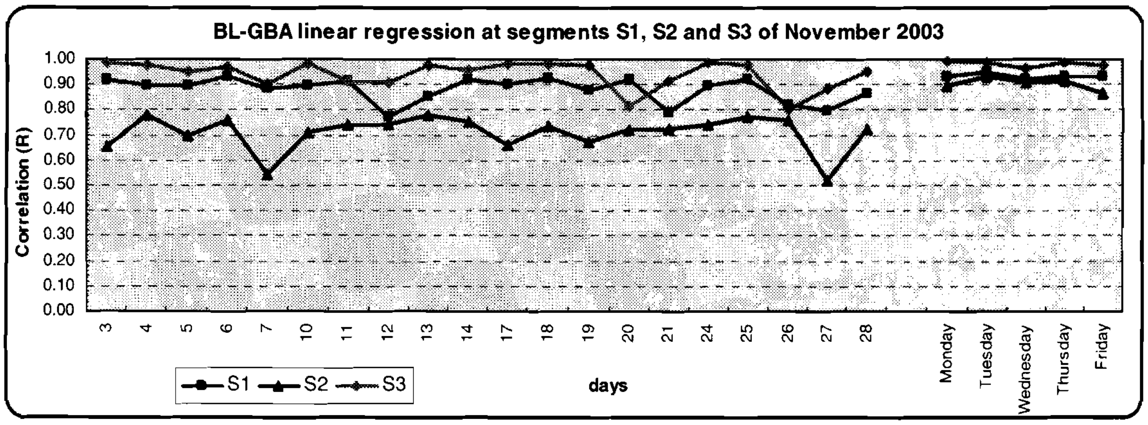

Figure 4.7 - Analysis of the BLGBA by linear regression of November 2003.

in the analyzed segments. Table 4.1 shows the percentage of growth in the segment S1 from the network of UEL, from January 2003 to January 2004. In the other analyzed segments, a small percentage of growth was also observed.

$$IVBL = \left(\sum_{i=1}^{86400} BL'_i - BL''_i \right) \Big/ 86400 \quad (1)$$

Where IVBL = variation index of a baseline in relation to another

The IVBL was also used to calculate the variation of DSNS generated from n weeks and compared to a DSNS of (n - 1) weeks, and in the comparison between the DSNS of 1 week with the DSNS of n weeks. These calculations using weekly DSNS were carried out with the purpose of evaluating and demonstrating the minimum quantity of samples necessary for the formation of the DSNS. With the comparison of the DSNS of n weeks with the one of (n – 1), during 24 weeks, it was observed that the percentage of variation tends to stabilize from the 12th week on, and not being significant for the formation of the DSNS. And when a DSNS of 1 week was established and a comparison was carried out for 24 weeks, it was also noticed that, from the 12th week on, the percentage of variation tends to stabilize around 20%, showing no more significant variations that could be added to the DSNS from this point on. The figure

Table 4.1 –Variation of the DSNS from January 2003 to January 2004, for segment S_1

%of growth of the baseline/DSNS camparede with the previous month													
	jan/03	fev/03	mar/03	Apr/03	May/03	jun/03	jul/03	Aug/03	Sep/03	Oct/03	nov/03	Dec/03	jan/04
IVBL	1.10%	1.51%	5.38%	0.07%	8.66%	2.94%	5.83%	6.38%	4.95%	4.12%	2.78%	2.89%	3.02%

