Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

Baseline Aplicado a Gerência de Redes

Autor: Mario Lemes Proença Jr.

Orientador: Prof. Dr. Leonardo de Souza Mendes

Tese de Doutorado Apresentada a Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos para obtenção do titulo de Doutor em Engenharia Elétrica. Área de Concentração: **Telecomunicações e Telemática**.

Banca Examinadora

Leonardo de Souza Mendes, Ph.D	DECOM/FEEC/UNICAMP
Dalton Soares Arantes, Dr.	DECOM/FEEC/UNICAMP
José Valdeni de Lima, Dr.	Informática/UFRGS
Mauricio Ferreira Magalhães, Dr.	DCA/FEEC/UNICAMP
Renato Baldini Filho, Ph.D.	DECOM/FEEC/UNICAMP
Rubens Nascimento Melo, Dr.	DI/PUC-Rio

Campinas, SP
Julho / 2005

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

P942b

Proença Junior, Mario Lemes
Baseline Aplicado a gerência de redes / Mario Lemes Proença Junior. --Campinas, SP: [s.n.], 2005.

Orientador: Leonardo de Souza Mendes.
Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Gerência. 2. Redes de Computação – Gerência. 3. Telecomunicações tráfego. 4. Redes de computação – Medidas de segurança. I. Mendes, Leonardo de Souza. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Titulo em Inglês: *Baseline* applied to network management
Palavras-chave em Inglês: Network management, Baseline, Alarms, Traffic characterization
Área de concentração: Telecomunicações e Telemática
Titulação: Doutor em Engenharia Elétrica
Banca examinadora: Dalton Soares Arantes, José Valdeni de Lima, Mauricio Ferreira Magalhães, Renato Baldini Filho e Rubens Nascimento Melo
Data da defesa: 28/07/2005

RESUMO

Nesta tese é apresentado o modelo BLGBA (*baseline* GBA), que se destina a geração de *baseline* para segmentos de rede. O modelo foi desenvolvido e implementado na ferramenta Gerenciamento de *Backbone* Automatizado (GBA) que se destina a auxiliar no gerenciamento de redes. Ele apresenta como seu maior benefício a geração automática de *baseline*, com base em análises realizadas de objetos residentes nas MIBs, de agentes SNMP, contidos nos equipamentos de rede.

Um estudo sobre trabalhos relacionados ao tema desta tese, referente à caracterização de tráfego e a detecção de anomalias para auxiliar no gerenciamento de redes, também é exposto neste trabalho.

Outra contribuição desta tese é o sistema para detecção de anomalias (ADGBA), que utiliza o *baseline* gerado pelo modelo BLGBA, em conjunto com o movimento coletado em tempo real, nos segmentos de rede. O objetivo principal é informar ao administrador da rede, somente no caso de ocorrência de algum evento significativo não previsto pelo *baseline*.

Para validação do modelo BLGBA e do sistema ADGBA foram realizados testes analíticos e práticos, com dados reais, coletados das redes da Universidade Estadual de Londrina e da Universidade Estadual de Campinas. Os resultados obtidos mostraram tanto a validade do modelo quanto à eficiência do sistema, proporcionando de forma prática e objetiva vantagens significativas para gerência de redes.

Palavras-chave: Gerência de Redes, *Baseline*, Alarmes, Caracterização de Tráfego.

Abstract

In this thesis the BLGBA (GBA Baseline) model is presented, which intends to create a baseline for network segments. The model was developed and implemented using the GBA tool, which is used as an aid in network management. The major advantage of this model is the automatic generation of the baseline. The baseline was generated based on analyses of SNMP objects of network equipment MIBs.

A study about related works to the subject of this thesis is presented, referring to the traffic characterization and anomalies detection aiming to help network management.

Another contribution of this thesis is the anomalies detection system (ADGBA), that use the baseline generated by BLGBA model and the real movement collected in real time of the network segments. The main objective is to inform the administrator only in case of occurrences of significant events not foreseen by the baseline.

Analytical and practical tests have been carried out using real data collected from the State University of Londrina and State University of Campinas networks, aiming to evaluate the BLGBA model and ADGBA system. The obtained results shown the validate of the model as also the efficiency of the system and show in practice significant advantages in network management.

Keywords*: Network Management, Baseline, Alarms, Traffic characterization.*

A minha esposa Aida e aos meus filhos Mahara e Pedro pelo amor, carinho e compreensão que me proporcionam.

A minha Mãe por seu exemplo de vida.

Agradecimentos

Agradeço a DEUS por ter me ajudado e dado força para a realização deste trabalho.

Agradeço a minha esposa AIDA pelo amor, apoio, carinho e compreensão durante os quatro anos que duraram este Doutorado.

Agradeço aos meus filhos MAHARA e PEDRO pelo amor, carinho e compreensão durante as minhas ausências.

Agradeço aos meus pais pela formação, amor e carinho que me proporcionaram e continuam proporcionando.

Agradeço aos meus amigos e colegas do NPD/ATI e do Departamento de Computação pelo apoio e compreensão durante a realização deste trabalho.

Ao meu orientador Leonardo, pela confiança depositada em mim e principalmente pela amizade que se estabeleceu.

A CAPES pelo apoio no projeto BELLATRIX, o qual proporcionou apoio financeiro, através de bolsa para realização deste projeto e das publicações efetuadas.

Ao CNPq pelo apoio no projeto ORION, que tornou possível a implementação de bolsistas e a aquisição de equipamentos fundamentais na execução deste trabalho.

Sumário

Lista de Figuras

Lista de Tabelas

Glossário

ADGBA	*Anomaly Detection System GBA* (sistema de detecção de anomalia GBA)
ANSP	Rede Acadêmica de São Paulo
API	*Aplication Program Interfaces*
ATM	*Asynchronous Transfer Mode*
bl-3	*baseline* de três dias
bl-7	*baseline* de 7 dias
BLGBA	*baseline* GBA (modelo para geração de *baseline* usando a ferramenta GBA)
Bl_i	*baseline i* de um segundo ao longo do dia
CPU	*Unit control Processor*
Diffserv	Arquitetura de serviços diferenciados
DSNS	*Digital Signature of Network Segment* (assinatura digital de segmentos de rede)
GBA	Gerenciamento de Backbone Automático
HTML	*HyperText Markup Language*
Intserv	Arquitetura de serviços integrados
IP	*Internet Protocol*
IQBL	Índice de qualidade do *baseline*
ISO	*International Organization for Standardization*
ISP	Internet Service Provider
IVBL	Índice de variação do *baseline*
JPEG	*Joint Photographic Experts Group*
MIB	*Management Information Base*
MPLS	*Multi-Protocol Label Switching*
MRTG	*Multi Router Traffic Grapher* (ferramenta de gerência)

NET-SNMP	Conjunto de aplicações utilizadas para acessar ao SNMP
NMS	*Network Management Systems*
NOC	*Network Operations Center*
P2P	*Peer to Peer* (aplicações ponto a ponto)
QoS	*Quality of Service*
RMON	*Remote Network Monitoring*
RMON2	*Remote Network Monitoring 2*
SMS	*Short Message Service*
SNMP	*Simple Network Management Protocol*
TCP	*Transmission Control Protocol*
TCP/IP	*Transmission Control Protocol/Internet Protocol*
UDP	*User Datagram Protocol*
UEL	Universidade Estadual de Londrina
UNICAMP	Universidade Estadual de Campinas
VoIP	*Voice over IP*

Trabalhos Publicados Pelo Autor

1. Proença Jr., Mario Lemes; Sakuray, Fabio; Mendes, Leonardo **Uma Experiência de Gerenciamento de Rede com *Backbone* ATM através da Ferramenta GBA**, Artigo publicado no congresso, XIX Simpósio Brasileiro de Telecomunicações – SBrT 2001, Fortaleza 03-06/09/2001.
2. Proença Jr., Mario Lemes; Coppelmans C.; Alberti, A.; Bottoli, Mauricio; Mendes, Leonardo, ***The Hurst Parameter for Digital Signature of Network Segment***, *11 International Conference On Telecommunications* - ICT 2004, Fortaleza. Springer-Verlag in the LNCS 3124. 08/2004, ISBN 3-540-22571-4. Pag. 772-781.
3. Proença Jr., Mario Lemes; Coppelmans C.; Bottoli, Mauricio; Mendes, Leonardo; ***Baseline to Help With Network Management*****,** artigo publicado no ICETE 2004 – *International Conference on E-business and Telecommunication Networks*, Setubal – Portugal – August 24-28. Proceedings of ICETE, INSTICC Press, ISBN 972-8865-15-5. Volume 2. Pag. 152-160.
4. Proença Jr., Mario Lemes; Zarpelão, B. Bruno; Mendes, Leonardo; ***Anomaly Detection for Network Servers using Digital Signature of Network Segment*****,** artigo publicado na *IEEE Advanced Industrial Conference on Telecommunications – AICT 2005,* Lisboa, Portugal, de 17 a 21/07/2005. IEEE Computer Society, Proceedings of AICT-2005, ISBN 0-7695-2388-9. Pag. 290-295.
5. Proenca Jr., Mario Lemes; Zarpelão, B. Bruno; Mendes, S. Leonardo; ***Anomaly Detection Aiming Pro-Active Management of Computer Network Based on Digital Signature of Network Segment***, artigo publicado na *4th IEEE Latin American Network Operations and Management Symposium* - LANOMS 2005, Porto Alegre, Brasil, de 29 a 31 de agosto de 2005.
6. Proença Jr., Mario Lemes; Zarpelão, B. Bruno; Bottoli M.; Breda G.; Mendes, S. Leonardo **Correlação de Objetos SNMP na Detecção de Anomalias em Servidores de Rede**, XXII Simpósio Brasileiro de Telecomunicações SBrT 2005, Campinas de 04 a 08 de setembro de 2005.

7. Proença Jr., Mario Lemes; Coppelmans C.; Sakuray Fabio; Alberti A.; Bottoli, Mauricio; Mendes, Leonardo; ***A Practical Approach for Automatic Generation of Network Segment Traffic Baselines***, artigo publicado na revista do SBrT em abril de 2005, São Paulo – Brazil – 2005. ISBN. (prelo)
8. Proença Jr., Mario Lemes; Coppelmans C.; Bottoli, Mauricio; Mendes, Leonardo; ***Baseline to Help With Network Management*****,** artigo publicado pela Kluwer Academic Publishers – ICETE 2004 *Best Papers* - 2005 - *International Conference on E-business and Telecommunication Networks*, a ser publicado em 2005. (prelo)

1 Introdução

O tráfego que flui pela Internet nos dias de hoje se caracteriza basicamente por serviços como voz, dados e imagens. Novos serviços e mídias estão sendo continuamente desenvolvidos com intuito de melhorar a interação e a comunicação entre os usuários de rede. Serviços como voz sobre IP (VoIP), vídeo sob demanda, jogos on-line, mensagens instantâneas, sistemas ponto a ponto (P2P) para compartilhamento e troca de mídias, fazem parte deste conjunto heterogêneo de informações que são transportadas através da Internet e do TCP/IP.

As redes de computadores atualmente são de vital importância para a sociedade atual, semelhante aos serviços essenciais como água, luz e telefonia. Seus serviços não podem ser interrompidos devido a importância fundamental que representam para as pessoas e empresas que os utilizam.

Inúmeros trabalhos e discussões têm sido realizados com intento de implementar qualidade de serviço (QoS) e engenharia de tráfego ao longo do *backbone* da Internet (Duffield e Grossglauser, 2001). Tecnologias como arquitetura de serviços integrados (Inteserv), arquitetura de serviços diferenciados (Diffserv) e o Multi-Protocol Label Switching (MPLS) foram desenvolvidos, implementados e estão sendo testados com o propósito de fornecer qualidade de serviço ao longo do *backbone* da Internet (El-Gendy *et al.,* 2003), (Firoiu, 2002), (RFC 3272, 2002), (Trimintzios *et al.* 2003).

No cenário atual, a utilização de redes e da Internet se tornaram imprescindíveis, fazendo com que seu crescimento continue a acontecer de forma exponencial. A busca pela implantação de QoS entre suas conexões e a utilização de novos serviços, exigem que a automação nas funções de gerenciamento sejam realizadas visando, essencialmente, a não interrupção dos serviços, a otimização na utilização dos recursos, redução de custos e a detecção antecipada de falhas a fim de evitar o congestionamento ao longo de seu *backbone* (Hajji, 2003), (Barford *et al.,* 2002).

O estabelecimento ou a previsão do que seria o comportamento normal do tráfego de rede, em um segmento ou componente da rede, como *switches*, servidores ou estações de trabalho, é uma função básica, porém fundamental, para auxiliar os gerentes de redes. Neste trabalho será apresentado o modelo chamado de *Baseline* GBA (BLGBA), desenvolvido para caracterizar o tráfego em segmentos de rede, com objetivo de gerar *baseline* destes segmentos, que deverão conter informações referentes a previsão sobre o comportamento do mesmo ao longo do dia.

Muitas ferramentas e sistemas de gerenciamento de redes (*Network Management Systems* - NMS) têm como objetivo auxiliar nas cinco áreas definidas originalmente como fundamentais para a gerência de redes pela *International Organization for Standardization* (ISO) (ISO, 1989), (ISO, 1992), quais sejam falhas, configuração, contabilização, segurança e performance. Contudo, não existe disponível nestas ferramentas e sistemas tradicionais de gerenciamento uma opção fundamental que se destina à construção de *baseline*, que represente as características singulares de cada segmento da rede.

O *baseline* pode ser definido como o conjunto de informações que representam o perfil do tráfego de um segmento de rede, através de limiares mínimos e máximos sobre o volume de tráfego, quantidade de erros, tipos de protocolos e serviços que trafegam por este segmento ao longo do dia.

A previsão real ou mesmo aproximada em um determinado instante sobre as características do tráfego, esperadas em um segmento ou elemento de rede, tornam as decisões de gerenciamento mais fáceis e confiáveis. Um dos principais objetivos deste trabalho é desenvolver ferramentas para que o administrador de rede solucione rapidamente e de forma eficiente os problemas que possam vir acontecer. O ideal, sempre procurado, é a realização da gerência de forma pró-ativa, antecipando soluções antes mesmo que os problemas aconteçam, porém nem sempre isto é possível, uma vez que a falta de informações contextuais são inerentes a cada situação intrínseca da gerência de redes. O *baseline* não seria em si a única solução para o gerenciamento pró-ativo, mas sem dúvida se constitui em um dos componentes principais para que ela seja alcançada.

Outro aspecto fundamental e até vital na gerência de redes, se refere à detecção e localização de anomalias. A definição de quais eventos representam uma anomalia e que

devem ser reportados ao administrador da rede, ainda é uma questão em aberto e que esta fruto de estudo com intuito de ser padronizada, tendo em vista que elas são diretamente relacionadas as políticas estabelecidas para a rede (Lakhina *et al.,* 2004), (Lakhina *et al.,* 2004b), (Thottan, 2003), (Barford *et al.,* 2002), (Roughan *et al.,* 2004). Estes eventos podem ser caracterizados como uma falha física ou mesmo de software que levam a uma interrupção ou degradação no serviço oferecido ao usuário final.

Roughan *et al.* (Roughan *et al.,* 2004) constatou que algumas anomalias detectadas, apesar de não aparecerem nos *logs* do sistema, necessitavam ser reportadas ao administrador de rede. Esses eventos apresentavam uma grande variação no comportamento dos dados monitorados. Lakhina *et al.* (Lakhina *et al.,* 2004) assumiu que para um evento ser considerado uma anomalia, não necessita causar grandes impactos na rede. O fato desse evento representar uma degradação no serviço oferecido ao usuário final, já justifica a sua notificação ao administrador da rede.

Este trabalho está organizado da seguinte maneira. No Capítulo 2, será apresentado um estudo sobre trabalhos relacionados à área desta pesquisa. No Capítulo 3, serão apresentados detalhes sobre a motivação para utilização de *baseline* na gerência de redes, bem como detalhes da implementação do modelo BLGBA proposto para geração de *baselines* ou assinatura digital para segmentos de rede (DSNS). Serão apresentados também resultados práticos da aplicação do modelo BLGBA. No Capítulo 4, serão apresentados um sistema de alarmes multinível em conjunto com um sistema para detecção de anomalias (ADGBA), que utilizam o *baseline* gerado pelo modelo BLGBA para monitoramento e controle de segmentos de rede. Por último, no Capítulo 5, são apresentadas conclusões e sugestões para futuros estudos na área.

2 Trabalhos Relacionados

Existem muitos trabalhos realizados sobre caracterização e medição de tráfego que são relacionados com a proposta apresentada neste trabalho (Rueda, 1996), (Adas, 1997), (Sugih *et al.,* 1997). A caracterização e a medição de tráfego são aspectos importantes que devem ser considerados na gerência e no controle de redes com a finalidade de auxiliar na engenharia de tráfego, detectar problemas de performance e de congestionamento (Jain, 2004). Em (Rueda, 1996) e (Adas, 1997) é apresentado uma síntese sobre diversos trabalhos realizados para caracterização de tráfego em redes de telecomunicações. Segundo (Adas, 1997), os modelos que não utilizam características estatísticas do tráfego apresentam uma performance pobre por sub ou super estimar a performance da rede.

Os modelos mais tradicionais e antigos para modelagem de tráfego eram inicialmente chamados de modelos de renovação. Assumiam o tráfego como sendo um processo de renovação, onde eram utilizados modelos baseados em processos de Bernoulli e na distribuição de Poisson (Adas, 1997), (Jagerman *et al.,* 1996).

Os modelos de tráfego classificados como estacionários, com dependência de curta duração, que incluem modelos baseados em cadeias de Markov, introduziram uma dependência aleatória que possibilitaram a captura do nível de rajadas de tráfego por apresentar uma correlação diferente de zero (Adas, 1997), (Jagerman *et al.,* 1996).

Já os modelos que tratam da dependência de longa duração encontrada em tráfego multimídia, são caracterizados por uma auto-correlação que decai mais lentamente do que os de curta duração, o que é definido de acordo com a análise de variância amostral. Modelos como ARIMA foram utilizados para descrever este tipo de tráfego (Adas, 1997), (Jagerman *et al.,* 1996). Em (Yen-Wen e Chou, 2001) é apresentado uma proposta para modelagem do tráfego em sub-redes utilizando ARIMA, com o objetivo de capturar as características do tráfego de Internet nestas sub-redes para a realização de análises de performance.

Leland *et al*. (Leland *et al.,* 1994) apresentou um modelo auto-similar para modelagem de tráfego. Esse trabalho foi um marco que colocou em cheque as abordagens adotadas até 1994, baseadas em distribuição de Poisson e cadeias de Markov. Demonstraram que o comportamento do tráfego, desde a sua origem a seu destino, não apresentava uma regularidade estatística e por isso não convergiam para um processo gaussiano comum. Para eles o tráfego de rede apresenta propriedade de rajadas em diferentes dimensões, fazendo com que as características estatísticas se degenerem de forma independente da escala inicial observada. A principal propriedade da auto-similaridade se refere à propriedade de um objeto manter as mesmas características, mesmo quando observado em escalas diferentes. Uma estrutura ou objeto será auto-similar se um pedaço dela se assemelhar ao todo, ou seja, cada pedaço irá conter réplicas de si mesmo. Um exemplo disso pode ser observado na Figura 2.1, que apresenta um conjunto 2D de Cantor. O conceito de auto-similaridade ou de fractais foi apresentado inicialmente por Mandelbrot (Mandelbrot, 1965).

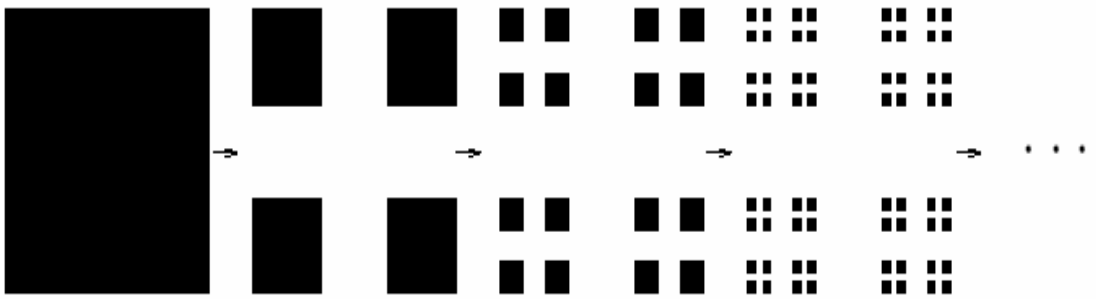


Figura 2.1 - Conjunto 2D de Cantor.

A maioria dos modelos mencionados acima se destinam a modelar o tráfego de rede usando ferramentas estatísticas de uma maneira abstrata e genérica, enquanto a proposta apresentada nesta tese, tem como objetivo gerar uma caracterização singular para cada segmento de rede, gerada a partir de dados reais, coletados nestes segmentos. O objetivo almejado foi a criação de um perfil particular para cada segmento analisado, perfil este que chamamos de *baseline* ou assinatura digital de segmentos de rede (DSNS).