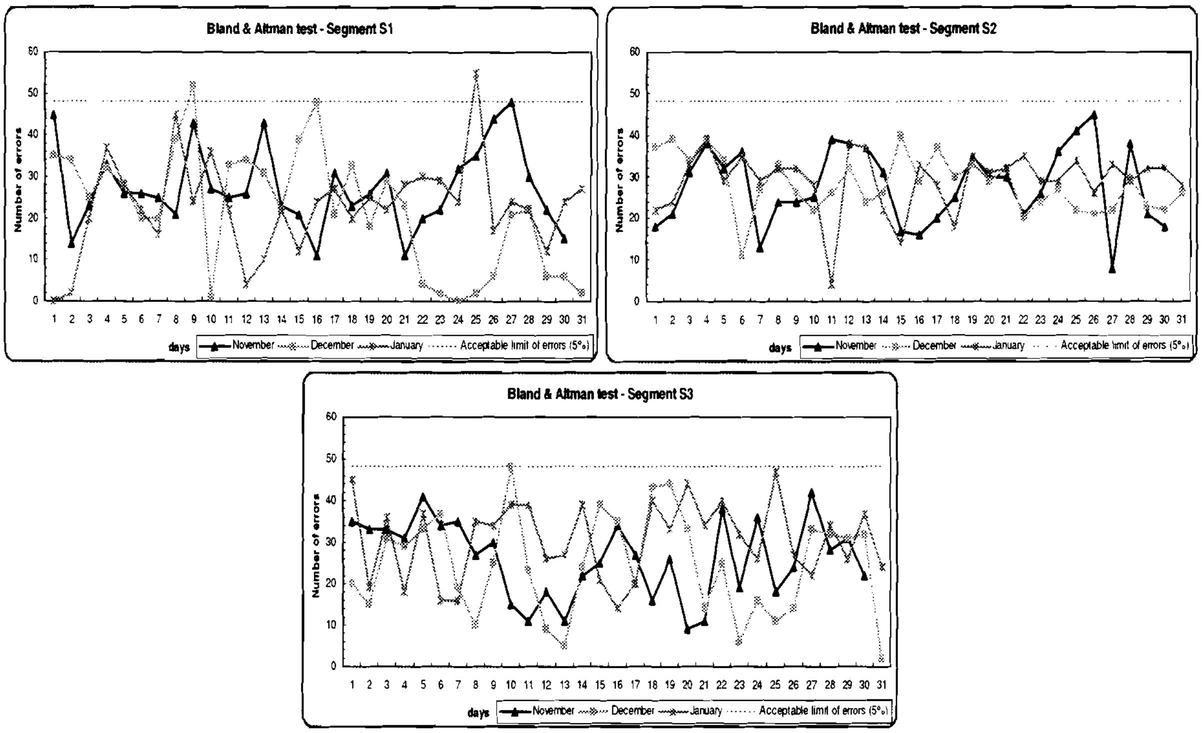

Figure 4.8 – Bland & Altman test from November 2003 to January 2004 for segments S1, S2, and S3.

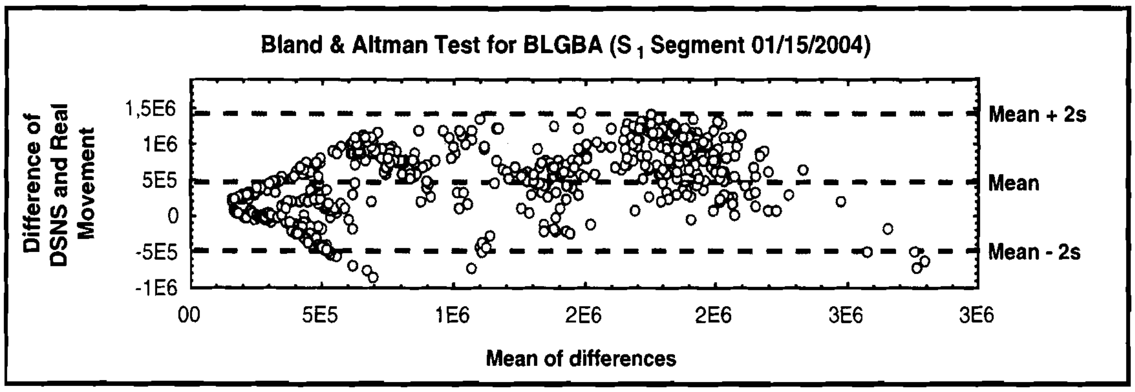

Figure 4.9 – Differences between the baseline and real movement for 01/15/2004.

4.6 shows the results of these comparisons. The IVBL test shows that would be necessary at least 4 and no more than 12 weeks for the formation of the DSNS.

Besides the visual evaluation of the results, other analytical tests have been carried out aiming to evaluate the reliability of the DSNS generated by the BLGBA in relation to the real movement. The tests were carried out from January 2003 to January 2004 using:

I. Linear Regression [29][31]: Figure 4.7 presents the results demonstrating a high correlation and adjustment between the movement that occurred those days in relation to their DSNS;

II. Test purposed by Bland & Altman [27][28]: Refer to the deviations analysis that occur between the DSNS and the real movement. 95 % of the deviations/errors observed during all days from January 2003 to January 2004, in segments S_1, S_2, and S_3, are between the required limits of $\overline{d} \pm 2 * s$, where $\overline{d}$ is the mean and s is the standard deviation of the differences between the DSNS and the real movement, confirm the reliability of the model, as shown in figure 4.8. Figure 4.9 presents an example of a daily analysis of deviations/errors, for the S_1 segment at 01/15/2004.

III. Hurst parameter (H): Tests carried out with the real movement and the DSNS generated by the BLGBA, using the statistical methods Variance-time, Local Whittle and Periodogram 0, generate the hurst parameter H. The analysis confirms that the traffic is self-similar and the DSNS is also self-similar, however presenting a lower hurst parameter. Figure 4.10 illustrates an example of these calculations for real movement and its DSNS (BL-7) for S_1, S_2 and S_3 segments during January 2004. In most of the cases, these tests also allow us to notice that in segments with lower number of computers like S_2, the hurst parameter presents a lower rate, between 0.6 and 0.7, in segments with great aggregated traffic like the S_1 and S_3 it presents a rate between 0.8 and 1.0. The Hurst parameter evaluation was made using the samples collected second by second with the GBA tool. Calculations were made for

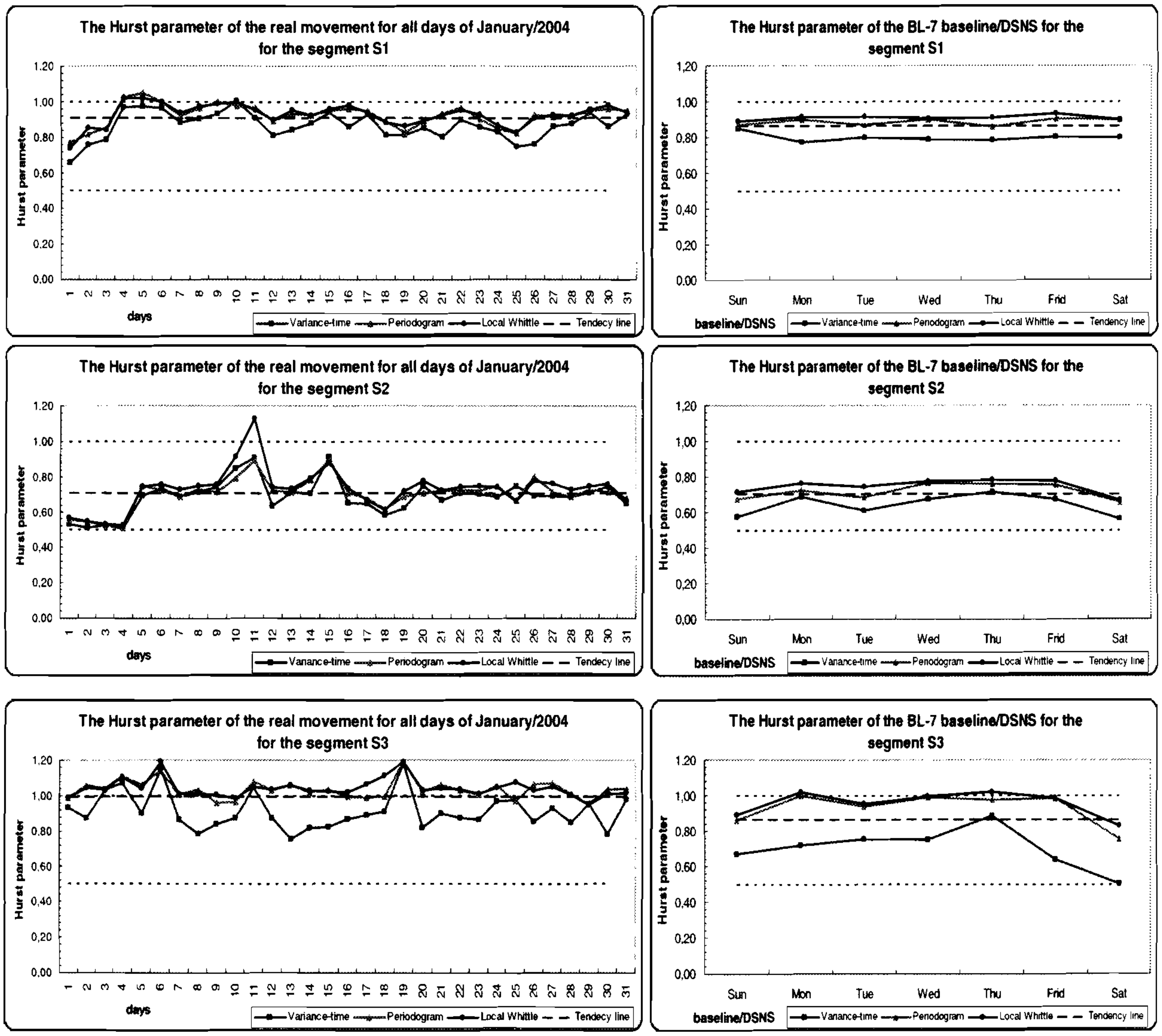

Figure 4.10 – The Hurst Parameter (H) for the real movement and its DSNS of S_1, S_2 and S_3 segments in January 2004.

each day between 8:00 and 18:00 hours, the period when the traffic is more similar to a stationary stochastic process 0. Its utilization makes possible the evaluation of the DSNS quality in segments of different burstiness. Indicating that the greater the burstiness of the segment the bigger the Hurst parameter and the better the characterization shown by the DSNS. And the lower the burstiness of the segment, the smaller the Hurst parameter and worse the results shown by the DSNS. These results are corroborated by the other tests utilized to validate the DSNS that also indicate an increase of the DSNS quality in segments with a higher burstiness.