Em (Hajji, 2003) é apresentada uma proposta similar à nossa, que usa uma distribuição assintótica da diferença entre sucessivas estimativas de um modelo para o tráfego de rede. Um problema que ocorre no modelo apresentado por Hajji é que ele supõe que os dados utilizados para o treinamento e testes de seu modelo são puros e sem anomalias. Em nosso caso o *baseline* é calculado e gerado a partir de dados reais coletados no segmento analisado.

Outra importante área que está relacionada a este trabalho é referente à detecção de anomalia em redes (Thottan, 2003), (Papavassiliou, 2000), (Krishnamurthy *et al.,* 2003). A detecção de anomalias no tráfego de redes é uma função fundamental. Tem como objetivo identificar rapidamente e de forma eficiente problemas advindos de falhas, ataques ou mesmo excesso de carga nos segmentos de rede que possam gerar até mesmo a descontinuidade do serviço.

Existem basicamente duas abordagens para detecção de anomalias. A primeira baseada em assinaturas e a segunda em análises estatísticas. A detecção baseada em assinatura já foi amplamente estudada e explorada. Basicamente funciona com base no estabelecimento prévio de assinaturas de possíveis anomalias. Esta abordagem é amplamente utilizada em sistemas para detecção de intrusos, contudo ela apresenta uma desvantagem relacionada à necessidade de se conhecer a *priori*, o padrão da anomalia ou mesmo do ataque. Outra abordagem para detecção de anomalias se refere às abordagens baseadas em análises estatísticas. Neste enfoque não é necessário o estabelecimento prévio de padrões de anomalias ou ataques, pois elas serão detectadas com base em uma mudança do comportamento normal, que se baseia em análises estatísticas realizadas em dados passados. O sistema de alarmes apresentado neste trabalho se assemelha a esse tipo de abordagem, justamente por se utilizar o *baseline*/DSNS que é criado com base em análise estatística de dados históricos monitorados na rede.

Em (Thottan, 2003) é apresentado um estudo sobre métodos de vários autores para detecção de anomalias. O modelo para detecção de anomalias proposto por ela, utiliza uma técnica de processamento estatístico de sinal, baseado na detecção de mudanças abruptas, analisando dados coletados de variáveis residentes em MIBs SNMP. Thottan utiliza amostras com uma freqüência de 15 segundos e supõe como questão em aberto, que algumas mudanças no comportamento das variáveis residentes nas MIBs SNMP não

correspondem às anomalias de rede. A utilização de um *baseline* gerado a partir de dados reais do segmento analisado, pode ajudar a resolver este problema. Eles sinalizam para a importância de monitorar e obter sucesso na caracterização de vários objetos da MIB, já que cada um deles pode responder a diferentes tipos de anomalias. Para eles, um importante requisito na detecção de anomalia se deve ao fato que um objeto possa responder por mais de uma anomalia, assim como mais de um objeto possa responder por uma única anomalia. A utilização de conjuntos apropriados de variáveis da MIB e o relacionamento existente entre elas, foram apontados como fator de sucesso na detecção de anomalias. Em nosso trabalho procuramos fazer a correlação na detecção de anomalias através de análises de vários objetos SNMP para o servidor analisado.

Papavassiliou *et al.* (Papavassiliou, 2000) e Lawrence Ho apresentaram uma ferramenta que tem como objetivo facilitar o gerenciamento de rede, reduzindo custos e minimizando erros humanos. Eles utilizam uma abordagem similar à apresentada neste trabalho para a construção dos *baselines*, que são calculados em dias separados, um para dias úteis e outro para finais de semana. Em (Ho Lawrence *et al.,* 2000) é apresentado de forma mais detalhada o software chamado TRISTAN, que foi utilizado para analisar transações em tempo real com base em dados históricos. Para eles a intensidade do tráfego pode ser calculada diretamente pela duração das transações que ocorrem. Eles calculam o *baseline* com base em dados históricos das transações, que são computados através de uma série temporal de uma dimensão. Assim como nós, eles geram um *baseline* e um limite de tolerância superior e inferior ao que é utilizado para a geração de alarmes. Um problema que constatamos é que a técnica ainda não foi testada na prática com tráfego de redes como IP ou mesmo *Ethernet*, em contraste com o estudo que apresentamos, onde focamos o desenvolvimento e testes realizados totalmente com dados reais coletados de redes IP e *Ethernet*.

Recentemente, Krishnamurthy *et al.* (Krishnamurthy *et al.,* 2003) apresentou uma interessante proposta para detecção de anomalias baseadas em *sketchs*, que utiliza uma variação de séries temporais para realização da previsão ou caracterização do tráfego. Esta técnica se concentra na detecção de mudanças significativas em grande fluxo de dados, com custo operacional pequeno. A eficiência desta técnica ainda não foi totalmente provada em operações em tempo real, pois depende fortemente da precisão e da periodicidade do

recálculo do parâmetro de previsão do modelo. O objetivo de nossa pesquisa foi a construção de um modelo simples que possa trabalhar em tempo real, auxiliando efetivamente o gerente de rede de forma eficiente e confiável a monitorar e detectar problemas que possam estar ocorrendo na rede.

Barford *et al.* (Barford *et al.,* 2002), apresenta um estudo sobre a aplicação da análise de sinal, utilizando *wavelet filters* em dados coletados através do SNMP e de fluxos IP. Eles conseguiram realizar a separação, através de diferentes freqüências do sinal, em alguns grupos com diferentes características, tornando possível com isso, a identificação de anomalias no tráfego. A partir dessa abordagem eles constataram, assim como nós, que o tráfego de rede apresenta ciclos diários e semanais bem definidos. Assim como eles, acreditamos ser essencial estabelecer um *baseline* sobre o comportamento padrão do tráfego, visando facilitar a identificação no futuro de anomalias que possam ocorrer.

Lakhina *et al.* (Lakhina *et al.,* 2004), apresentaram uma técnica para o diagnóstico de anomalias, que inclui a detecção, identificação e a quantificação das mesmas. Para eles uma simples indicação que a rede enfrenta problemas já é suficiente para que uma anomalia seja sinalizada. A identificação tem como objetivo determinar a real localização da fonte dos problemas frente aos eventos sinalizados como anômalos. Eles também procuram determinar quão discrepante está o tráfego real do comportamento esperado. Para eles, apesar da grande quantidade de literatura disponível sobre caracterização de tráfego, o entendimento sobre anomalias de tráfego ainda requer especial atenção. Tendo em vista a natureza do tráfego ser multidimensional incluindo ruídos, ou seja, uma grande diversidade de fluxos, protocolos e serviços que dificultam a descoberta de anomalias. Um detalhe importante apresentado por eles e que deve ser melhor estudado é a definição sobre o volume de anomalias, pois para eles ela se refere simplesmente à mudança repentina positiva ou negativa no fluxo de tráfego analisado.

Wu *et al.* (Wu, 2003) caracteriza o padrão de operação normal do tráfego através da análise fatorial. Esse processamento estatístico permite que um grande conjunto de dados seja traduzido em um conjunto menor de fatores comuns. As anomalias são detectadas através do cálculo da distância de Mahalanobis, que é aplicado para comparar os fatores que traduzem o comportamento normal, com o tráfego realmente encontrado na rede. Os

resultados apresentados pelo autor se relacionam com a detecção de ataques e questões de segurança das informações que trafegam na rede monitorada.

Roughan *et al.* (Roughan *et al.,* 2004) apresenta uma técnica para detecção de anomalias do tipo *IP Forwarding*, utilizando a técnica Holt-Winters e uma outra de decomposição de dados. Eles adotaram uma abordagem simples para diminuir o número de alarmes falsos utilizando duas fontes de dados: o protocolo de gerenciamento SNMP e o protocolo de roteamento externo BGP. Baseado na premissa que os alarmes falsos, encontrados nas duas fontes de dados não estão relacionados, ou seja, não são simultâneos, o sistema gera alarmes apenas quando encontra desvios de comportamento nas duas fontes de dados para a mesma situação. Em nosso trabalho uma anomalia somente é sinalizada caso ocorra alarmes simultâneos em mais de um objeto no mesmo período.

Sekar *et al.* (Sekar *et al.,* 2002) apresentam uma abordagem para detecção de anomalias baseada em especificação, que é composta por uma fase de treinamento e outra de detecção. Para eles, a detecção de anomalias é afetada fortemente pela definição de quais aspectos do sistema monitorado devem ser aprendidos na fase de treinamento, que visa criar um perfil e com isso estabelecer um comportamento normal sobre esse sistema. Posteriormente, na fase de detecção, esse perfil é comparado com o que ele chama de comportamento corrente, com intuito de se detectar desvios que serão reportados ao administrador da rede como anomalias.

Jiang e Papavassiliou (Jiang e Papavassiliou, 2003) apresentaram um algoritmo para diagnóstico de falhas baseado em séries temporais não estacionárias. As idéias apresentadas por eles são interessantes e semelhantes às apresentadas neste trabalho. Eles utilizam limites que se adaptam ao *baseline* gerado, preservando características temporais. O que não fica claro no trabalho apresentado por eles é a viabilidade e aplicabilidade do modelo em casos reais.

Recentemente Pang *et al.* apresentaram um interessante estudo sobre caracterização de tráfego não produtivo que também chamaram de ruídos no tráfego normal de Internet (*Internet background Radiation*) ou ainda tráfego que ocorre na rede mas não é solicitado pelo usuário. Eles procuraram criar um *baseline* do tráfego não solicitado. Sua abordagem é voltada para aplicação de filtros de porta, que tem por objetivo separar o tráfego normal do

não solicitado, criando uma caracterização do chamado tráfego não solicitado. A exposição deste trabalho demonstra que a caracterização de tráfego continua sendo um aspecto fundamental na gerência de rede e tende a ser necessária nos mais diferentes e intrínsecos aspectos que fazem parte do tráfego de rede.

Apesar da extensa bibliografia disponível referente à caracterização e medição de tráfego e detecção de anomalias que estão diretamente ligados à gerência de performance e segurança, inúmeros trabalhos ainda continuam sendo desenvolvidos com a finalidade de criar mecanismos eficientes, confiáveis, consistentes e de fácil aplicabilidade para gerência de redes. O objetivo que se busca é fazer com que a gerência de segurança, performance e falhas possam ser realizadas de forma eficiente, segura e com baixo custo.

3 *Baseline* ou DSNS

O *baseline* ou assinatura digital do segmento de rede analisado (DSNS) pode ser definido como o conjunto básico de informações que definem o perfil do tráfego nestes segmentos. Este perfil é composto por limiares mínimos e máximos que indicam qual o comportamento normal para o tráfego ao longo do dia. Este conceito aplica-se também a qualquer elemento de rede que possa ser monitorado, como servidores, estações, *switches* e roteadores.

A previsão real ou mesma aproximada sobre as características que constituem o tráfego de rede em um segmento analisado, possibilita que as decisões de gerência sobre anomalias ou problemas que possam estar acontecendo, sejam mais confiáveis e seguras (Thottan, 2003), (Papavassiliou, 2000).

O uso do *baseline* pode ajudar o administrador de rede a identificar limitações e controlar melhor o uso dos recursos que são críticos para serviços sensíveis à latência, como voz sobre IP (VoIP) e transporte de vídeo, justamente por eles não suportarem retransmissão ou mesmo congestionamentos de rede. Além de melhorar o controle dos recursos, sua utilização também facilita o planejamento e expansão da rede, por claramente auxiliar ao administrador da rede a identificar o real uso dos recursos e os pontos críticos ao longo do *backbone*, evitando também problemas de performance e falhas que possam acontecer.

O uso do DSNS proporciona ao administrador de rede, vantagens relacionadas à gerência de performance que são obtidas através do conhecimento prévio da quantidade máxima e mínima de tráfego no segmento ao longo do dia. Isto torna possível o estabelecimento de alarmes e controles mais efetivos e funcionais, por estarem utilizando limites que se adaptam ao DSNS, respeitando as variações do tráfego ao longo do dia, ao invés do uso de limites lineares que são programados baseados na experiência do administrador de rede (Hajji, 2003), (Thottan, 1998). Desvios em relação ao que está sendo monitorado em tempo real e o que o DSNS expressa devem ser observados e analisados

com cuidado, podendo ou não ser considerados problemas. Para isso a utilização de um sistema de alarmes integrado ao DSNS e o monitoramento da rede em tempo real deverá tratar desses problemas e somente sinalizando ao administrador da rede quando necessário.

No tocante à gerência de segurança, a utilização do DSNS pode oferecer informações relacionadas à análise do comportamento dos usuários. O conhecimento a *priori* sobre o comportamento e as características do tráfego de um determinado segmento está diretamente relacionado ao perfil de utilização de seus usuários. Estas informações podem ser utilizadas para identificar anomalias, prevenir aspectos de intrusões ou mesmo ataques à rede, reduzindo com isso o tempo da rede fora do ar e aumentando a sua confiabilidade (Cabrera *et al.,* 2001), (Northcutt, 2002), (Xinzhou, 2002).

Hoje em dia é muito comum a utilização de ferramentas para auxiliar no gerenciamento da rede, como o MRTG (*Multi Router Traffic Grapher*) (MRTG, 2005) e o GBA (Gerenciamento de *Backbone* Automatizado) (GBA, 2005). Essas ferramentas geram gráficos com análises estatísticas baseadas em médias ao longo de um determinado período sobre um segmento ou objeto analisado, conforme pode ser observado respectivamente nas figuras 3.1 e 3.2. Contudo, a simples utilização deste tipo de ferramentas impõe sérias limitações ao administrador da rede no tocante à descoberta e solução de problemas, por ela ser realizada de forma manual, através de controle visual, utilizando somente o conhecimento empírico adquirido pelo administrador de rede no dia a dia de seu trabalho.

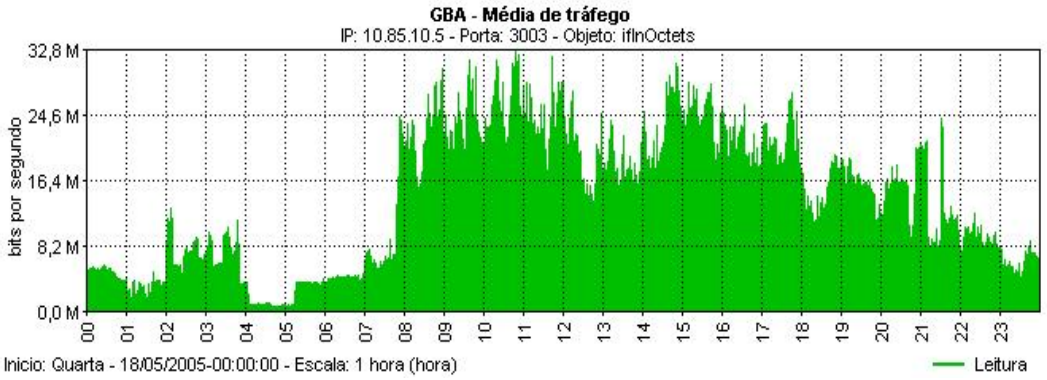
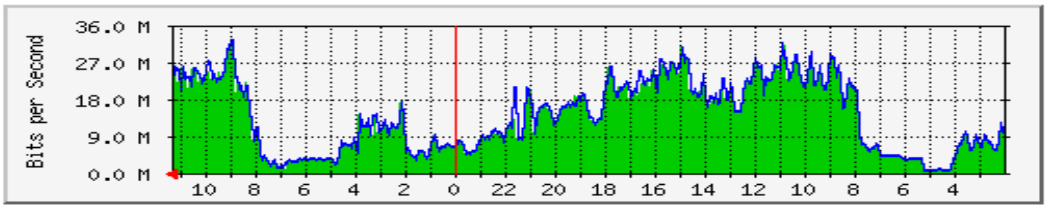


Figura 3.1 - Análise diária (média de 5 min.) realizada com ferramenta GBA.

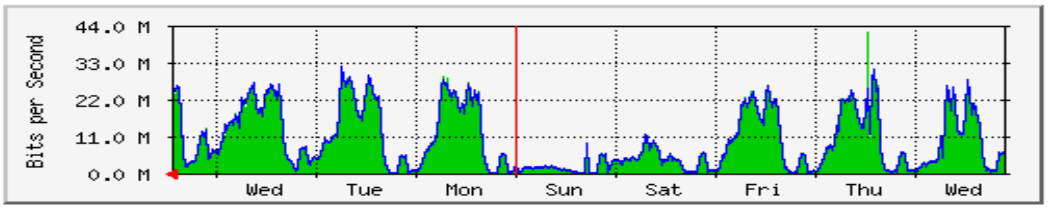
The statistics were last updated **Thursday, 19 May 2005 at 11:20**,
at which time the device had been up for **24 days, 12:11:17**.

`Daily' Graph (5 Minute Average)



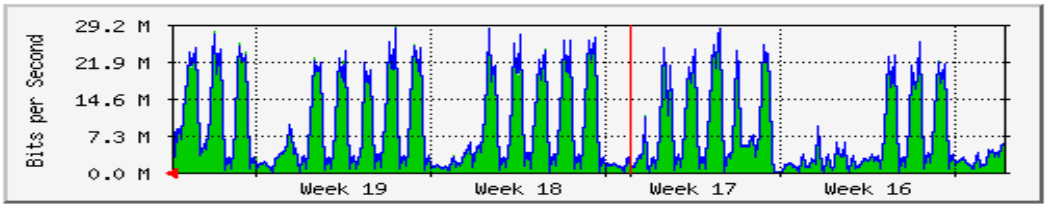
Max In:32.8 Mb/s (21.9%) Average In:14.1 Mb/s (9.4%) Current In:22.2 Mb/s (14.8%)
Max Out:32.6 Mb/s (21.7%) Average Out:14.1 Mb/s (9.4%) Current Out:22.0 Mb/s (14.7%)

`Weekly' Graph (30 Minute Average)



Max In:42.4 Mb/s (28.3%) Average In:10.1 Mb/s (6.7%) Current In:24.1 Mb/s (16.1%)
Max Out:32.1 Mb/s (21.4%) Average Out:9882.0 kb/s (6.6%) Current Out:24.2 Mb/s (16.2%)

`Monthly' Graph (2 Hour Average)



Max In:29.0 Mb/s (19.3%) Average In:8576.0 kb/s (5.7%) Current In:12.5 Mb/s (8.3%)
Max Out:28.6 Mb/s (19.1%) Average Out:8475.7 kb/s (5.7%) Current Out:12.7 Mb/s (8.5%)

Figura 3.2 - Análises diária (média de 5 min.), semanal (média de 30 min.) e mensal (média de 2 horas) MRTG.

Redes com grande quantidade de segmentos tornam ainda mais complexa sua administração, considerando a grande quantidade de gráficos a serem analisados (Papavassiliou, 2000) (Barford *et al.,* 2002). Normalmente os gráficos demonstram informações sobre o volume de tráfego médio que entra ou sai em um determinado segmento, não disponibilizando nenhum recurso a mais que possa auxiliar ao administrador de rede na tomada de decisões, referente a descoberta e solução de problemas que possam estar ocorrendo ou mesmo que já tenham ocorrido. Neste caso a utilização do *baseline*/DSNS pode proporcionar um benefício, justamente por fornecer uma previsão do que seria o comportamento padrão para o segmento ou objeto analisado.

No caso da ferramenta GBA, foi possível implantar a automação no processo de monitoramento, através da utilização de um sistema de alarmes multinível que avalia o *baseline* em relação ao movimento real, informando ao administrador da rede caso algum desvio significativo seja encontrado. O sistema de alarmes multinível será abordado com mais detalhes no Capítulo 4 desse trabalho. A principal vantagem obtida através da utilização do *baseline* neste caso é justamente o fato de o administrador não ter que ficar monitorando os gráficos gerados com objetivo de encontrar erros. O módulo implementado para esta função, disponibiliza inicialmente informações visuais nos gráficos fornecidos pela ferramenta GBA, sinalizando a ocorrência de alarmes que podem indicar uma situação anômala. Também já esta sendo implementado um módulo de envio de mensagens através da rede de dados e de mensagens do tipo SMS (*Short Message Service*) utilizando a rede de telefonia celular para avisar ao administrador da rede tão logo um alarme ou anomalia seja detectada.

3.1 Implementação do *Baseline*

O objetivo principal a ser alcançado com a construção do modelo para geração do *baseline* (BLGBA), foi a caracterização do tráfego do segmento de rede a que ele se refere. Esta caracterização se constitui em uma previsão do comportamento normal esperado para o tráfego durante todo o dia, compondo o perfil do tráfego ou *baseline* que flui pelo segmento analisado. O objetivo é que o este perfil seja composto de várias informações como: tipos de protocolos, tipos de aplicações, serviços, quantidade de erros e volume de tráfego. A meta a ser alcançada era a construção de uma ferramenta que pudesse gerar uma previsão da expectativa do tráfego e de outras variáveis que compusessem o seu perfil básico, com grau de acerto e confiabilidade aceitáveis aos padrões requeridos pelo administrador da rede. Obviamente, 100 % de acerto seria a situação ótima e ideal a qual foi perseguida durante a construção do modelo BLGBA, porém, sabe-se que isso é uma possibilidade nem sempre possível de ser alcançada tendo em vista a natureza aleatória dos dados estudados, sendo assim, procurou-se realizar uma grande quantidade de testes práticos com objetivo de validar e mesmo demonstrar a viabilidade da utilização prática do modelo proposto.