5. ALARMS

The use of the DSNS makes it possible to achieve another important point related to the addition of accuracy in the mechanisms of alarms used by management tools. This also enables the automation of the monitoring of network segments, performed visually through the graphs generated by tools such as MRTG and GBA, mentioned in section 2 of this work. In order to do so, the construction of a mechanism of alarm was chosen, based on a modification of the hysteresis mechanism, shown initially in the RMON [26][32], and the thresholds established by the DSNS.

The key idea is that the network manager will only receive an alarm in case a deviation from normal network behavior occurs that justifies his attention. These possible anomalies are restricted to significant differences of the daily movement in relation to its DSNS. This system is called the GBA Generates Alarms and working real time with the GBA Collects Data, as illustrated in figure 4.1, modules (b) and (c).

The alarm mechanism establishes a window of time t for anomalies detection, that we call hysteresis window. In this window anomalies from the normal behavior and the forecast by the DSNS will be analyzed. The intent of histeresis window is to reduce the probability of false alarms, generated by transient behavior of burst traffic.

During the tests, it was possible to observe that the use of the thresholds generated by the DSNS for each second of the day cause 15% of alarms which demonstrates, in practice, the random behavior of the traffic in networks as Ethernet and makes the use of monitoring through alarms simply based on the thresholds established in the DSNS