Como plataforma de desenvolvimento para o modelo BLGBA foi utilizada a ferramenta GBA (Gerenciamento Automático de *Backbone*), principalmente devido a grande quantidade de informações históricas referentes a monitoramentos realizados na rede da Universidade Estadual de Londrina (UEL) desde 1998. A ferramenta GBA, foi desenvolvida como resultado de um projeto de mestrado, com objetivo de auxiliar no gerenciamento de redes com *backbone* ATM. A ferramenta GBA cumpriu seus objetivos e foi além, na medida em que se tornou uma plataforma de aprendizagem e desenvolvimento para projetos de redes, auxiliando na pesquisa e no gerenciamento de redes. Mais informações sobre a ferramenta GBA podem ser encontradas em http://proenca.uel.br/gba.

Para realização dos testes com o propósito de validar e avaliar o modelo BLGBA, foram utilizados principalmente dados de segmentos da rede da UEL e dois segmentos da rede da Universidade Estadual de Campinas (UNICAMP). Os testes foram realizados

principalmente durante os anos de 2003 e 2004. Durante esse período pôde-se observar variações importantes impostas por períodos sazonais de férias de verão e inverno, greves, feriados e até mesmo resultado de vestibular. Esses fatores sazonais têm grande influência no comportamento normal esperado para os segmentos da rede. Os dados analisados são de redes locais com TCP/IP baseadas em *Ethernet* e no caso da UEL a rede é baseada em um *backbone* ATM operando Ethernet e TCP/IP sobre LAN *Emulation*. Abaixo estão descritos os segmentos onde os testes foram realizados:

1. O 1º segmento estudado, chamado de S_1 é responsável por interligar o firewall da rede da UEL à Internet, ele concentra o tráfego de aproximadamente 3000 computadores. O monitoramento foi realizado diretamente no servidor de firewall.

2. O 2º segmento onde foi realizada avaliação, chamado de S_2, interliga o principal servidor WEB da rede da UEL à sua rede local e a Internet. O monitoramento foi realizado diretamente no servidor de WEB.

3. O 3º segmento estudado, chamado de S_3 é responsável por interligar todos o computadores da rede da UEL ao servidor de Proxy. O monitoramento foi realizado diretamente no servidor de Proxy.

4. O 4º segmento chamado de S_4 é responsável por interligar toda a rede da UEL a seu roteador. O monitoramento foi realizado diretamente no servidor de *switch* ATM que interliga ao roteador.

5. O 5º segmento estudado, chamado de S_5 interliga a rede da UEL a sua pró-reitoria de assuntos acadêmicos. Esse segmento concentra o tráfego de aproximadamente 60 computadores.

6. O 6º segmento S_6 interliga a rede da UNICAMP à rede acadêmica de São Paulo (ANSP). Por esse segmento se concentra todo o tráfego da UNICAMP para a Internet. O monitoramento foi realizado diretamente no *switch* que interliga ao roteador.

3.2 Ambiente de desenvolvimento

O ambiente de desenvolvimento utilizado para a execução do modelo BLGBA é o mesmo utilizado pela ferramenta GBA, que atualmente se encontra na versão 5.0. A plataforma utilizada para o desenvolvimento é baseada no sistema operacional Windows XP. Para desenvolvimento foram utilizadas as linguagens de programação MS Visual C++ versão 7.0 e a linguagem JAVA utilizada para o módulo Gera_Graficos principalmente devido a sua facilidade para geração de figuras do tipo JPEG.

A interface com o usuário da ferramenta é baseada em janelas MS-DOS através de linha de comando e da WEB utilizando páginas HTML com suporte de ASP e JAVA Script.

Para acesso ao SNMP e realização de consultas as MIBs residentes nos agentes foi utilizada a biblioteca NET-SNMP na versão 5.0.8 (Net-SNMP, 2005), que disponibiliza um vasto conjunto de aplicações e *Aplication Program Interfaces* (APIs).

A estação utilizada como plataforma de gerência onde foram e continuam sendo realizados os monitoramentos e testes, é um micro que utiliza processador Pentium 4 com velocidade de 3.00 GHz e 512 MB de memória RAM, disco SATA de 120 GBytes e sistema operacional Windows XP. A estação está ligada à rede através de um link *Ethernet* de 100 Mbit/s. O overhead imposto pelo sistema à rede é de aproximadamente 0,1 % da banda disponível para realização de 25 processos de monitoramento continuamente, segundo a segundo, durante as 24 horas do dia. O consumo de CPU é da ordem de 25 % somente para realização dos monitoramentos.

A rede da Universidade estadual de Londrina utiliza atualmente, *backbone* ATM em seu núcleo e *switches ethernet* nas bordas executando *Lan Emulation*. Os testes nos permitiram identificar que os agentes SNMP dos equipamentos utilizados possibilitam até 40 consultas a objetos em 1 segundo. Este limite pode variar dependendo das condições da rede, do equipamento de monitoramento e do equipamento de rede onde o agente SNMP reside.

3.3 Ferramenta GBA

A Figura 3.3 demonstra o diagrama operacional da ferramenta GBA utilizada para implementar o modelo BLGBA. A ferramenta não é comercial e tem sido utilizada como plataforma de pesquisa com objetivo de auxiliar no gerenciamento de rede. Atualmente, a ferramenta GBA se encontra na versão 5.0 e é composta pelos seguintes módulos:

1. Lê-switches: Módulo responsável pela coleta de dados dos agentes SNMP nos equipamentos de rede.
2. Gera_*baseline*: módulo responsável pela geração de *baseline*/DSNS segundo o modelo BLGBA.
3. Gera_alarmes: módulo responsável pela geração de alarmes com base em análise do *baseline* e do movimento real coletado pelo módulo lê-switch. Este módulo e o módulo lê-switchs trabalham em conjunto.
4. Gera_graficos: módulo responsável pela geração de gráficos referentes ao monitoramento realizado e também com o *baseline* que apresenta o movimento esperado.
5. Gera_analises: módulo responsável pela realização de testes analíticos que têm como objetivo avaliar o *baseline* gerado. Os testes que são realizados por esse módulo são: Regressão linear, análise residual, teste de Bland & Altman, além do IVBL.
6. GBA_funções_básicas: responsável por executar todo o I/O dos arquivos utilizados pela ferramenta GBA. Este módulo é utilizado por todos os outros que compõem a ferramenta.

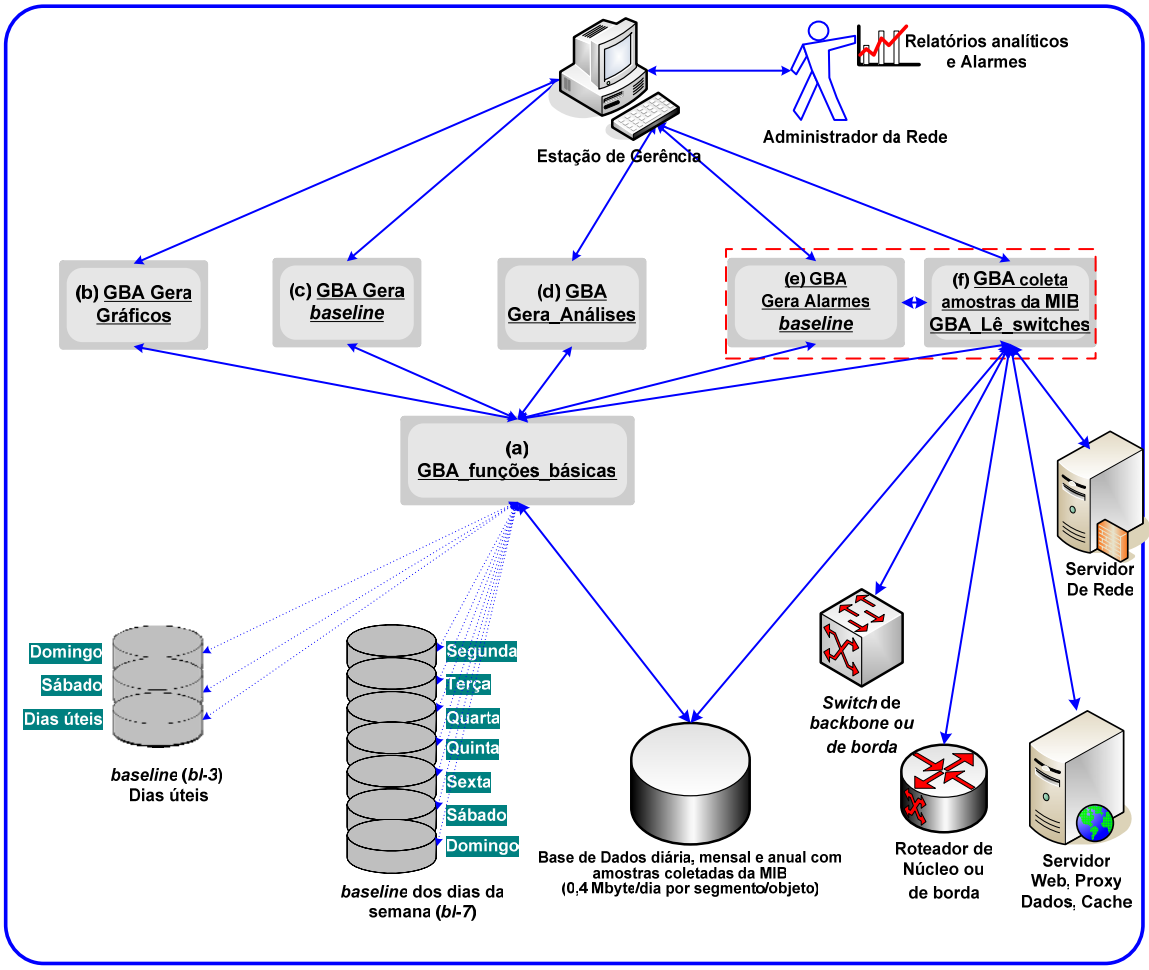


Figura 3.3 - Diagrama operacional de funcionamento da ferramenta GBA.

As interações disponíveis entre a ferramenta e seus usuários são realizadas através de uma interface de linha de comando e uma interface disponível na WEB, conforme ilustrado na Figura 3.4 e Figura 3.5. Os usuários podem configurar parâmetros referentes aos objetos monitorados, escolha de tipo e período referente ao *baseline* que será gerado, parâmetros referentes ao sistema de alarmes e referentes aos testes analíticos disponibilizados pela ferramenta GBA.

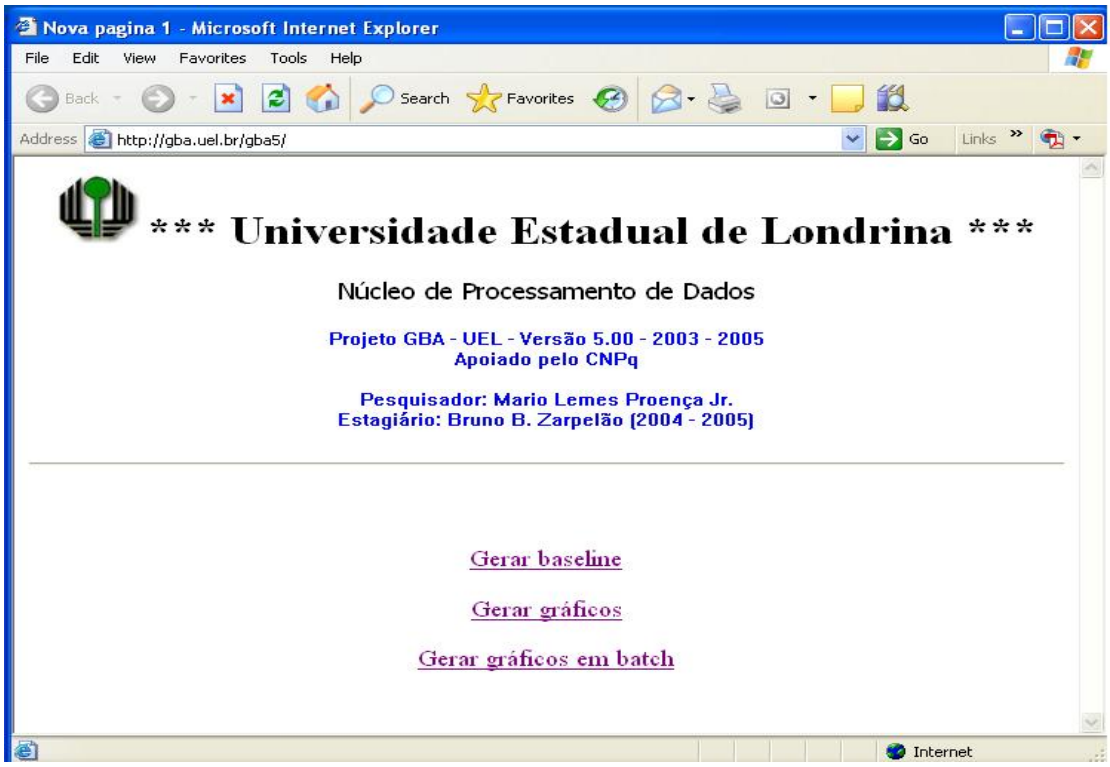


Figura 3.4 - Interface da Ferramenta GBA.

A Figura 3.5 apresenta um exemplo da interface utilizada para geração de gráficos que tem por objetivo reportar tanto o movimento real coletado através do módulo Lê-*switches* como seu *baseline*/DSNS esperado. Como observado nesta figura, a interface oferece opções para:

1. Tipo do gráfico: somente o *baseline*, o movimento ou ambos;
2. Opção para escolha de data inicial e final para apresentação;
3. Intervalo de precisão do gráfico que pode variar entre dias, horas, minutos e até segundos;
4. Escolha do idioma inglês ou português;
5. Inclusão ou não dos alarmes gerados pelo sistema de alarmes;
6. Tipo do *baseline bl-7* ou *bl-3*;
7. Visualização por média ou por picos do movimento ocorrido.

Figura 3.5 - Interface do módulo Gera_Gráficos da ferramenta GBA

3.4 Modelo BLGBA

3.4.1 Coleta das informações

O modelo BLGBA foi desenvolvido para realizar análises estatísticas em dados históricos coletados das MIBs pertencentes aos agentes SNMP residentes nos equipamentos de rede. Este monitoramento é realizado segundo a segundo durante o dia todo.

Optou-se pelo monitoramento segundo a segundo, justamente por se buscar preservar as características de cada momento, levando-se em conta uma granularidade mínima e aceitável em função do *overhead* gerado. Apesar de alguns estudos buscarem justamente a redução do custo de *pooling* na rede decorrentes de necessidades gerenciais, como o apresentado em , concluí-se ser mais importante preservar a precisão no monitoramento em detrimento do custo associado referente a utilização de banda.

Em (Proença, 2001), é apresentado o resultado de uma pesquisa onde foi verificado que gráficos gerados a partir de monitoramento do tráfego de rede, realizado com amostras SNMP coletadas em intervalos de 5 minutos, por exemplo, tendem a ocultar informações importantes referentes ao movimento real que flui no segmento analisado. O que normalmente ocorre é a perda de precisão e exatidão da informação apresentada. Softwares como MRTG (MRTG, 2005) e a própria ferramenta GBA são utilizados para coleta e monitoramento de segmentos de rede, apresentando na maioria das vezes gráficos gerados a partir de monitoramentos realizados com intervalo de amostragem da ordem de 5 minutos. A informação apresentada neste tipo de análise normalmente é resultante de uma média em função do tempo de consulta, as Figuras 3.1 e 3.2 são exemplos deste tipo de gráficos. O estudo realizado em 2001 demonstrou que análises que utilizam médias para monitoramento de redes frequentemente apresentam diferenças de até 200% em relação ao ocorrido na realidade.

A largura de banda necessária para o monitoramento de cada objeto SNMP, realizada segundo a segundo com a ferramenta GBA, é muito pequena. Cada consulta

resulta em um pacote UDP de aproximadamente 96 bytes. Este valor foi considerado pouco significativo levando-se em conta um *link* Ethernet de 100 Mbit/s, utilizado nas redes onde os testes foram realizados. Obviamente que com o aumento da quantidade de objetos e segmentos monitorados, este valor deve ser reconsiderado em função do benefício oferecido pela ferramenta. No estudo que é apresentado neste trabalho, optou-se pela amostragem segundo a segundo justamente pela possibilidade de aumentar a precisão nas avaliações do *baseline* e mesmo no sistema de alarmes e detecção de anomalia.

3.4.2 Definição do modelo

Com intuito de preservar características particulares de cada dia da semana, optou-se por criar dois tipos de *baseline*/DSNS. O primeiro é chamado de ***bl-7*** e consiste de sete arquivos de *baseline*/DSNS, um para cada dia da semana. O segundo é chamado de ***bl-3***, consiste de 3 arquivos de *baseline*/DSNS, um para os dias úteis que corresponde de segunda a sexta feira, um para o sábado e outro para o domingo. A Figura 3.6, ilustra em forma gráfica os dois modelos de *baseline* gerados pelo BLGBA. Esta divisão foi criada com intuito de tornar o modelo mais maleável e adaptável a situações que requerem diferenciações entre os dias.

A opção por separar o *baseline* dos dias úteis em relação ao dos finais de semana, foi realizada com objetivo de minimizar a margem de erro no resultado final apresentado. Maxion (Maxion, 1993), já havia comprovado o que os nossos resultados práticos também demonstraram, que a variação de tráfego existente entre um dia útil e os finais de semana tende a ser muito grande. No caso da Universidade Estadual de Londrina chega a ser superior a 200 %, conforme pode ser observado na Figura 3.9 (a) e (g) que demonstram o movimento e seu respectivo *baseline*/DSNS referente a sábado e domingo em relação às Figuras 3.9 (b),(c),(d),(e) e (f) que especifica o movimento e seu respectivo *baseline*/DSNS para segunda, terça, quarta, quinta e sexta feira. Na Figura 3.10 (a) e (g) em relação a

Figura 3.10 (b),(c),(d),(e) e (f) também pode ser observado a mesma diferença. Este foi o motivo que nos levou a criar os *baselines* independentes para cada dia da semana.

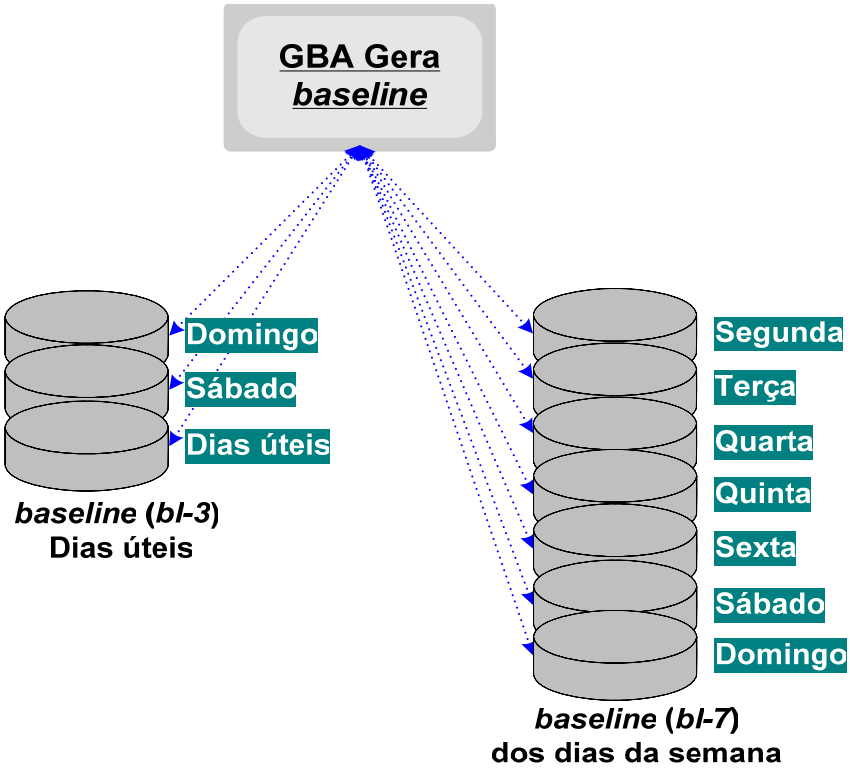


Figura 3.6 - modelo de *baseline* ***bl***-7 e ***bl***-3.

A Figura 3.9 e a Figura 3.10 mostram uma semana inteira do dia 03/04/2005 a 09/04/2005 contendo o movimento real e os respectivos *baselines* gerados para os modelos ***bl-3*** e ***bl-7*** respectivamente. Neste caso, pode ser observado que para os finais de semana, tanto o *baseline* gerado pelo modelo ***bl-3*** como o gerado pelo ***bl-7*** apresentam o mesmo resultado, ou seja, o *baseline* para sábado e para domingo serão os mesmo quando gerados para o modelo ***bl-3*** e ***bl-7***. Contudo, para os dias úteis de segunda a sexta feira o resultado dos *baselines* gerados a partir do modelo ***bl-7*** procura preservar as características particulares de cada dia. No modelo ***bl-3*** é gerado um único *baseline* para todos os dias úteis. As figuras 3.9 e 3.10 mostram um exemplo completo da aplicação dos dois modelos para uma mesma semana. Nota-se que o *baseline* da Figura 3.9 é o mesmo em todos os dias úteis da semana, de segunda a sexta, enquanto os *baselines* dos dias úteis apresentados na

Figura 3.10 demonstram pequenas diferenças entre os dias de segunda a sexta feira. Neste exemplo específico apresentado, tanto a utilização do modelo ***bl-7*** como o do ***bl-3*** se mostram adequados para serem utilizados, porém existem situações onde isso não será possível, devendo o administrador da rede optar por qual tipo de *baseline* seja mais adequado.

O Modelo BLGBA desenvolvido neste trabalho, executa análises estatísticas sobre os valores coletados dos objetos SNMP, respeitando o exato momento da coleta, segundo a segundo, durante as 24 horas do dia. A idéia principal é que o *baseline* deverá preservar as características do tráfego, levando em conta as variações temporais ao longo do dia.

Para geração do *baseline*/DSNS, optou-se por não incluir períodos sazonais, como férias e greves, que influenciam nas amostras em conseqüência da baixa utilização da rede nesses períodos. Além disso, no cálculo para geração do *baseline*/DSNS, também são excluídos erros que normalmente ocorrem no monitoramento de objetos SNMP, devido a problemas de congestionamento, perdas de pacotes ou mesmo falha no agente SNMP. A ferramenta GBA realiza coleta de informações, segundo a segundo, do agente SNMP residente no equipamento de rede monitorado, com isso são esperadas que 86.400 amostras sejam coletadas por dia. Observou-se que normalmente ocorrem problemas que acabam por inviabilizar pelo menos 0,05 % das amostras ao longo do dia. Os testes foram realizados em diversos tipos de equipamentos como *switches*, roteadores e servidores de diversos fabricantes e constatou-se a persistência deste problema referente a erros ou perdas de amostras ao longo do dia, por isso eles foram considerados como *outliers* das amostras.

O processamento para geração do *baseline/DSNS* é feito inicialmente em batch visando sua criação com dados referentes a um período pré-estabelecido. O *baseline*/DSNS é gerado segundo a segundo para um período de dias representados por **N**, que compõe o conjunto n_j (j = 1, 2, 3, 4, ..., *N*). Com a coleta diária tem-se um conjunto de amostras do dia, representadas por a_i (i = 1, 1, 2, ..., 86400), com isto construiu-se a matriz bidimensional com 86400 linhas e ***N*** colunas, que deve ser previamente ordenada e que será representada por $M_{ij.} = a_{1,}n_1$, conforme ilustrado na Figura 3.**7**.

$$M_{ij} = \begin{pmatrix} a_{00001},n_1,\ a_{00001},n_2,\ \ldots,\ a_{00001},n_{n-1},\ a_{00001},n_n \\ a_{00002},n_1,\ a_{00002},n_2,\ \ldots,\ a_{00002},n_{n-1},\ a_{00002},n_n \\ a_{00003},n_1,\ a_{00003},n_2,\ \ldots,\ a_{00003},n_{n-1},\ a_{00003},n_n \\ a_{00004},n_1,\ a_{00004},n_2,\ \ldots,\ a_{0004},n_{n-1},\ a_{00004},n_n \\ \ldots \\ a_{86400},n_1,\ a_{86400},n_2,\ \ldots,\ a_{86400},n_{n-1},\ a_{86400},n_n \end{pmatrix}$$

Figura 3.**7** - Matriz de escolha do *Bli.*

O algoritmo utilizado para o cálculo do *baseline*/DSNS é baseado em uma variação do cálculo da moda, que originalmente é definida como a amostra mais freqüente de um conjunto de amostras observadas (Bussab, 2003). O algoritmo leva em consideração a freqüência das classes inferiores bem como da classe modal. O cálculo leva em consideração a distribuição dos elementos em freqüências com base na diferença entre o maior G_{aj} e o menor S_{aj} elemento da amostra, utilizando somente 5 classes modais. Esta diferença dividida por cinco, forma a amplitude *h* entre as classes, $h = \left(\frac{G_{aj} - S_{aj}}{5} \right)$,

onde h = amplitude, G_{aj} = maior elemento, S_{aj} = menor elemento da amostra analisada.

Com isto, obtém-se os limites de cada classe modal L_{Ck}, que são calculados conforme $L_{ck} = S_{aj} + h * k$, onde L_{Ck} é o limite da classe modal *k; S_{aj}* é o menor elemento da amostra analisada; *h* é a amplitude; *k* é a classe modal e *Ck* representa a classe *k* que pode assumir valores (***k* = 1..5**). A Figura 3.8 ilustra de forma gráfica a execução do algoritmo BLGBA para escolha do Bl_i, onde são geradas 5 classes modais e alojados os elementos de cada linha da matriz ordenada $M_{ij.}$.

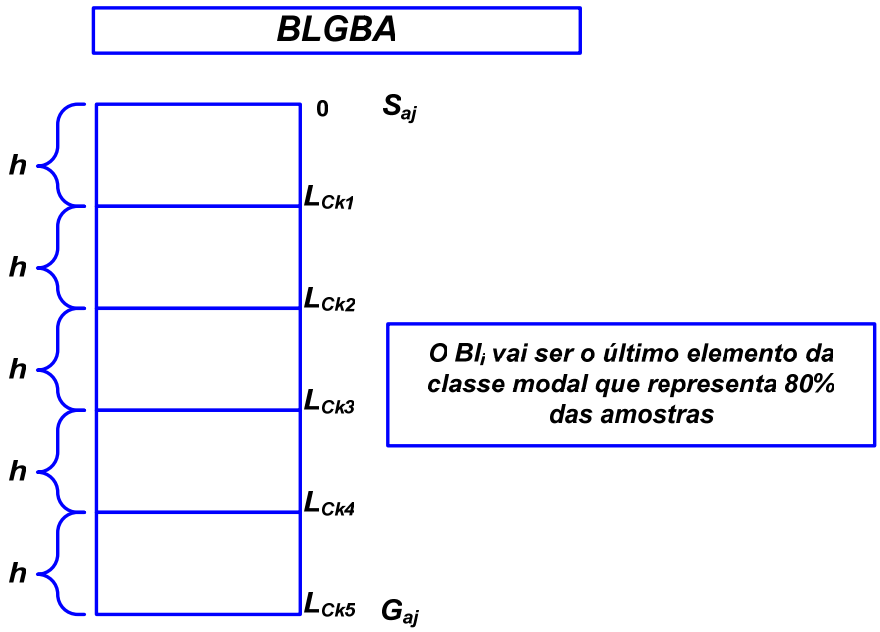


Figura 3.8 - Diagrama de execução do BLGBA

O objetivo do algoritmo BLGBA para o cálculo do *baseline* de cada segundo Bl_i se destina a obter o elemento que representa 80 % das amostras analisadas. O Bl_i será definido como o maior elemento inserido na classe, com freqüência acumulada maior ou igual a 80%. O objetivo é obter o elemento que se situasse acima da maioria das amostras, respeitado o limite de 80 %. Este processo é utilizado para a geração dos *baselines* modelo ***bl-7*** **e** ***bl-3.***

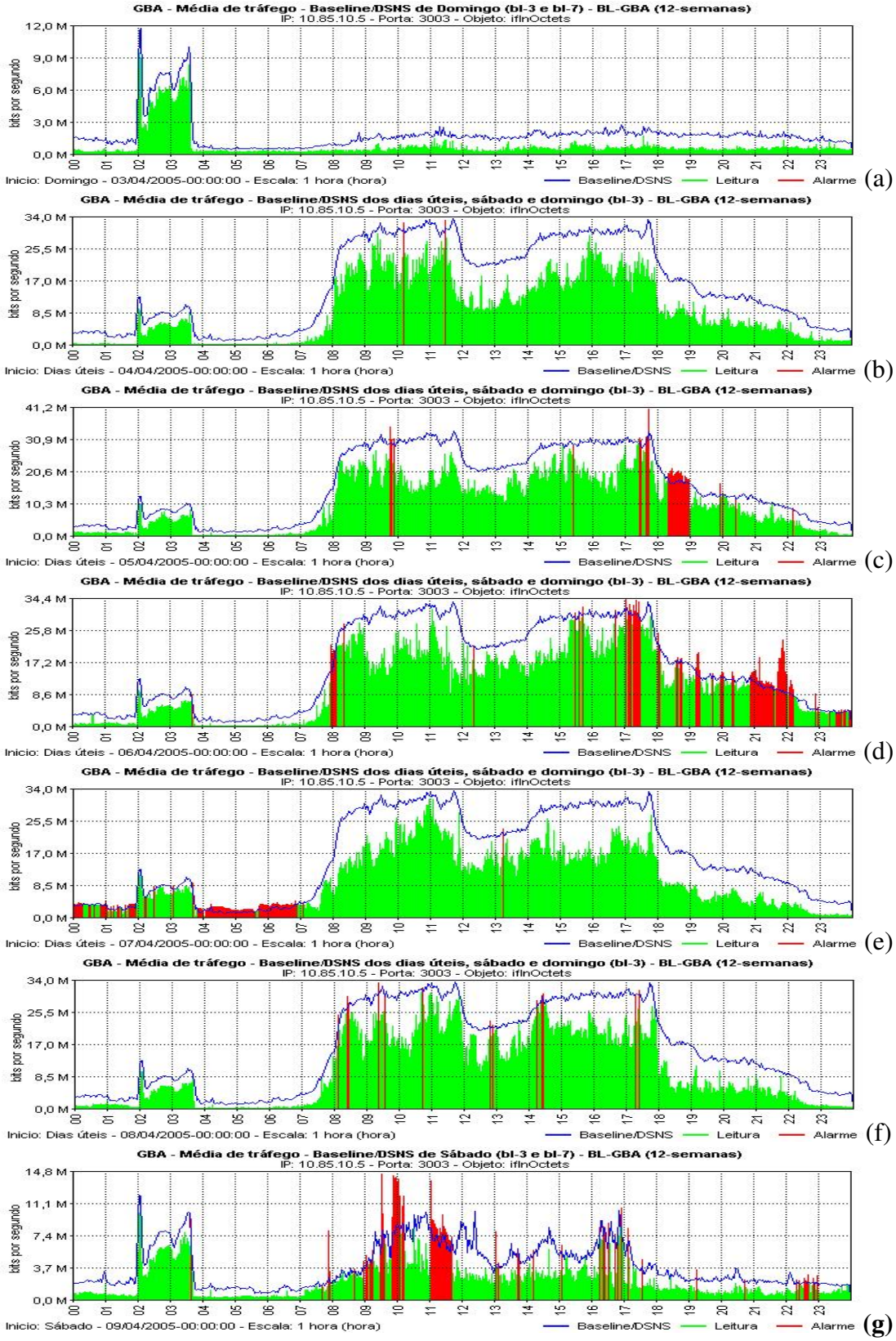


Figura 3.9 - *Baseline* e o movimento real para o segmento S_4 para a semana de 03/04/2005 a 09/04/2005 utilizando *bl-3*.

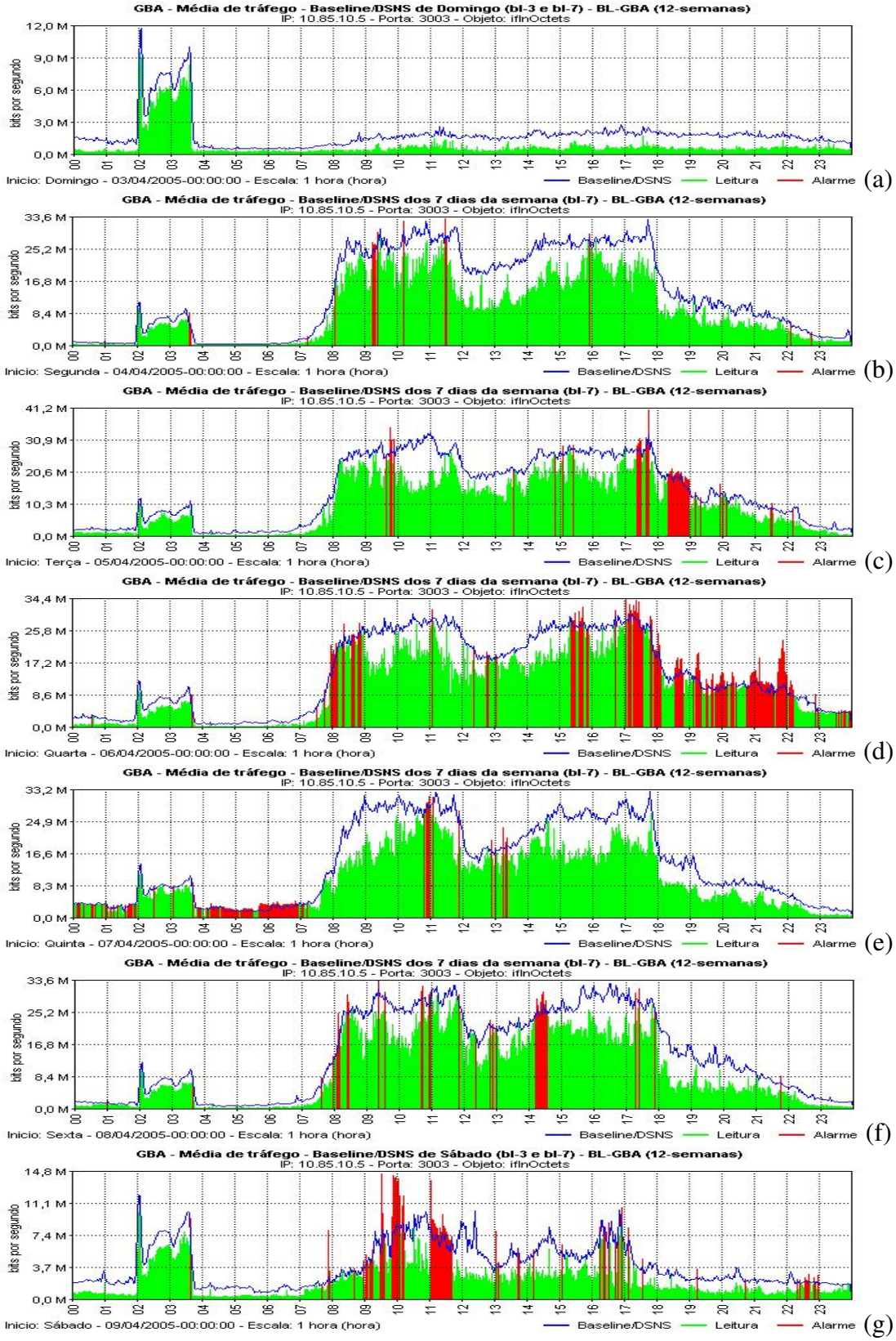


Figura 3.10 - *Baseline* e o movimento real para o segmento S_4 para a semana de 03/04/2005 a 09/04/2005 utilizando *bl-7*.

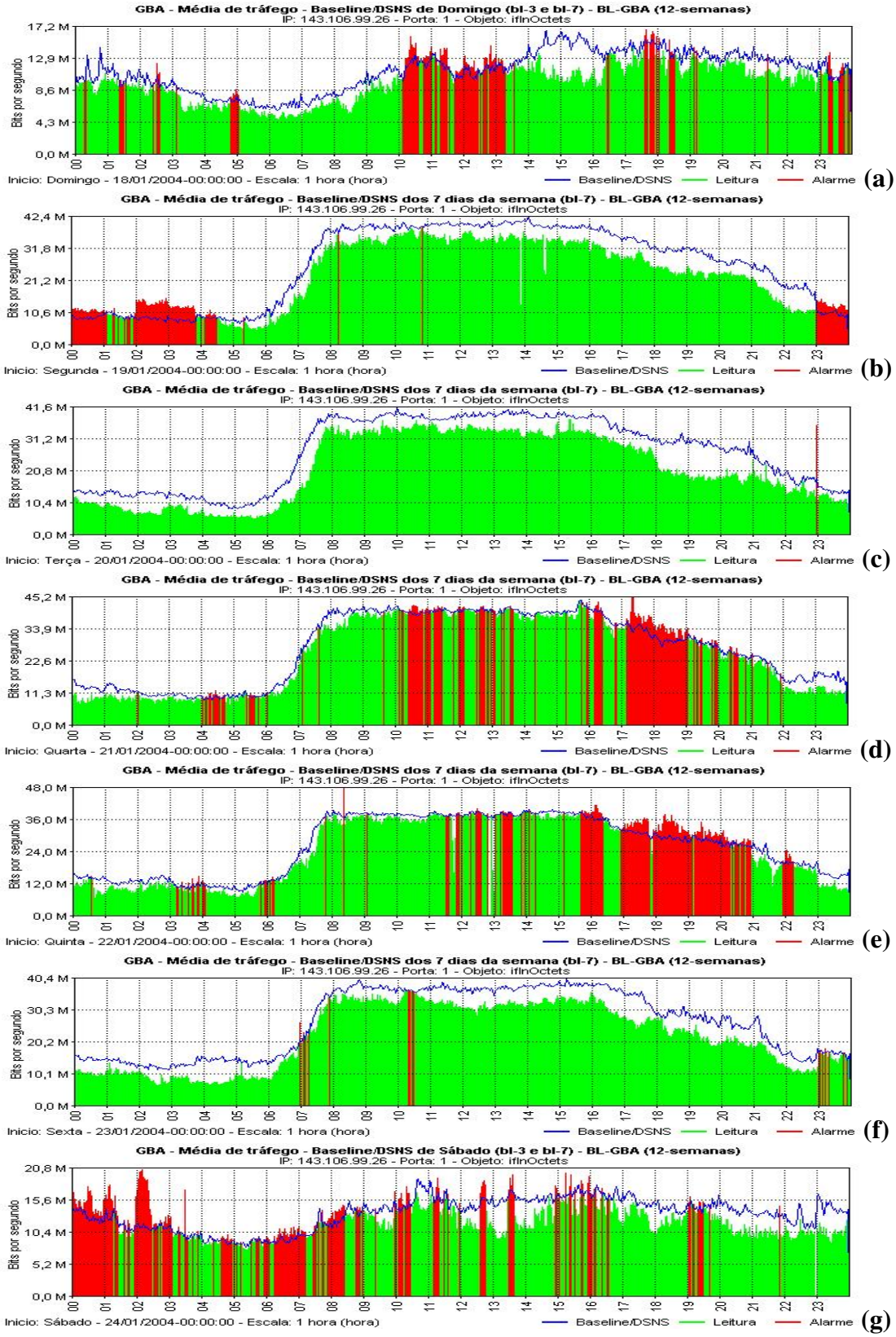


Figura 3.11 - Movimento real e respectivo *baseline* para semana de 18 a 24/01/2004 referentes ao segmento S_6.

A definição do algoritmo para cálculo do BLGBA, foi realizada após inúmeros testes com outros modelos estatísticos, baseados em média, moda e avaliação dos percentis que se concentraram no octil e na média do último decíl. Após a realização de simulações com dados reais coletados através da ferramenta GBA, observou-se que uma grande maioria das amostras se concentrava acima do octil, ou seja, na faixa que representa de 80 a 100 % do valor nominal referente às amostras analisadas. Este fato norteou à criação do algoritmo BLGBA, que baseado no cálculo da moda, tem como objetivo obter o elemento que representa 80% das amostras analisadas para ser o representante do *baseline* do segundo *i*.

Com objetivo de validar e verificar a confiabilidade do algoritmo proposto para o BLGBA, foram realizados alguns testes analíticos com os diversos modelos citados acima, que corroboraram para a escolha definitiva do BLGBA. O que se buscou foi a comparação entre os *baselines* resultantes dos diversos modelos, para verificar qual a melhor opção. Os resultados dos testes podem ser sintetizados como segue:

1. Análise visual dos gráficos gerados que continham o *baseline*/DSNS e o movimento real do dia. Segundo , a análise visual assim como todos os testes estatísticos oferecem resultados aproximados. Este tipo de análise proporciona duas vantagens em relação aos testes numéricos: A primeira, diz respeito à facilidade no entendimento e na tomada de decisões proporcionadas justamente pela forma de apresentação. A segunda, que este tipo de análise oferece mais informações do que um simples resultado positivo ou negativo obtido a partir de um teste. A velha retórica de que uma imagem vale mais do que mil palavras se aplica neste caso. A Figura 3.9, Figura 3.10 e Figura 3.11 ilustram exemplos reais de situações que demonstram o bom ajuste entre o previsto pelo *baseline* gerado pelo BLGBA e o movimento real. Os testes realizados com *baselines* gerados a partir da média e moda simples, apresentavam previsões muito abaixo do movimento real enquanto os *baselines* gerados por algoritmo baseado em octil e média do decil, apresentavam previsões muito acima do movimento real. A Figura **3.12**, demonstra o resultado dos estudos para avaliação dos diversos modelos estudados para a geração do *baseline*. Nela podem ser visualizados os

resultados dos *baselines* gerados a partir da média, moda, octil, média do decil e do algoritmo proposto neste trabalho, que chamamos de BLGBA. As linhas demonstradas na figura são referentes ao conjunto de amostras utilizadas para escolha do *baseline*. Como pode ser observado, os resultados tanto do octil como da média do decil se aproximam muito dos gerados pelo modelo BLGBA, porém eles apresentam uma característica de acomodação vertical, na escolha do elemento que será definido como o *baseline* das amostras. Esta acomodação nos levou a constatar que os *baselines* gerados para estes modelos (octil e média do decil), tendem a gerar uma tendência de manter picos ocorridos no tráfego que não se mostraram adequados nas análises e testes realizados.