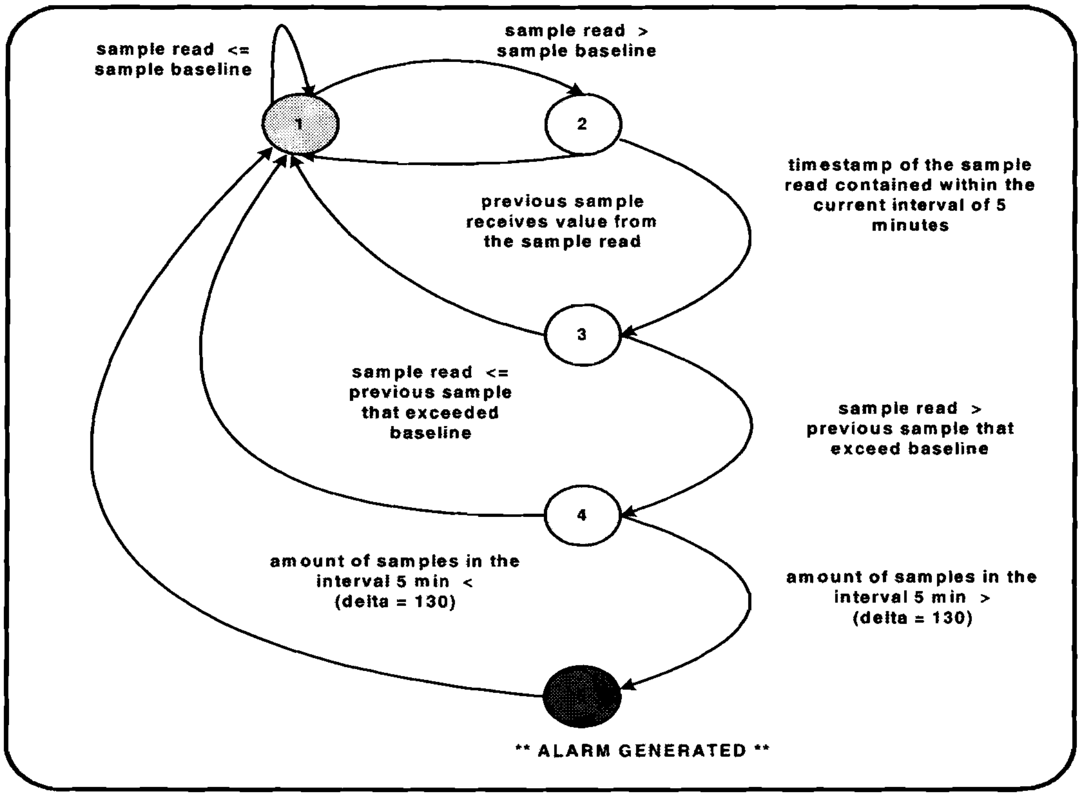
sample read <=
sample baseline
sample read >
sample baseline
1
2
timestamp of the sample read contained within the current interval of 5 minutes
previous sample receives value from the sample read
3
sample read <= previous sample that exceeded baseline
sample read > previous sample that exceed baseline
4
amount of samples in the interval 5 min < (delta = 130)
amount of samples in the interval 5 min > (delta = 130)
** ALARM GENERATED **

Figure 5.1 – Automaton that defines the functioning of the mechanism of alarms

Table 5.1 – Results of the alarms from January to December 2003 for the segment S_1

%of alarms in the month using Baseline calculated until the previous month												
Mechanisms of Alarms	January	February	March	April	May	June	July	August	September	October	October	December
Greater than Baseline	12.684	21.681	17.497	14.048	20.203	12.194	13.051	15.088	19.075	15.009	11.000	5.950
Original hysteresis (RMON)	6.422	6.422	8.914	7.142	10.134	6.178	6.664	7.733	9.799	7.710	5.680	3.060
Hysteresis modification	0.002	0.002	0.014	0.009	0.016	0.009	0.009	0.011	0.029	0.021	0.010	0.001

impossible. The same thing happened for the inferior thresholds of the DSNS.

The use of the hysteresis mechanism, in its original format implied that 7% of the samples would generate alarms, which would make its use also impossible due to a great quantity of alarms that would be generated. Our modification of the hysteresis mechanism for the control of alarms, generated a model that considers intervals of time t for the monitoring in which alarms are generated, only if the three rules below are broken:

- Rule 1: the analyzed sample is higher or lower than the superior or inferior thresholds established in the DSNS;
- Rule 2: the analyzed sample is higher or lower than the previous sample that broke rule 1 within the interval t;
- Rule 3: the quantity of samples that broke rules 1 and 2 is higher than δ.

Rule number 3 was included in order to prevent excessive number of alarms occasioned by momentary bursts. The tests showed a relation which is inversely proportional to δ in relation to the quantity of alarms generated. In other words, the bigger the δ, the smaller will be the quantity of generated alarms. After several analyses, it was noticed that an acceptable value for the relation of alarms and problems that occurred, implied in a δ equal to 130, considering an interval of hysteresis of 5 minutes. Figure 5.1 shows the automaton related to the implemented model for the alarm system in GBA tool.

After implementing the modification of the mechanism of hysteresis that forecasts the use of the δ accumulator for the generation of alarms, it was possible to notice the occurrence of a very small number of alarms per day, signaling effectively that something different was happening in the analyzed segment at that moment. In table 5.1 is shown the summary of the alarms that happened during 2003 in the segment S_1, using the model of DSNS ***bl-7*** presented in this work. As it could be seen, the modification in the hysteresis mechanism for the control of the alarms reduced the quantity of alarms to less than 0,01% a day, making its use possible. In this table is also shown the monthly summary of the alarms caused by the real movement higher than the DSNS and higher than the DSNS using the hysteresis mechanism proposed in the RMON.

The alarm system that was presented achieved its goal, helping with the specific management of UEL's network, that in this case used the t parameters equal to 5 minutes and $\delta = 130$. It also showed the practical evidence of one of the advantages of automatic monitoring which could be offered to the network management. The proactive management is only achieved because the alarm system that uses the DSNS for the generation of alarms is used together with the program that gathers information of the switch each second of the day, being able to activate the alarm as soon as it happens.

The experience also shows that the t and δ parameters can be customized in order to make the alarm system more or less sensitive to the variations of traffic in relation to its DSNS and consequently have a more or less rigid monitoring, depending on the necessity.

6. CONCLUSION

This work presents two important contributions: the first one related to the automatic generation of digital signature of network segments (DSNS), which constitutes itself into an important mechanism for the characterization of the traffic of the analyzed segment, through thresholds that reflect the real expectation of the volume of traffic respecting the time characteristics along the day and the week. This enables the network manager to identify the limitations and the crucial points in the network, control the use of the network resources, establish the real use of the resources, besides contributing to the planning of the needs and demands along the backbone.

The second contribution is the alarm system, integrated to the DSNS as well as to the monitoring performed real time by the GBA, figure 4.1 (b) and (c), that makes it possible for the network manager to be informed through messages, at the exact moment a significant difference related to the traffic and the DSNS is detected. This possibility is fundamental for the segments or crucial points of the networks that demand perfect control and proactive management in order to avoid the unavailability of the services rendered.

The use of graphs such as the ones shown in figures 4.2 and 4.3 with information about the DSNS and about the daily movement, makes a better control over the segments possible.

It could be noticed that the behavior of the traffic of the Ethernet networks is random, self-similar and extremely influenced by the quantity of bursts, which intensify as the number of hosts connected to the segment increase, as shown in [30], It also showed that the model chosen for the DSNS, presented in this work, is viable for the characterization of the traffic in backbone segments that concentrate the traffic of a great number of hosts, as shown in the examples of section 4.

Tests were also realized with DSNS from other MIB objects, like ipInReceives, icmpInMsgs, udpInDatagrams. The results have been satisfactory and demonstrate that the BLGBA model can be used for other MIB objects, however more tests must be done aiming to evaluate this possibility.

Besides the tests performed at the networks of UEL and in the Communications Department of the Electric Engineering Faculty of UNICAMP, with results validating the model presented in this work, tests with different types of networks, such as factories, large Internet providers and industries should be performed, aiming to evaluate and perfect the model proposed for generation of DSNS.

A research being developed is the creation of a multiparametric model for alarm generation aiming to aid the security, performance and fault management, using a set of some monitored objects DSNS, such as IP, TCP, UDP and ICMP packet traffic, traffic volume in bytes and number of errors. The model consists in the utilization of a DSNS set, information about possible network anomalies and rules for alarm generation. These are based on thresholds in differentiated levels, which would indicate specific conditions to customizable problems to the network. A creation of an efficient mechanism of anomaly detection and alarm generation is expected.

REFERENCES

[1] Firoiu, V.; Le Boudec, J.-Y.; Towsley, D.; Zhi-Li Zhang; Theories and models for Internet quality of service, Proceedings of the IEEE, Volume: 90, Issue: 9, Sept. 2002, Pages: 1565 – 1591.

[2] El-Gendy, M.A.; Bose, A.; Shin, K.G, Evolution of the Internet QoS and support for soft real-time applications, Proceedings of the IEEE, Vol.91, Iss.7, July 2003, Pages: 1086- 1104.

[3] Duffield, N.G.; Grossglauser, M.; Trajectory sampling for direct traffic observation; Networking, IEEE/ACM Transactions on, Volume: 9, Issue: 3, June 2001, Pages: 280 - 292.

[4] INTERNET ENGINEERING TASK FORCE (IETF). Overview and Principles of Internet Traffic Engineering, RFC 3272, may.2002.

[5] Trimintzios, P.; Pavlou, G.; Flegkas, P.; Georgatsos, P.; Asgari, A.; Mykoniati, E.; Service-driven traffic engineering for intradomain quality of service management; Network, IEEE , Volume: 17 , Issue: 3 , May-June 2003, Pages:29 - 36.