2. Análise dos desvios proposta por Bland e Altman , que leva em consideração a diferença entre o movimento predito e o observado. Estas diferenças devem estar dentro do intervalo definido por $\overline{d} \pm 2 * s$, onde $\overline{d}$ é a diferença entre a média e o s é o desvio padrão entre as diferenças. A proposta de Bland e Altman estabelece um limite superior e inferior onde os desvios devem estar contidos. Para considerar o bom ajuste do modelo, ele deve apresentar 95 % das diferenças estudadas entre estes limites. Nos casos estudados, o BLGBA foi o modelo que apresentou melhor ajuste segundo este teste. Resultados de testes do modelo BLGBA utilizando a análise dos desvios proposta por Bland e Altman serão apresentados posteriormente neste trabalho.

3. Regressão linear entre o movimento real e o predito pelo *baseline* gerado para os modelos estudados. Novamente o modelo que apresentou melhores resultados foi o BLGBA. A Figura 3.13 demonstra o resultado do teste de regressão realizado para o segmento $\boldsymbol{S_4}$ nos meses de setembro a novembro de 2003. Nesta figura pode observar-se que o modelo BLGBA apresenta melhor coeficiente de correlação entre o movimento real e o *baseline* predito.

4. Análise dos resíduos: Este cálculo foi realizado para os modelos estudados durante a escolha do algoritmo para cálculo do *baseline*. Novamente constatamos que os resultados apresentados pelo BLGBA foram os que melhor se apresentaram para a escolha. A Figura 3.14, demonstra o resultado de uma análise de resíduos, realizadas para o BLGBA no segmento S_4. Os resultados não demonstram nenhum tipo de tendência, indicando o bom ajuste do modelo.

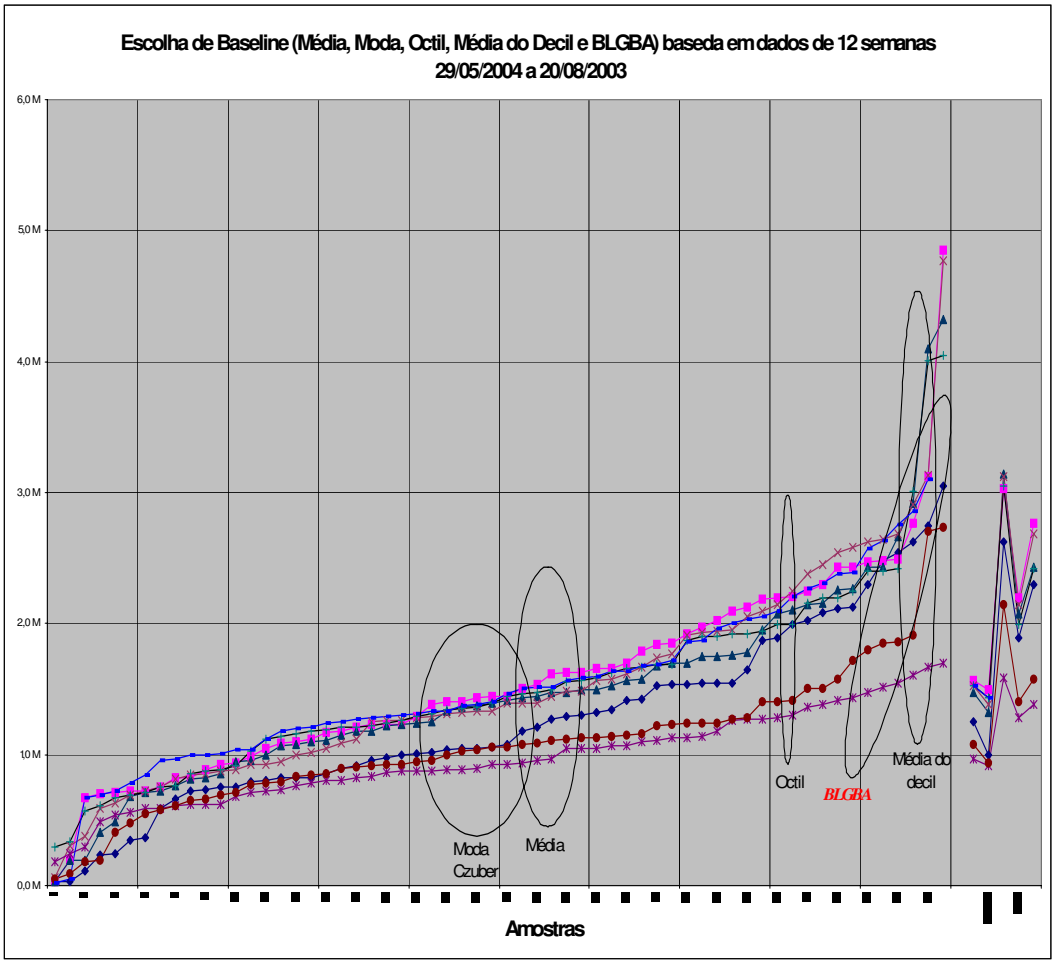


Figura 3.12 - Avaliação dos diferentes modelos para geração de *baselines.*

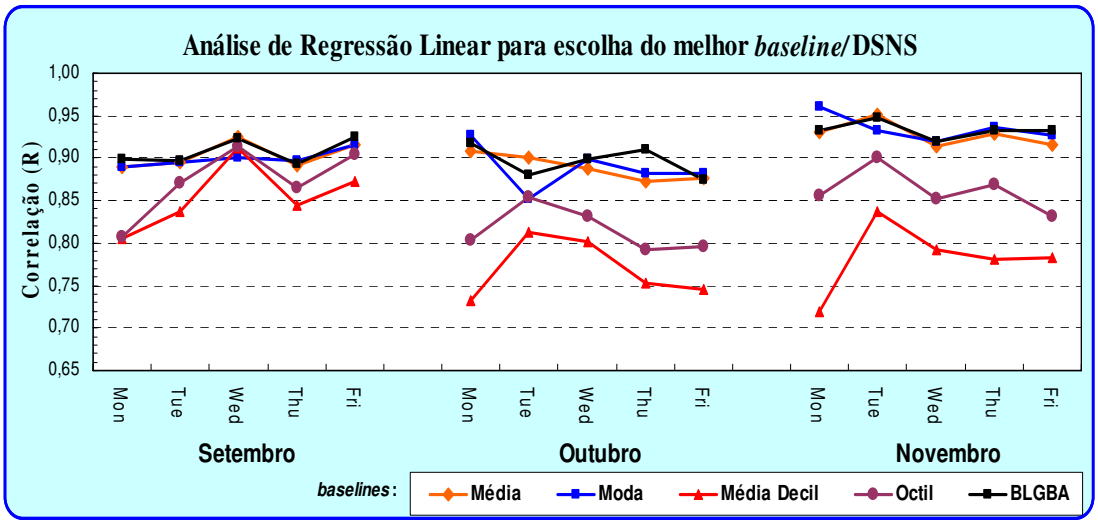


Figura 3.13 - Análise de Regressão dos diferentes métodos para geração do *baseline*/DSNS.

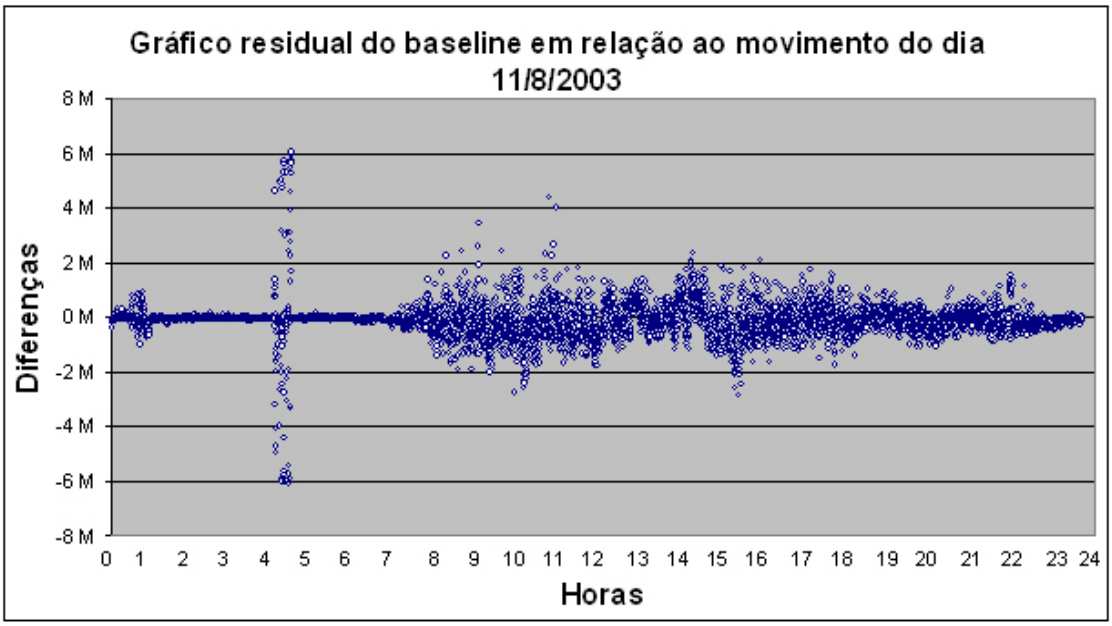


Figura 3.14 - Análise de resíduos para o segmento s4 em 11/8/2003.

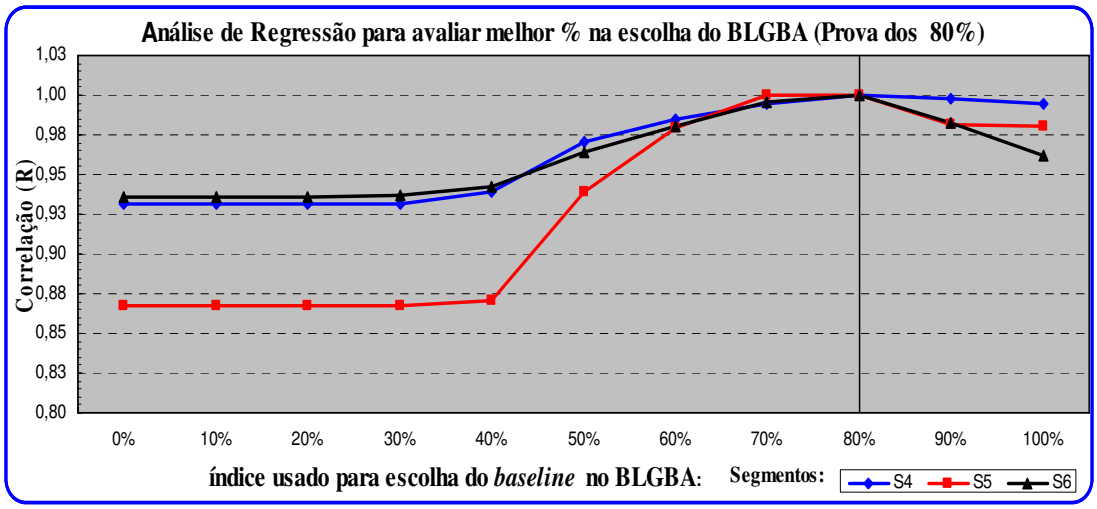


Figura 3.15 - Análise de Regressão para validar escolha do índice de escolha para BLGBA (prova 80%).

O processo para escolha do índice de representatividade do elemento que representa 80 % das amostras analisadas para ser escolhido como o *baseline*/DSNS $\boldsymbol{Bl_i}$, no segundo ***i*** de um dia, foi realizado inicialmente de forma empírica, e com base em análises visuais, efetuadas com índices variando de 0 a 100 % das amostras. Além da análise visual, também foram realizados testes analíticos, através de regressão linear entre o movimento e o *baseline*/DSNS, com objetivo de verificar e validar a escolha do elemento, que representava 80 % das amostras para o modelo BLGBA na escolha do elemento que deveria representar o $\boldsymbol{Bl_i}$. A Figura 3.15 demonstra o coeficiente ***R*** de correlação entre o *baseline*/DSNS e o movimento real, utilizando índice de escolha para o BLGBA, variando de 0 a 100 %. Estes testes foram realizados para os segmentos S4, S5 e S6 durante 6 meses do ano de 2003. Os testes de regressão, assim como os de análises visuais realizados, comprovaram que a escolha do elemento que representa 80 % das amostras para o algoritmo do BLGBA foi o mais adequado.

Inicialmente, o modelo foi construído somente para analisar o volume de bits que entram e saem no segmento analisado, que fica armazenado no objeto *ifInOctets* e *ifOutOctets* pertencentes ao grupo *Interface* da MIB-II (RFC 1213, 1991) residente nos agentes SNMP dos equipamentos de rede. Porém, após a consolidação do modelo e a comprovação dos resultados obtidos, os testes foram estendidos para os seguintes objetos:

número total de datagramas recebidos pela interface (*ipInReceives*) e o número total de segmentos TCP recebidos pela interface de rede (*tcpInSegs*).

Definição dos segmentos e respectivos objetos SNMP utilizados para os testes realizados nesta tese:

1. Primeiro segmento estudado, chamado de S_1 foram analisados os objetos *ifInOctets*, *ifOutOctets* e *ipInReceives*.
2. Segundo segmento, onde foi realizada avaliação, chamado de S_2, foram utilizados os objetos *ifInOctets*, *ifOutOctets* e *ipInReceives* e *tcpInSegs*.
3. Terceiro segmento estudado, chamado de S_3 foram utilizados os objetos *ifInOctets*, *ifOutOctets* e *ipInReceives* e *tcpInSegs*.
4. Quarto segmento, chamado de S_4, foram utilizados os objetos *ifInOctets*, *ifOutOctets* e *ipInReceives*.
5. Quinto segmento estudado, chamado de S_5, foram analisados os objetos *ifInOctets*, *ifOutOctets.*
6. Sexto segmento, S_6, foram realizadas análises utilizando os objetos *ifInOctets*, *ifOutOctets.*

3.5 Avaliação do modelo BLGBA

Na avaliação do modelo BLGBA foram realizados vários testes analíticos, com o objetivo de demonstrar a eficácia e confiabilidade dos *baselines* gerados. Inicialmente, foi criado um teste para avaliar o coeficiente de variação de um *baseline* de um mês em relação a outro. Este índice foi chamado de IVBL (índice de variação do *baseline*). O IVBL é calculado com base na diferença média entre um *baseline* e outro, de acordo com a expressão (3.3). A utilização do IVBL possibilitou observar que existiu uma pequena, porém positiva variação no volume de tráfego nos segmentos analisados, durante o andamento desta pesquisa. A Tabela **3.1** ilustra a variação percentual de janeiro de 2003 a janeiro de 2004, no segmento S_4, que interliga ao roteador principal da rede da UEL. Nos outros segmentos analisados, também foi encontrado um pequeno percentual de variação, indicando um crescimento no volume de tráfego. Entende-se que este cálculo não é totalmente conclusivo, por haver a possibilidade de ocorrer alterações positivas e negativas em períodos distintos, decorrentes de fatores sazonais. Entretanto, o objetivo que se procurava atingir, era justamente avaliar um indicativo da tendência do volume de tráfego, retratado no *baseline* de um mês em relação ao *baseline* de outro.

$$IVBL = \left(\sum_{i=1}^{86400} BL'_i - BL''_i \right) \Big/ 86400 \qquad (3.3)$$

Onde *IVBL* = índice de variação de um *baseline* em relação a outro.

Tabela 3.1 - Variação do *baseline* de Janeiro 2003 a Janeiro de 2004 no segmento S_4.

% de crescimento do baseline em relacao ao mês anterior													
	jan/03	fev/03	mar/03	abr/03	mai/03	jun/03	jul/03	ago/03	set/03	out/03	nov/03	dez/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%

O IVBL também foi utilizado para calcular o percentual de variação de um *baseline* gerado para ***n*** semanas, comparado com outro *baseline* gerado para (***n*** - 1) semanas. Estes cálculos utilizando *baselines* semanais, foram realizados com o objetivo de avaliar qual a quantidade mínima de semanas de amostras necessárias para o cálculo do *baseline*. Neste caso, o objetivo era observar a partir de qual semana não haveria mudanças significativas que pudessem contribuir para formação do *baseline*.

Durante 24 semanas realizou-se a comparação do *baseline* de ***n*** semanas com o *baseline* de (***n*** – 1) semanas. Este teste consiste em calcular um *baseline* para ***n*** semanas e outro para (***n*** – 1) semanas, com objetivo de avaliar o percentual de variação médio entre eles, sendo que ***n*** irá variar de 24 a 1. Comparou-se:

1. O *baseline* de 24 semanas com o de 23 semanas;
2. O *baseline* de 23 semanas com o de 22 semanas;
3. O *baseline* de 22 semanas com o de 21 semanas;
4. ... ...
5. O *baseline* de 4 semanas com o de 3 semanas;
6. O *baseline* de 3 semanas com o de 2 semanas;
7. O *baseline* de 2 semanas com o de 1 semana;

Com este cálculo, constatou-se que o percentual de variação entre os *baselines* tende a se estabilizar a partir da 12ª semana, não sendo mais significativo para o cálculo do *baseline* as variações a partir deste resultado.

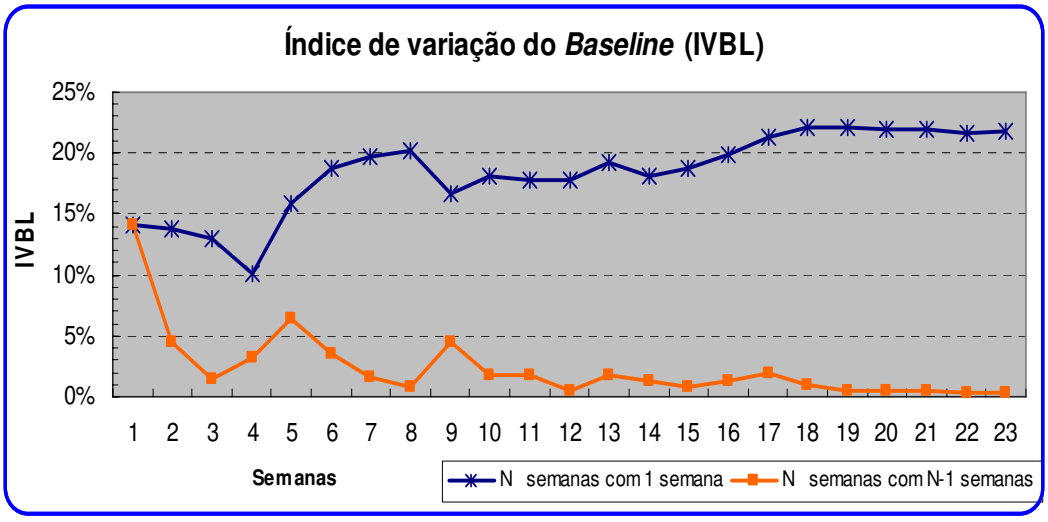


Figura 3.16 - Porcentagem de variação entre o *baseline* de n semanas com o de n-1 semanas e com o de 1 semana.

O cálculo inverso também foi realizado, onde se comparava o *baseline* de 1 semana com o *baseline* de ***n*** semanas, sendo que ***n*** variava de 2 a 24. Neste caso, também pode-se observar, que a partir da 12ª semana o percentual de variação se estabiliza em torno de 20%. Demonstrando, portanto que não ocorrem mais variações significativas necessárias para serem incluídas no cálculo do *baseline*, a partir deste resultado. A Figura 3.16 demonstra os resultados das comparações do *baseline* de ***n*** semanas com o de uma semana e o de ***n*** semanas com o de (***n*** – 1) semanas.

Os resultados destes cálculos ajudaram na conclusão de que seriam necessárias 12 semanas para o cálculo do *baseline*. Após este estudo, adotou-se como padrão para esta pesquisa o número de 12 semanas para o cálculo dos *baselines*. Observou-se também que a utilização de mais ou menos 12 semanas podem ser consideradas para o cálculo do *baseline*; porém o número de 12 semanas mostrou-se mais adequado ao modelo proposto, considerando-se nossos objetivos iniciais. Outros testes analíticos apresentados a seguir, mostram a validade do modelo BLGBA no quais os *baselines* foram calculados com 12 semanas de amostragem.

Além da análise visual entre o *baseline* e o movimento real, que demonstraram um bom ajuste do modelo ao proposto, foram realizados outros testes analíticos, visando avaliar a confiabilidade do modelo BLGBA para geração de *baseline*: regressão linear, teste proposto por *bland e Altman* e análise de resíduos.

3.5.1 Análise de Resíduos

Dentre os testes realizados, a análise dos resíduos (Bussab, 2003), , considerada como um teste simples, porém não menos importante, tem por objetivo demonstrar se o modelo proposto está adequado ou não. Neste caso, o grau de ajuste do modelo proposto é avaliado através de análise visual dos gráficos, gerados a partir das diferenças entre o que foi predito pelo modelo e o que realmente aconteceu no movimento real. Espera-se que os resíduos apresentados nos gráficos, estejam distribuídos de forma aleatória em torno do eixo ***0***. O objetivo é avaliar inadequações ou tendências neste gráfico, o que indicaria problemas no modelo.