[6] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). OSI Management Framework, ISO 7498-4, Geneva 1989.

[7] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). OSI Systems Management Overview, ISO 10040, Geneva 1992.

[8] Paul Barford, Jeffery Kline, David Plonka, Amos Ron; A signal analysis of network traffic anomalies, Internet Measurement Workshop; Proceedings of the second ACM SIGCOMM Workshop on Internet measurement, Marseille, France, Pages: 71 - 82, 2002, ISBN:1-58113-603-X

[9] Rueda, A.; Kinsner; A survey of traffic characterization techniques in telecommunication networks, Electrical and Computer Engineering, 1996. Canadian Conference on, Vol.2, Iss., 26-29 May 1996, Pages:830-833 vol.2.

[10] Adas, A.; Traffic models in broadband networks, Communications Magazine, IEEE, Vol.35, Iss.7, Jul 1997, Pages:82-89.

[11] Sugih Jamin; Danzig, P.B.; Shenker, S.J.; Lixia Zhang; A measurement-based admission control algorithm for integrated service packet networks, Networking, IEEE/ACM Transactions on, Volume: 5, Issue: 1 , Feb. 1997, Pages:56 - 70.

[12] Yen-Wen Chen; Chung-Chi Chou; Traffic modeling of a sub-network by using ARIMA; Info-tech and Info-net, 2001. Proceedings. ICII 2001 - Beijing. 2001 International Conferences on, Vol.2, Iss., 2001, Pages: 730-735 vol.2.

[13] Hajji, H.; Baselining network traffic and online faults detection; Communications, 2003. ICC '03. IEEE International Conference on, Volume: 1, 11-15 May 2003, Pages: 301 - 308.

[14] Thottan, M.; Chuanyi Ji; Anomaly detection in IP networks, Signal Processing, IEEE Transactions on Volume: 51, Issue: 8, Aug. 2003, Pages: 2191 – 2204.

[15] Papavassiliou, S.; Pace, M.; Zawadzki, A.; Ho, L.; Implementing enhanced network maintenance for transaction access services: tools and applications, Communications, 2000. ICC 2000. IEEE International Conference on, Volume: 1, 18-22 June 2000, Pages: 211 - 215 vol.1.

[16] Balachander Krishnamurthy, Subhabrata Sen, Yin Zhang, Yan Chen, Sketch-based change detection: methods, evaluation, and applications, Internet Measurement Workshop Proceedings of the 2003, ACM SIGCOMM conference on Internet measurement; Miami Beach, Pages: 234 - 247, ISBN:1-58113-773-7.

[17] Thottan, M.; Chuanyi Ji, Proactive anomaly detection using distributed intelligent agents; Network, IEEE, Volume: 12, Issue: 5, Sept.-Oct. 1998, Pages: 21 – 27.

[18] Cabrera, J.B.D.; Lewis, L.; Xinzhou Qin; Wenke Lee; Prasanth, R.K.; Ravichandran, B.; Mehra, R.K.; Proactive detection of distributed denial of service attacks using MIB traffic variables-a feasibility study, Integrated Network Management Proceedings, 2001 IEEE/IFIP International Symposium on ,14-18 May 2001 Pages:609 – 622.

[19] NORTHCUTT, Stephen, NOVAK Judy. Network Intrusion Detection, Third Edition, New Riders, 2002.

[20] Xinzhou Qin; Wenke Lee; Lewis, L.; Cabrera, J.B.D.; Integrating intrusion detection and network management, Network Operations and Management Symposium, 2002. NOMS 2002. 2002 IEEE/IFIP, 15-19 April 2002.

[21] Ferramenta para Auxílio no Gerenciamento Backbone Automatizado, Available by Web in http://proenca.uel.br/gba/ (07/03/2004).

[22] The Multi Router Traffic Grapher (MRTG), Available by Web in http://www.mrtg.com/ (07/03/2004).

[23] INTERNET ENGINEERING TASK FORCE (IETF). Management Information Base for Network Management of TCP/IP-based internets: MIB-II, RFC 1213, mar.1991.

[24] PROENÇA, Mario Lemes, Jr. "Uma Experiência de Gerenciamento de Rede com Backbone ATM através da Ferramenta GBA", paper publish in, XIX Simpósio Brasileiro de Telecomunicações – SBrT 2001, Fortaleza de 03-06 of Setembro 2001.

[25] Frank Feather, Dan Siewiorek, Roy Maxion, Fault detection in an Ethernet network using anomaly signature matching, Applications, Technologies, Conference proceedings on Communications architectures, protocols and applications, San Francisco, SIGCOMM 93, ISSN:0146-4833.

[26] INTERNET ENGINEERING TASK FORCE (IETF). Remote Monitoring Management Information Base Version, RFC 1757, fev.1995.

[27] Bland J. Martin and Altman Douglas G., Statistical Methods For Assessing Agreement Between Two Methods of Clinical Measurement, The LANCET i: 307-310, February 8, 1986.

[28] Bland J. Martin, Altman Douglas G. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 346, 1085-7, 1995.

[29] Bussab, Wilton O.; Morettin Pedro A. Estatística Básica, Editora Saraiva, 5a edição 2003.

[30] Leland Will E., Taqqu M. S., Willinger W., Wilson D. V., On the Self-Similar Nature of Ethernet Traffic (Extended Version), IEEE/ACM Transactions on Networking, volume 2, No 1, February 1994.

[31] PAPOULIS, Athanasios, Pillai S. Unnikrishna. Probability, Random Variables and Stochastic Processes, Fourth Edition, McGraw-Hill, 2002.

[32] PERKINS, David T. RMON- Remote Monitoring of SNMP-Managed LANs, Prentice Hall, 1999.

Mario Lemes Proença Jr. received the of M.Sc degree in Computer Science from the Computer Science Institute of Federal University of Rio Grande do Sul, Porto Alegre, Brazil, in 1998. He will obtain the Ph.D. degree at State University of Campinas in 28/07/2005. Also, he is a computer science professor since 1991 in state university of Londrina, Brazil. His research interests include Computer Network, Network Operations and Management and Security. He currently is leader of the group of research in computer networks of computer science department of State University of Londrina.

Fabio Sakuray received the M.Sc. degree in Computer Science from the Federal University of São Carlos, in 1994. He is currently working towards the Ph.D. degree at State University of Campinas. His research interests include transmission of Voice over IP networks and Quality of Service (QoS).

Camiel Coppelmans received the title of Bacharel in Computer Science from State University of Londrina, Brazil, in 2003. He is currently working towards the M.Sc degree in Electrical Engineering from the State University of Campinas.

Antonio Marcos Alberti received the title of Electrical Engineerig from the Federal University of Santa Maria, Santa Maria, Brazil, in 1996; the M.S. degree in Electronics and Communications from the State University of Campinas, Campinas, Brazil, in 1998; and the Ph.D. degree in Telecommunications and Telematics also from the State University of Campinas, in 2003. He is currently working in National Institute of Telecommunications, Brazil, as a teacher and researcher, Brazil.

Maurício Luis Bottoli received the title of Electrical Engineering from the Federal University of Santa Maria, in 1997, and the M.Sc. degree in Electrical Engineering from the State University of Campinas, in 1999. He is currently working towards the Ph.D. degree at State University of Campinas. His research interests include optical networks, modeling and optical simulation tools.

Leonardo Mendes received his B.S. degree in 1985 from the Gama Filho University, Rio de Janeiro, his M.S. degree in 1987 from the Catholic University of Rio de Janeiro, and his Ph.D. degree in 1991 from Syracuse University, all in electrical engineering. In 1992 he joined the Faculty of Electrical Engineering of the State University of Campinas, Brazil. Prof. Mendes's recent R&D focus is in the studies and development of Communications Engineering applications for metropolitan IP networks. Prof. Mendes created, at UNICAMP, the Laboratory of Communications Network (LaRCom), from which he is now the Director and also the main coordinator. At LaRCom , Prof. Mendes and his group have developed or are developing the following projects: 1) an optical system simulator to help in the analysis of optical networks; 2) an environment for the simulation of systems using event driven technique which allows the development of Atm, IP and CDMA simulators; 3) development of Internet set top boxes using J2ME for small devices; 4) communications description of Internet devices using Corba component modules for Telecommunications; 5) development of e-Learning objects for the PGL project. Nowadays, 4 professors, 2 post doctorate researchers and 17 graduate students participate in the projects developed in the LaRCom.