Assim como testes estatísticos são resultados aproximados os visuais também o são. O que se espera, é que os gráficos resultantes não apresentem tendências como curvaturas, linearidade ou mesmo multilinearidades. A outra vantagem da análise visual dos gráficos é que eles disponibilizam mais informações do que um simples resultado gerado de forma numérica.

A Figura 3.17 e a Figura 3.19**,** demonstram o resultado da análise de resíduos para os dias 14 e 15/08/2003 do segmento S_4 da rede da UEL. Neste caso, pode-se observar não haver tendências, os resíduos se apresentam de forma uniforme em torno do eixo 0. Observamos também diferenças significativas às 00:30 e 04:00 horas que sinalizam horários de backup realizados nos servidores da rede da UEL.

A Figura 3.18 ilustra o movimento real ocorrido no dia 14/08 de 2003 para o segmento S_4. Nesta figura observa-se a ocorrência dos *backups* às 00:30 e 04:00 horas. A Figura 3.17 que apresenta os resíduos referentes ao movimento deste mesmo dia, também demonstra a indicação de resíduos no horário das 00:30 e as 04:00 horas. Na Figura 3.20 pode ser observado o *baseline*/DSNS e o movimento real referentes ao dia 15/08/2003. Esta figura se refere à mesma situação apresentada na Figura 3.19 só que na forma de resíduos.

A seguir são apresentados exemplos em forma de gráfico de resíduos, gerados a partir do cálculo de resíduo entre o *baseline* e o movimento real, para cada um dos

segmentos onde os testes foram realizados durante a realização deste trabalho. Os resultados encontrados não demonstraram tendências que indicassem impropriedades no modelo.

São apresentados exemplos referentes aos resíduos de um dia inteiro em conjunto com outra figura que retrata o movimento e seu respectivo *baseline*/DSNS. A Figura 3.21 demonstra a análise de resíduos, enquanto a Figura 3.22 demonstra o *baseline* e o movimento real para o segmento S_1. A Figura 3.23 demonstra a análise de resíduos, enquanto a Figura 3.24 demonstra o *baseline* e o movimento real para o segmento S_2. A Figura 3.25 demonstra a análise de resíduos, enquanto a Figura 3.26 apresenta o *baseline* e o movimento real para o segmento S_3. A Figura 3.27 demonstra a análise de resíduos enquanto a Figura 3.28 demonstra o *baseline* e o movimento real para o segmento S_5. A Figura 3.29 apresenta a análise de resíduos, enquanto a Figura 3.30 demonstra o *baseline* e o movimento real para o segmento S_6.

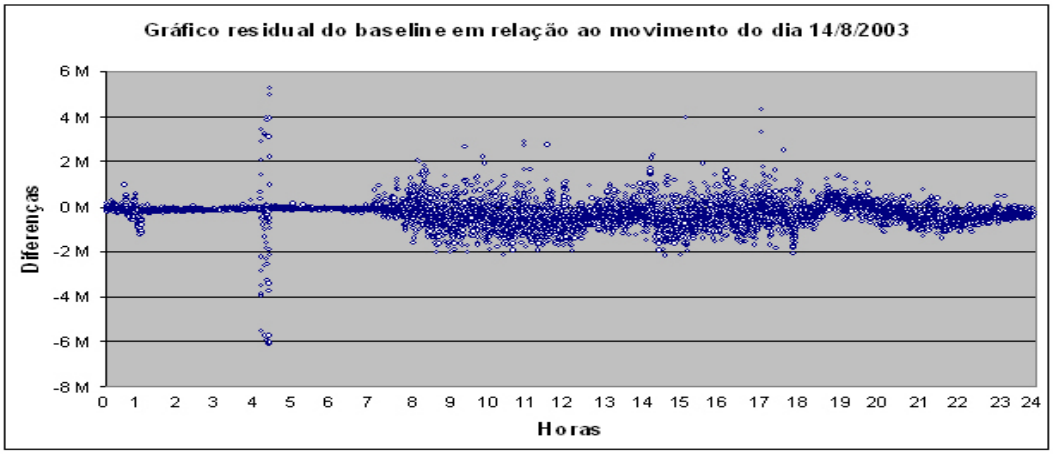


Figura 3.17 - Análise de resíduos para o segmento S_4 em 14/08/2003.

Figura 3.18 - Movimento ocorrido no dia 14/08/2003 no segmento S_4.

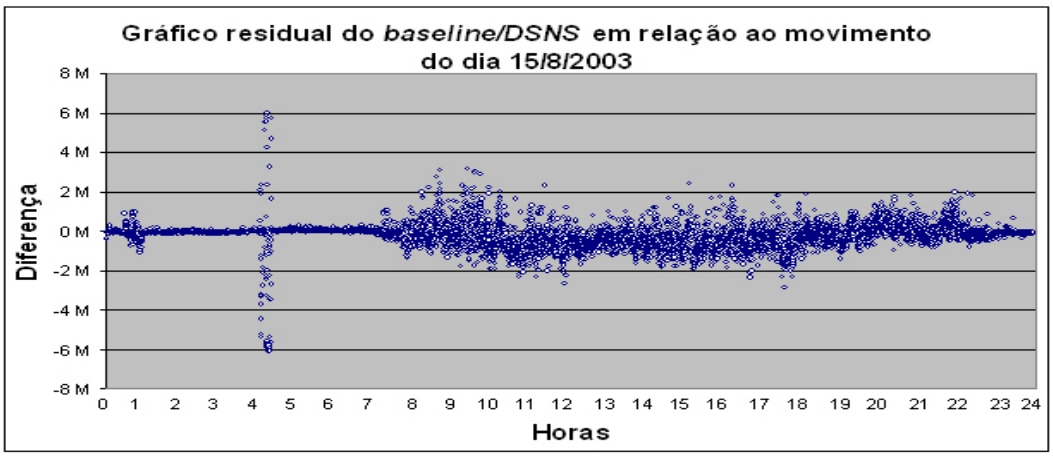


Figura 3.19 - Análise de resíduos para o segmento S_4 em 15/08/2003.

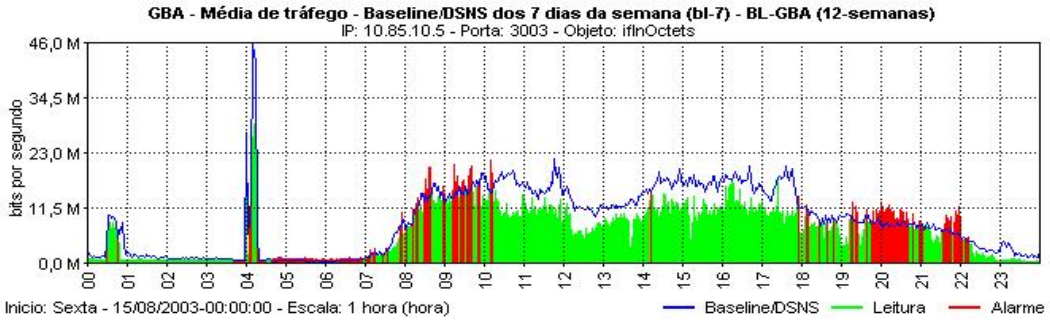


Figura 3.20 - Movimento e *baseline* do dia 15/08/2003 no segmento S_4.

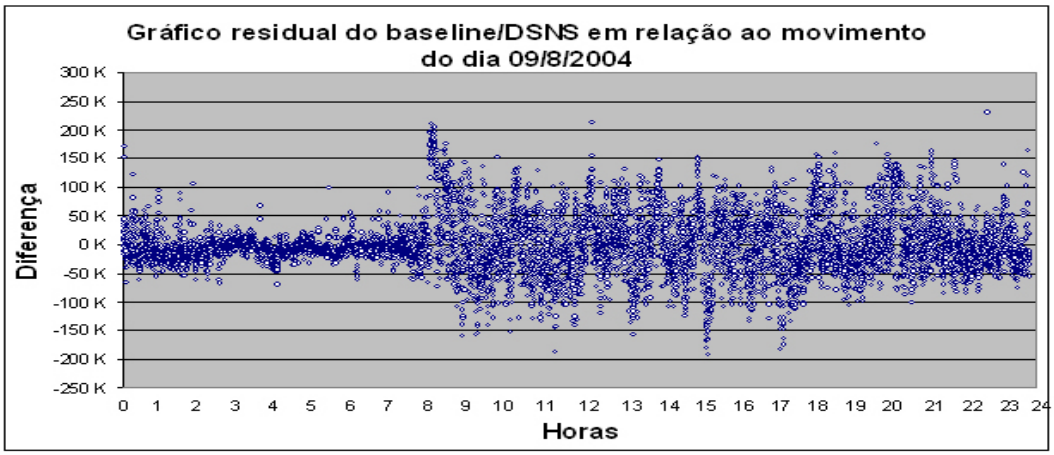


Figura 3.21 - Análise de resíduos para o segmento S_1 em 09/08/2004.

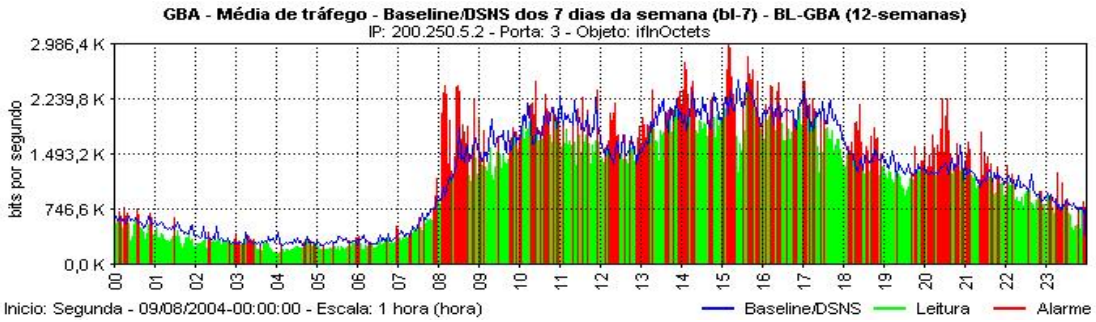


Figura 3.22 - Movimento e *baseline* do dia 09/08/2004 no segmento S_1.

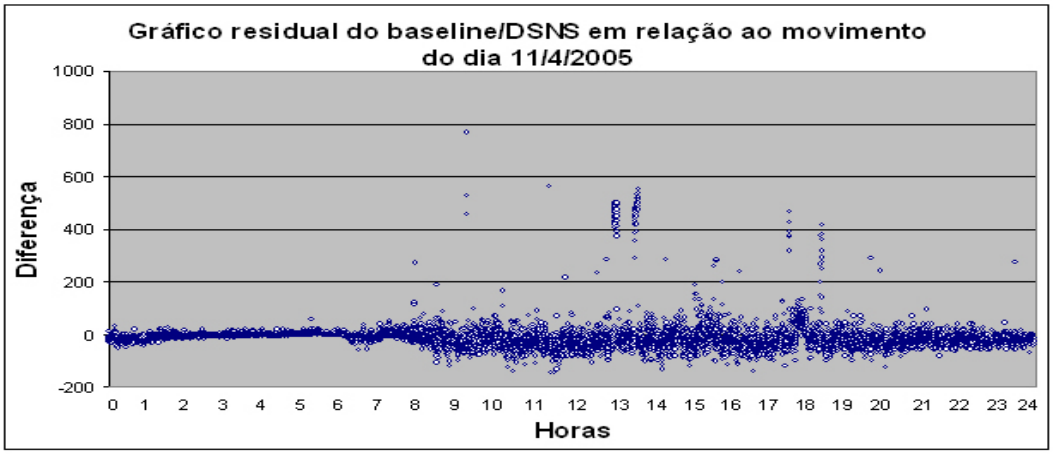


Figura 3.23 - Análise de resíduos para o segmento S_2 em 11/04/2005.

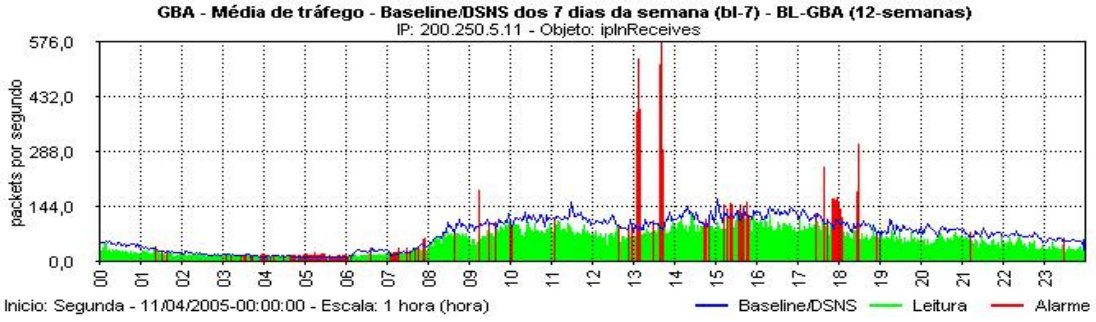


Figura 3.24 - Movimento e *baseline* do dia 11/04/2005 no segmento S_2.

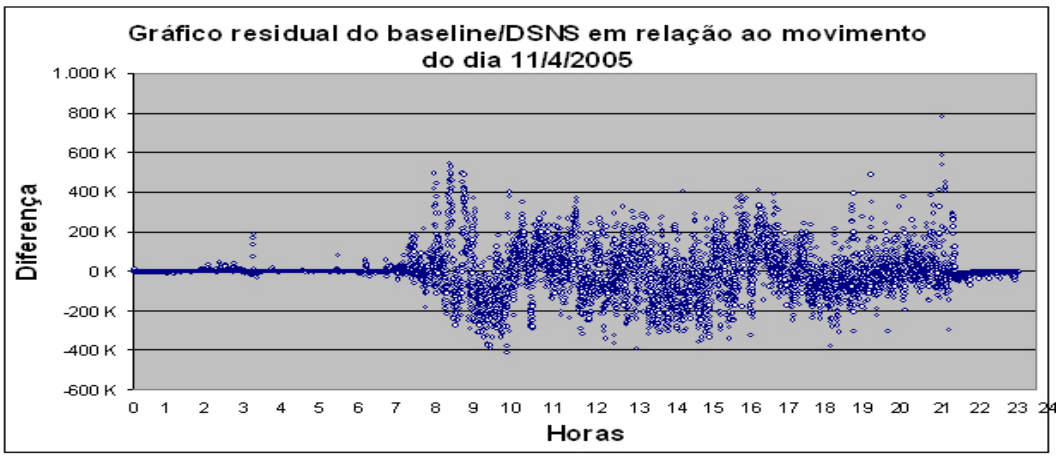


Figura 3.25 - Análise de resíduos para o segmento S_3 em 11/04/2005.

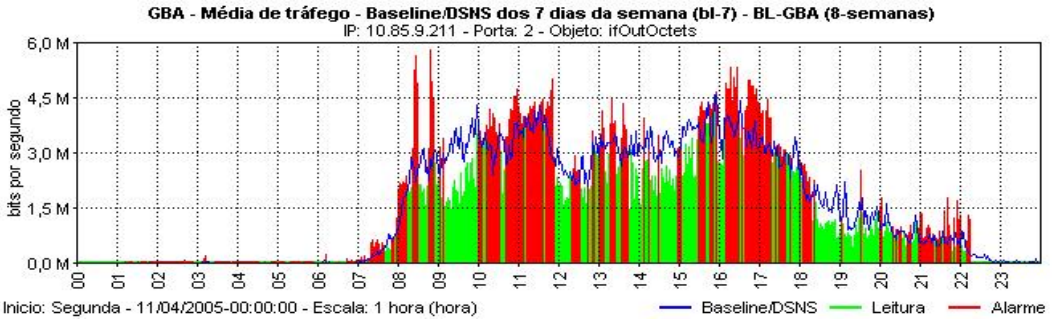


Figura 3.26 - Movimento e *baseline* do dia 11/04/2005 no segmento S_3.

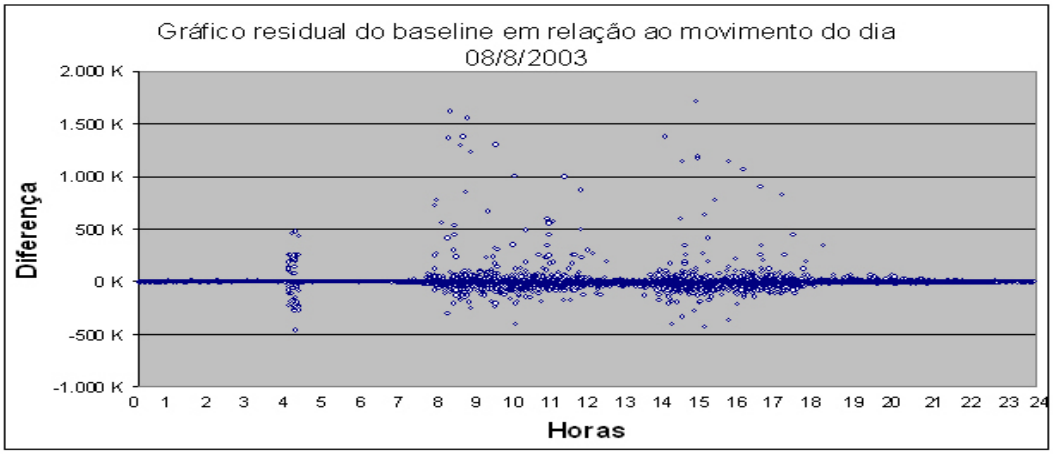


Figura 3.27 - Análise de resíduos para o segmento S_5 em 08/08/2003.

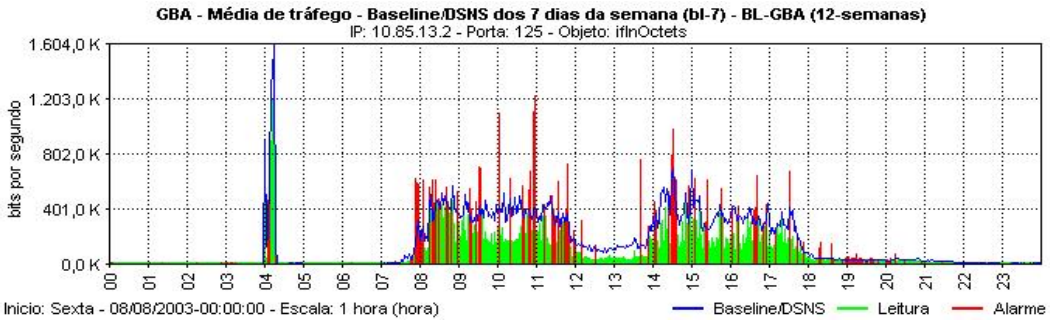


Figura 3.28 – *baseline* e movimento real do segmento S_5 em 08/08/2003.

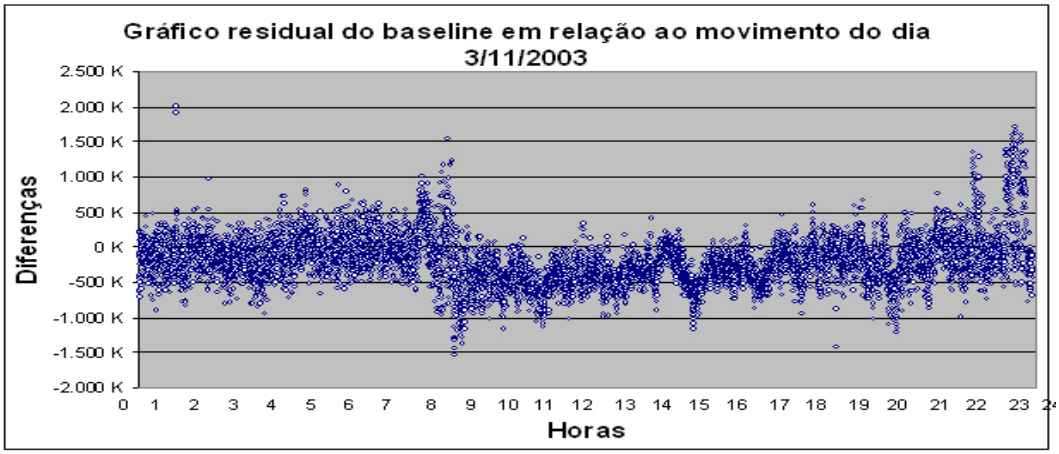


Figura 3.29 - Análise de resíduos para o segmento S_6 em 03/11/2003.

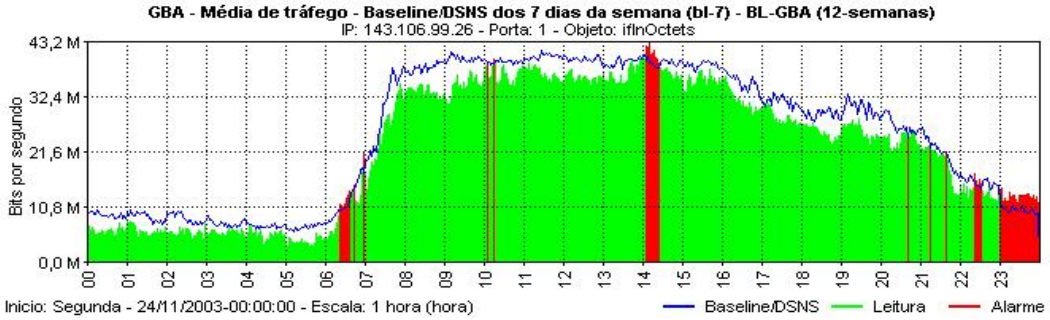


Figura 3.30 - Movimento e *baseline* do dia 24/11/2003 no segmento S_6.

3.5.2 Regressão Linear

O modelo de regressão nos possibilita realizar estimativa ou mesmo previsão de variáveis aleatórias em função de outras variáveis . O objetivo dos modelos de regressão é modelar o relacionamento de duas ou mais variáveis, utilizadas para predição em conjunto com uma variável como resposta. Esta modelagem poderá ser realizada através de uma equação linear ou de uma função não linear. Seu emprego requer o uso de conhecimento teórico, em conjunto com experiência em análise de dados estatísticos. O modelo de regressão linear é composto por um componente sistemático e um aleatório, $Y = f(X) + \varepsilon$, onde f descreve a relação entre Y e X ; Y é a variável resposta ou dependente; X é a variável independente ou preditora e ε são os erros aleatórios que podem ocorrer.

Neste trabalho, a regressão linear foi utilizada com o objetivo de avaliar o grau de ajuste, entre o predito pelo *baseline* gerado pelo BLGBA e o movimento real que ocorreu no segmento analisado. A seguir, são apresentados resultados de cálculos referentes a regressões para os seis segmentos, onde foram realizados os testes durante este trabalho. Serão apresentados gráficos referentes aos meses de julho a dezembro de 2004, contendo resultados referentes à regressão dos dias úteis destes meses, para todos os segmentos e objetos SNMP que foram monitorados neste período. Nos gráficos apresentados a seguir deve-se observar que os ***dias*** apresentados no eixo horizontal se referem somente aos dias úteis do mês, excluindo-se portanto os finais de semana. Esta opção foi adotada devido à baixa utilização da rede neste período, pois a análise destes resultados em conjunto com os dias úteis prejudicava a análise global do segmento estudado.

De modo geral os resultados da regressão, apresentam coeficiente de correlação (R) com valores superiores a 0,8, que indicam um bom ajuste do predito em relação ao movimento real. Resultados que apresentaram coeficiente de correlação (R) com valores inferiores a 0,8 são satisfatórios, porém demandam análises mais detalhadas para avaliar qual o problema que poderia estar ocorrendo.

Os finais de semana foram excluídos destes gráficos por não apresentarem bons resultados nos testes de regressão. Isto ocorreu devido ao baixo volume de tráfego e em conseqüência da não obrigatoriedade do expediente nos finais de semana nos departamentos ou segmentos estudados. Nos finais de semana, observou-se que pequenas alterações no tráfego eram bastante significativas para os testes de regressão, fazendo com que eles apresentassem resultados ruins, ou seja, coeficiente de correlação abaixo de 0,7. Pode-se observar também que este problema ocorreu fundamentalmente de forma bastante aleatória, vários setores ou departamentos trabalhavam nos finais de semana gerando tráfegos não conformes com o *baseline*.

A Figura 3.31, Figura 3.32 e Figura 3.33 demonstram o resultado do cálculo de regressão linear referente ao *baseline*/DSNS, em função do movimento real respectivamente para os objetos *IfInOctets*, *ifOutOctets* e *ipInReceives* que se destinavam ao monitoramento do segmento S_1**.** Os resultados apresentados nestas figuras são referentes aos meses de julho a dezembro de 2004. Como pode ser observado, eles demonstram um bom ajuste entre o *baseline* e o seu respectivo movimento. Alguns problemas surgiram em pontos isolados, retratados através do baixo valor apresentado para o coeficiente de correlação (R). Estes problemas se devem a alterações significativas que ocasionalmente ocorrem no tráfego no dia a dia de funcionamento de uma rede, em função de grandes transferências de arquivos ou mesmo devido a feriados ou recessos. Especificamente no mês de setembro ocorre um valor baixo para o coeficiente (R) no dia 5, que se trata na verdade do feriado do dia 07 de setembro. Em novembro o valor baixo apresentado pela regressão no dia 11 se refere ao feriado do dia 15 de novembro. E em dezembro o baixo valor de (R) entre os dias 18 e 23, referem-se ao recesso de final de ano que ocorre entre as festas do Natal e do Ano Novo. É interessante observar que estes fatores interferem tanto na geração como na avaliação da caracterização do tráfego de forma muito significativa.

As figuras 3.34, 3.35, 3.36 e 3.37 demonstram os resultados do cálculo da regressão linear referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets*, *ifOutOctets*, *ipInReceives*, e *tcpInSegs* utilizados no monitoramento do segmento S_2 que interliga o servidor Web da Universidade Estadual de Londrina. Neste segmento os resultados da regressão também foram satisfatórios, apresentando somente alguns problemas referentes a alterações significativas que ocorreram no tráfego. Estes

problemas ocorreram em função de mudanças no horário do *backup*, realizadas neste servidor, durante os meses de novembro e dezembro de 2004 e de outras operações como, por exemplo, grande quantidade de acessos ocorridos ao servidor Web por ocasião da divulgação de resultados ou mesmo inscrições no concurso vestibular, ou ainda feriados que ocorreram neste período.

As figuras 3.38, 3.39, 3.40 e 3.41, demonstram os resultados do cálculo da regressão linear referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets*, *ifOutOctets*, *ipInReceives*, e *tcpInSegs* utilizados no monitoramento do segmento S_3 que interliga ao *Proxy* da rede da UEL. Os baixos valores apresentados para regressão referentes ao objeto *IfInOctets*, ocorreram em conseqüência da realização do balanceamento de carga realizado neste servidor. Este servidor está interligado à Internet através de dois *links* externos. Neste caso, o próprio sistema operacional em conjunto com o software de Proxy, realiza o balanceamento de carga com intuito de otimizar a utilização dos *links*. Esta operação prejudicou a caracterização neste objeto. Os resultados obtidos do cálculo de regressão para os outros objetos SNMP monitorados foram mais significativos. Somente apresentando problemas que resultaram em coeficiente de correlação (R) baixo, quando ocorreram feriados e recessos ou ainda situações normais de grandes transferências não previstas pelo *baseline*/DSNS.

As figuras 3.42, 3.43 e 3.44 mostram resultados do cálculo da regressão linear referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets*, *ifOutOctets* e *ipInReceives*, utilizados no monitoramento do segmento S_4. Os resultados obtidos referente ao cálculo de regressão para este segmento, que é responsável por interligar a rede da UEL ao seu roteador principal foram satisfatórios, exceto alguns resultados coeficientes de correlação baixos em dias de feriados ou recessos ocorridos durante os meses analisados.

As figuras 3.45 e 3.46 mostram os resultados do cálculo da regressão linear referente aos meses de julho a dezembro de 2004, para os objetos *IfInOctets* e *ifOutOctets*, utilizados no monitoramento do segmento S_5. Os resultados alcançados neste segmento, que é composto por poucas estações e consecutivamente pouco tráfego agregado, foi notadamente inferior ao dos outros segmentos. Os baixos valores para o coeficiente de correlação apresentados foram mais uma comprovação do que se havia constatado, através

de avaliações visuais entre o *baseline*/DSNS gerado e o movimento real. Na verdade conclui-se que o modelo BLGBA apresenta um desempenho inferior em segmentos com baixo volume de tráfego agregado, devendo portanto ser realizado mais testes e análises com objetivo de ajustar o modelo a este tipo de segmentos.

As figuras 3.47 e 3.48 demonstram os resultados do cálculo da regressão linear referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets* e *ifOutOctets*, utilizados no monitoramento do segmento S_6. Os resultados obtidos referente ao cálculo de regressão para este segmento que interliga a rede da UNICAMP à Internet foram excelentes, apresentando problemas somente em feriados ou recessos ocorridos.

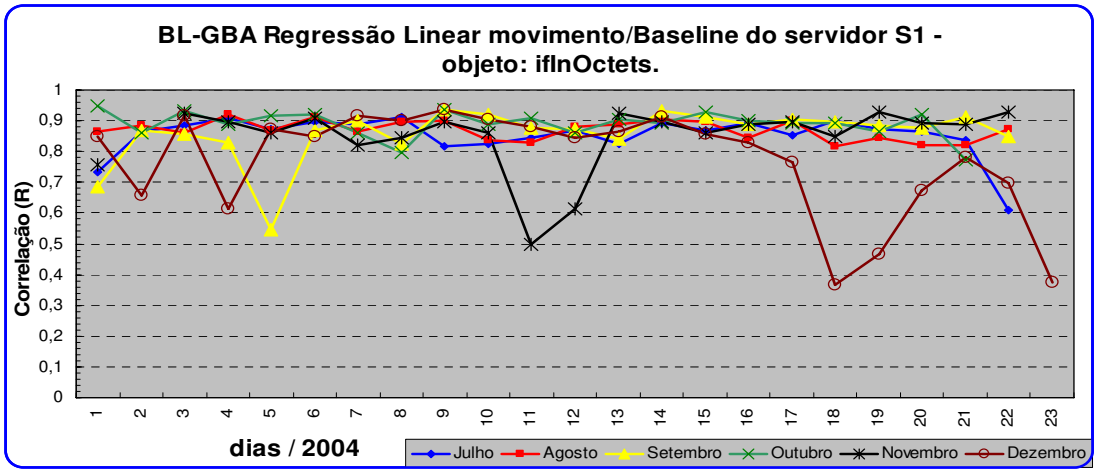


Figura 3.31 - Regressão Linear para objeto *IfInOctets* do servidor S_1 de julho a dezembro de 2004.

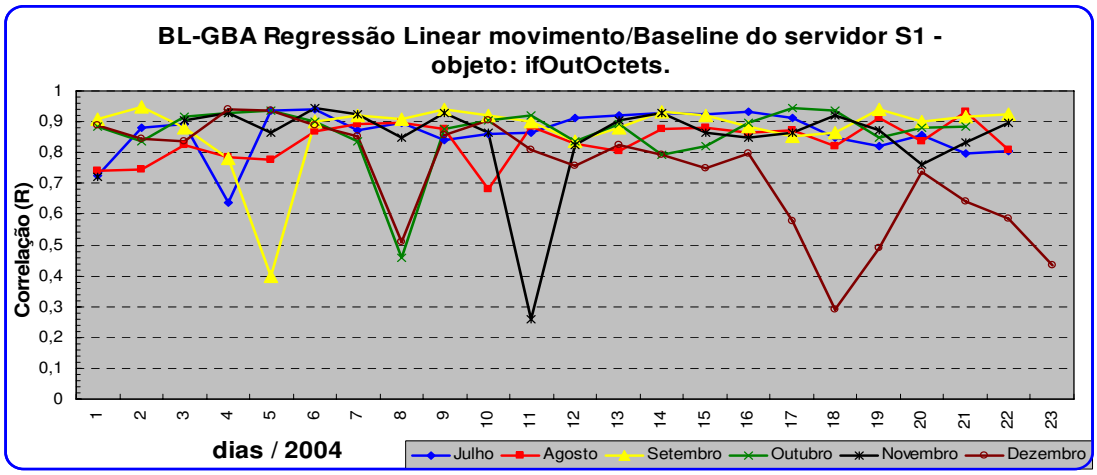


Figura 3.32 - Regressão Linear para objeto *ifOutOctets* do servidor S_1 de julho a dezembro de 2004.

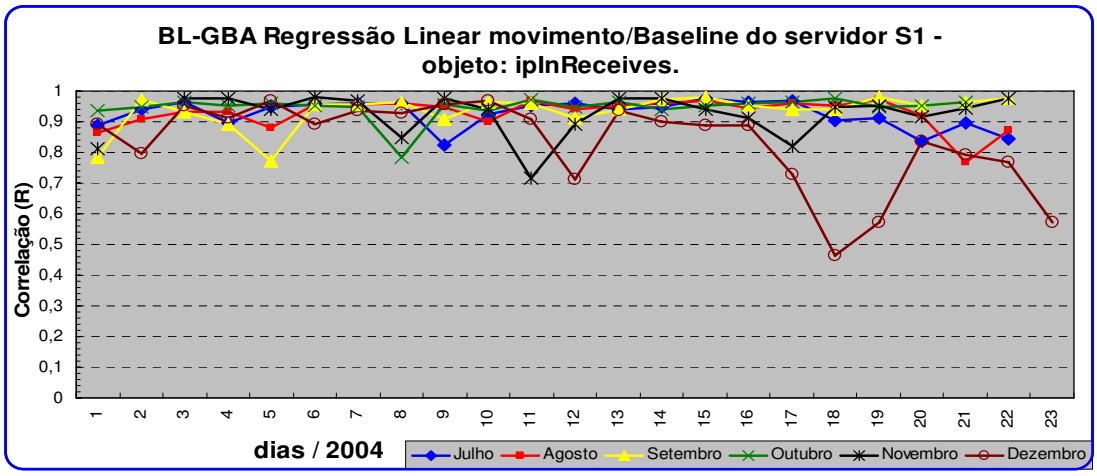


Figura 3.33 - Regressão Linear para objeto *ipInReceives* do servidor S_1 de julho a dezembro de 2004.

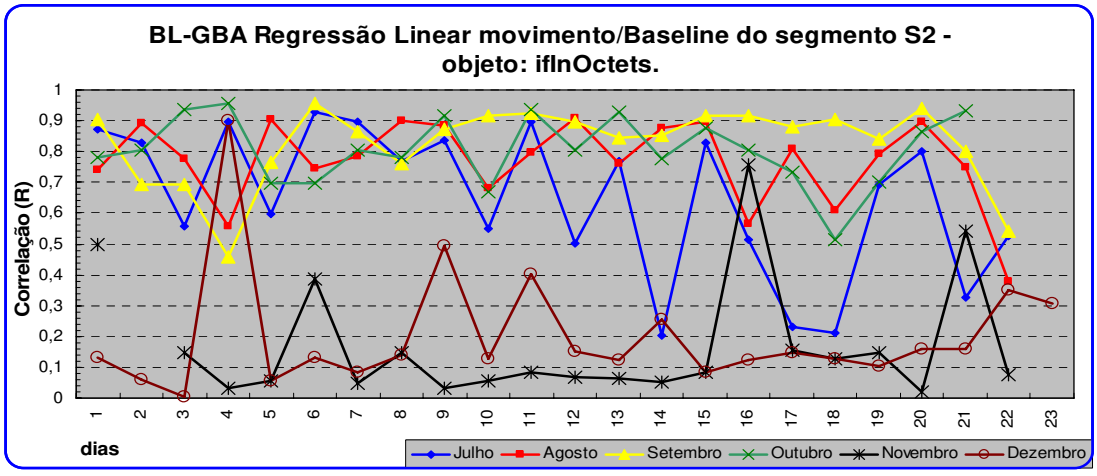


Figura 3.34 - Regressão Linear para objeto *IfInOctets* do servidor S_2 de julho a dezembro de 2004.

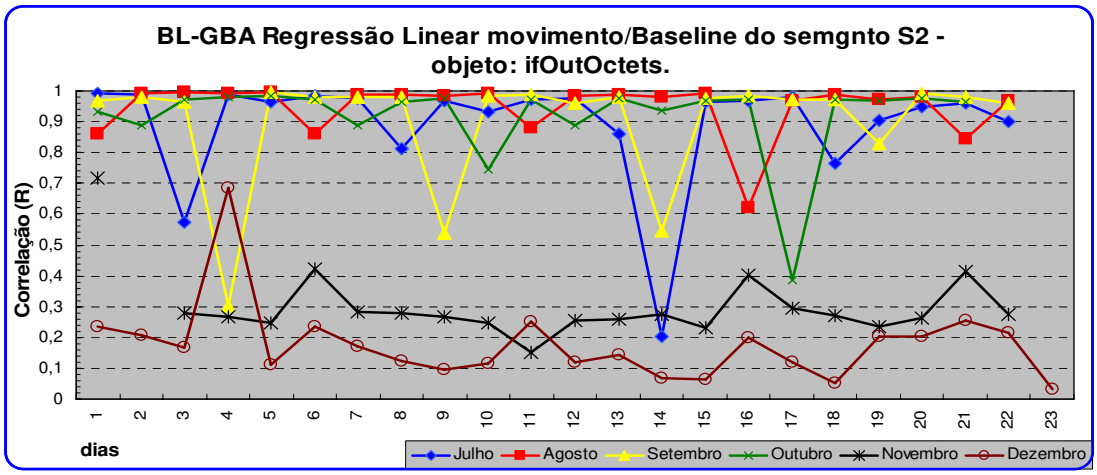


Figura 3.35 - Regressão Linear para objeto *ifOutOctets* do servidor S_2 de julho a dezembro de 2004.

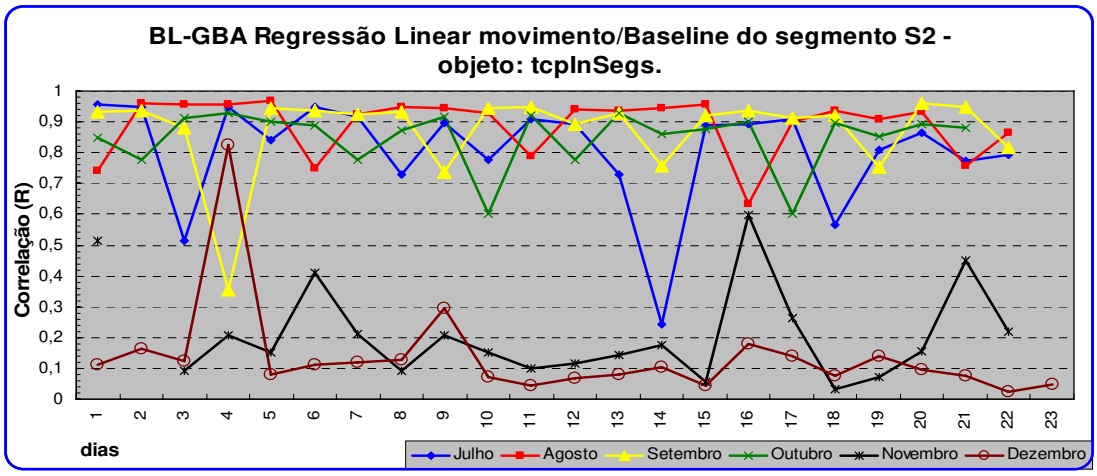


Figura 3.36 - Regressão Linear para objeto *tcpInSegs* do servidor S_2 de julho a dezembro de 2004.

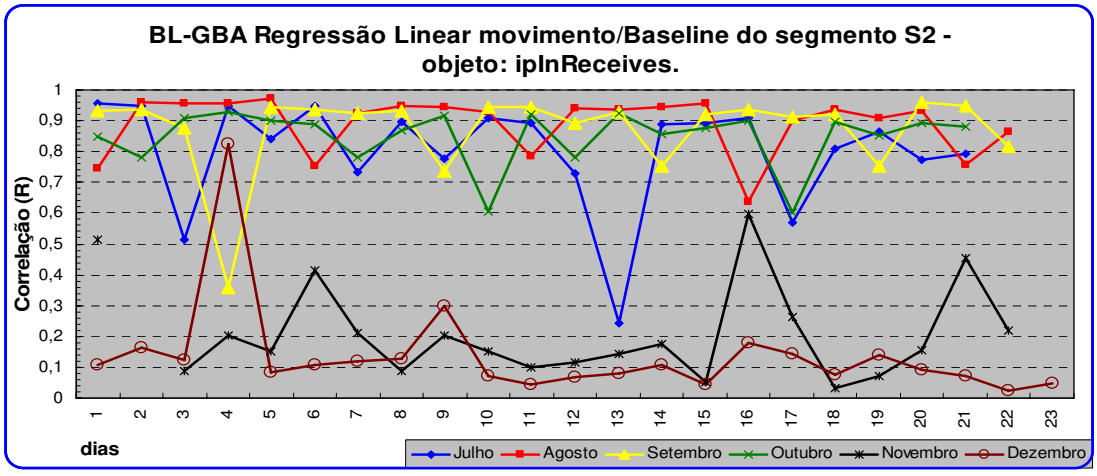


Figura 3.37 - Regressão Linear para objeto *ipInReceives* do servidor S_2 de julho a dezembro de 2004.

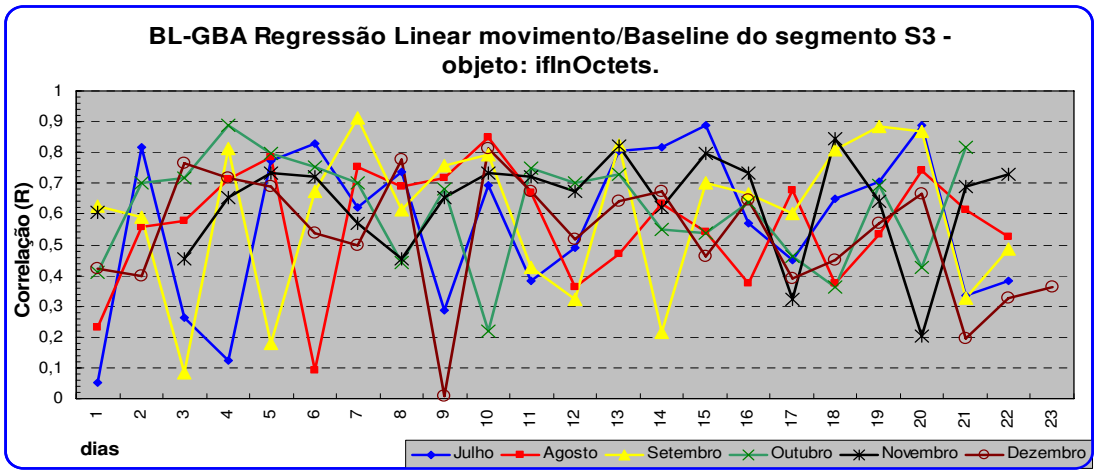


Figura 3.38 - Regressão Linear para objeto *IfInOctets* do servidor S_3 de julho a dezembro de 2004.

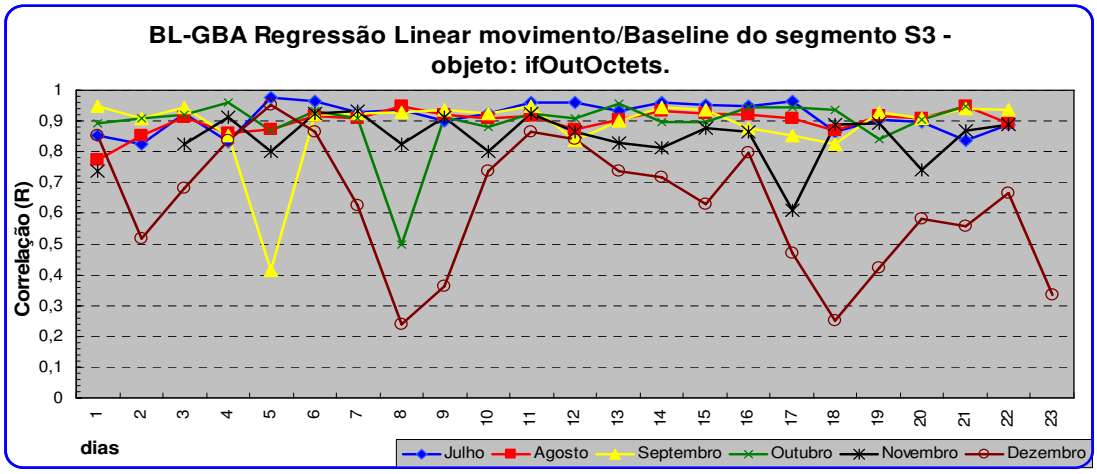


Figura 3.39 - Regressão Linear para objeto *ifOutOctets* do servidor S_3 de julho a dezembro de 2004.

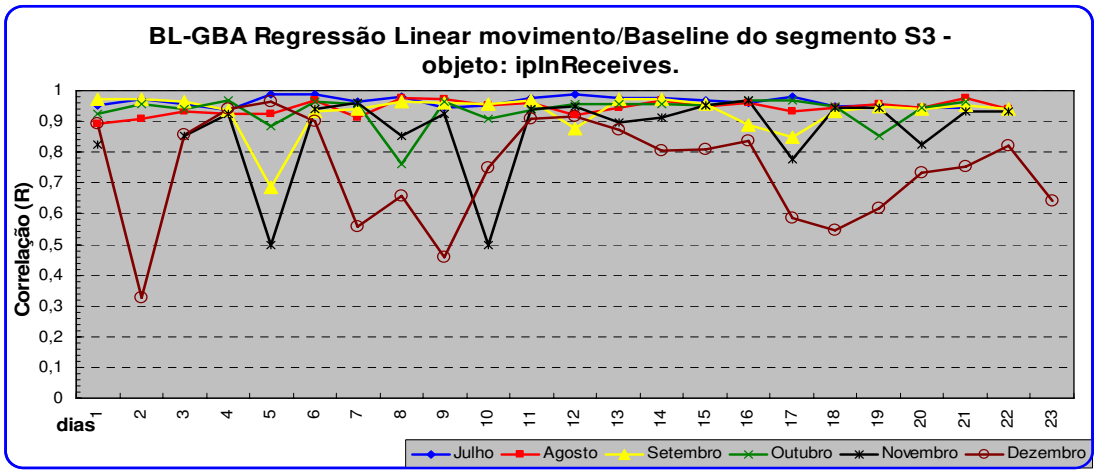


Figura 3.40 - Regressão Linear para objeto *ipInReceives* do servidor S_3 de julho a dezembro de 2004.

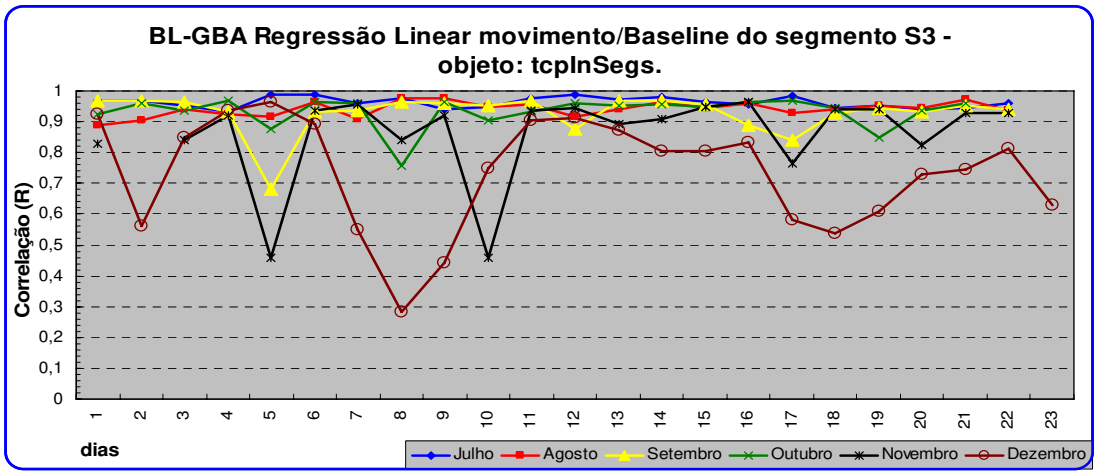


Figura 3.41 - Regressão Linear para objeto *tcpInSegs* do servidor S_3 de julho a dezembro de 2004.

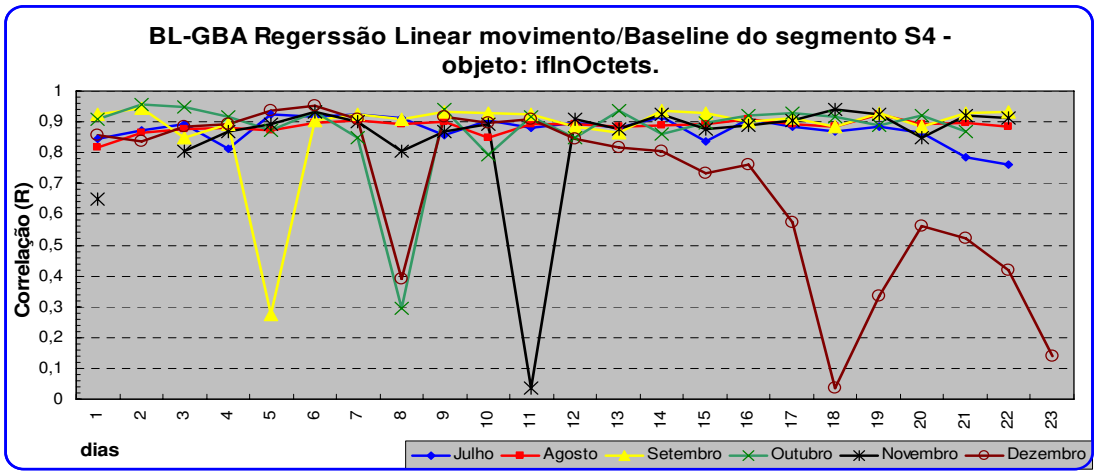


Figura 3.42 - Regressão Linear para objeto *IfInOctets* do servidor S_4 de julho a dezembro de 2004.

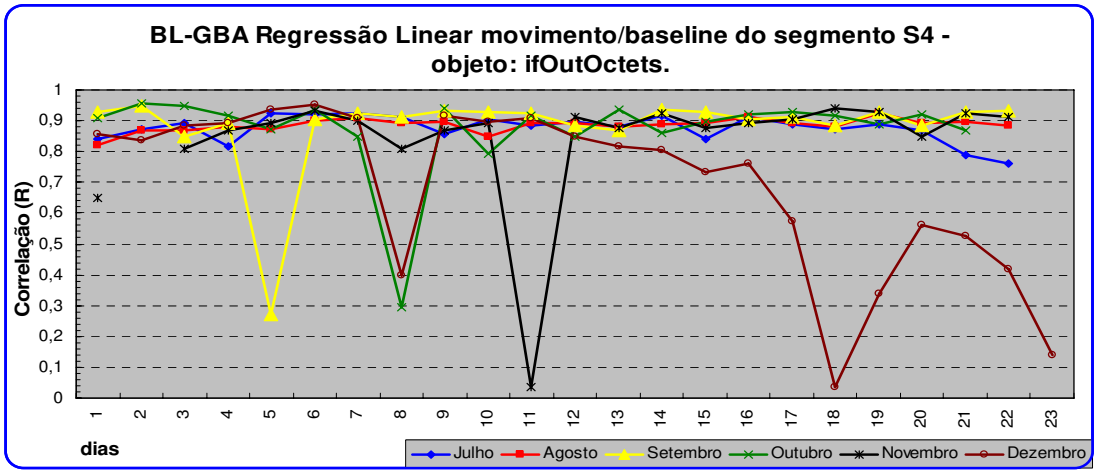


Figura 3.43- Regressão Linear para objeto *ifOutOctets* do servidor S_4 de julho a dezembro de 2004.

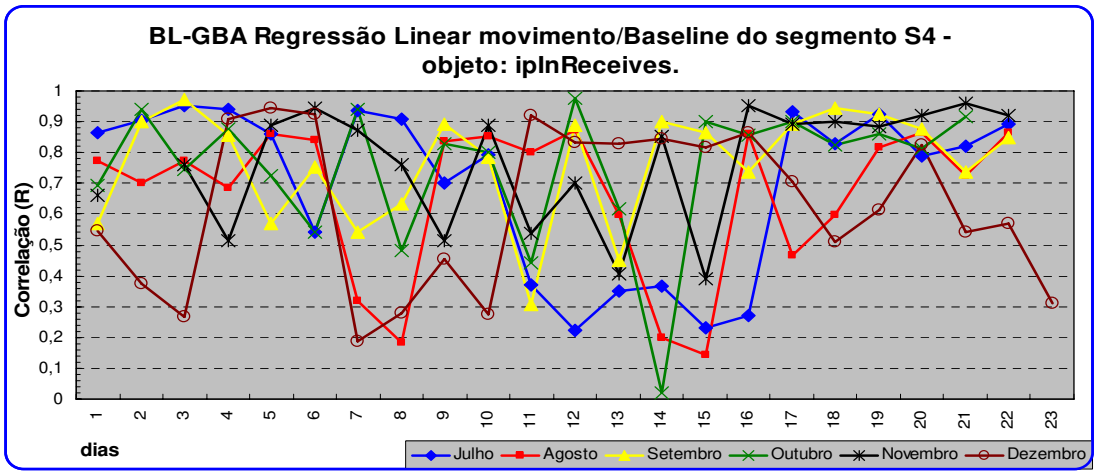


Figura 3.44 - Regressão Linear para objeto *ipInReceives* do servidor S_4 de julho a dezembro de 2004.

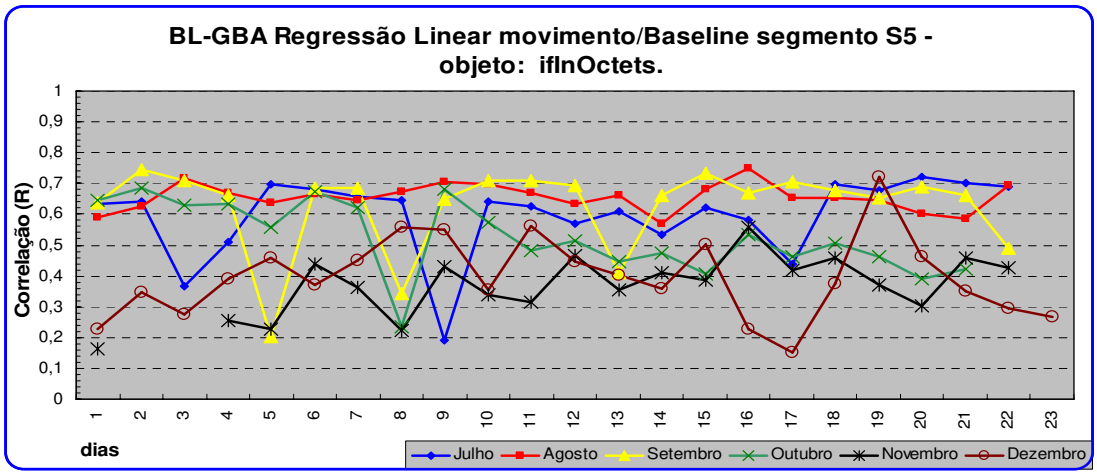


Figura 3.45 - Regressão Linear para objeto *IfInOctets* do servidor S_5 de julho a dezembro de 2004.

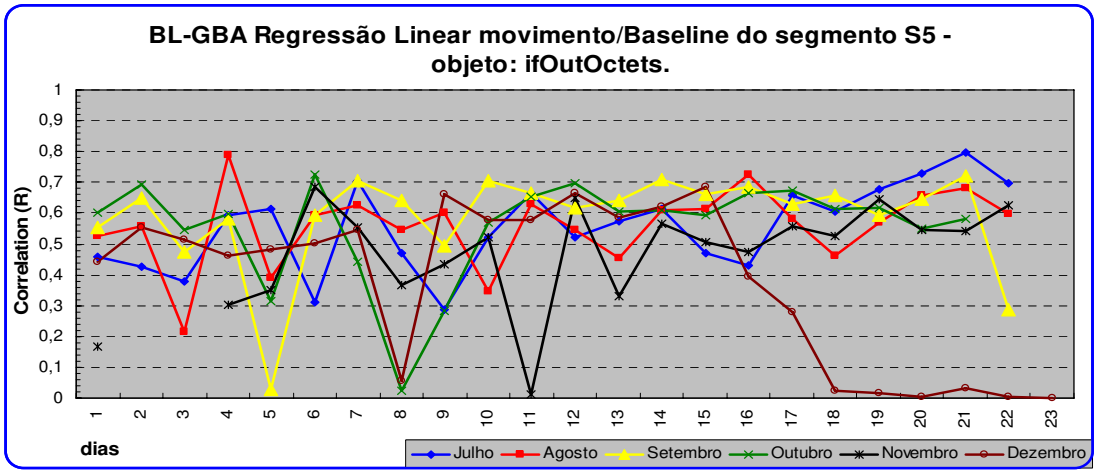


Figura 3.46 - Regressão Linear para objeto *ifOutOctets* do servidor S_5 de julho a dezembro de 2004.

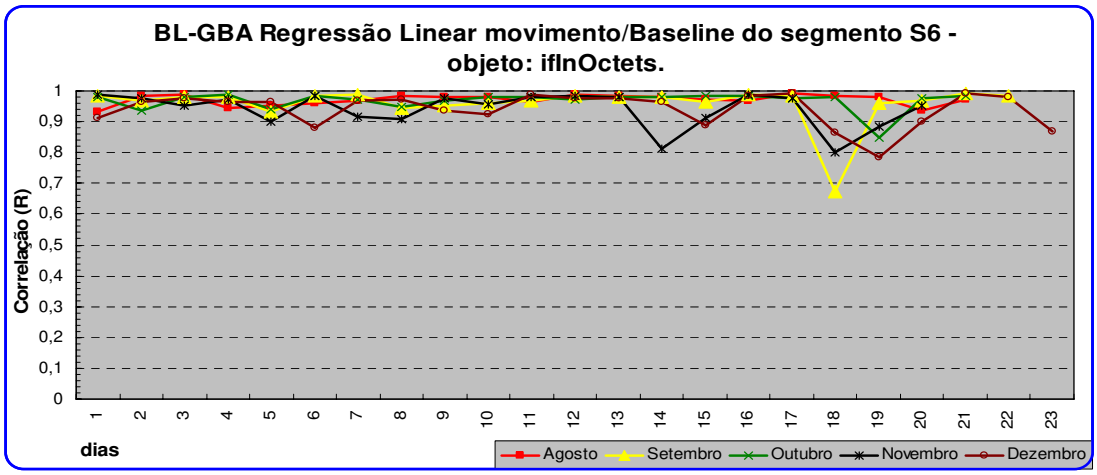


Figura 3.47 - Regressão Linear para objeto *IfInOctets* do servidor S_6 de agosto a dezembro de 2003.

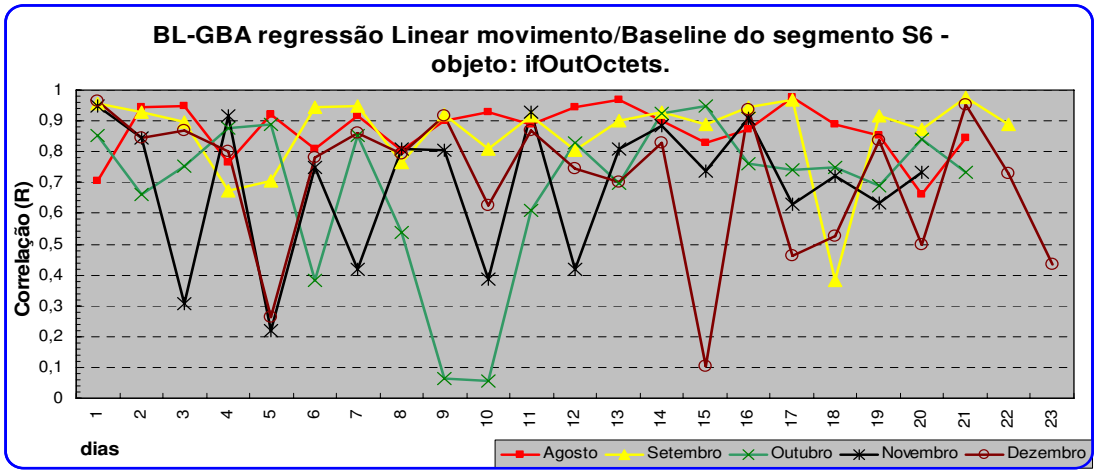


Figura 3.48 - Regressão Linear para objeto *ifOutOctets* do servidor S_6 de agosto a dezembro de 2003.

3.5.3 Teste de Bland e Altman

O teste de Bland e Altman (Bland and Altman, 1986), apesar de simples não deixa dúvida em ser um teste estatístico de fácil aplicabilidade e entendimento. Eles apresentam uma abordagem baseada em técnicas gráficas e cálculos simples. Foi desenvolvido como alternativa para avaliações estatísticas realizadas por testes tradicionais que utilizam coeficiente de correlação ou técnicas de análise de regressão.

Para Bland e Altman o cálculo do coeficiente de correlação (r) entre dois conjuntos de dados apresenta alguns problemas, e seu uso acaba sendo inapropriado. Eles acreditam que uma alta correlação não significa necessariamente que duas medições ou métodos possam concordar entre si, devido aos seguintes fatores:

1. O coeficiente de correlação (r), usado em regressões, mede a força da relação entre duas variáveis e não o ajuste entre elas;
2. Uma mudança na escala das medições realizadas não afeta a correlação, mais certamente irá afetar o ajuste entre duas medidas;
3. O teste de significado pode demonstrar se dois métodos estão relacionados. Seria surpreendente, se estes métodos utilizados para avaliar a mesma coisa, não estivessem relacionados. Para eles o teste de significado é irrelevante em questões de concordância.
4. Dados que parecem apresentar uma concordância ruim podem produzir altos índices de correlação.

Como alternativa aos métodos que utilizam o coeficiente de correlação (r), o teste proposto por Bland e Altman se refere à análise dos desvios que possam ocorrer entre uma medida e outra. Eles recomendam que as análises devem ser baseadas na diferença entre as medidas realizadas no mesmo objeto. A média da diferença é utilizada como desvio que sinaliza a diferença sistemática entre o previsto e o realizado. O desvio padrão das diferenças mede as flutuações aleatórias entre elas. Para eles, 95 % dos desvios ou erros

observados devem estar entre os limites de $\overline{d} \pm 2 * s$, onde $\overline{d}$ é a média e s é o desvio padrão das diferenças entre uma medida e outra. Em nosso caso seriam as diferenças entre o *baseline*/DSNS e o movimento real. Se os erros estiverem dentro dos limites propostos significa que o modelo apresenta um bom ajuste e alto grau de confiabilidade.

A Figura 3.49, demonstra o resultado do teste de Bland & Altman para o segmento S_1 no dia 15/01/2004. No gráfico podem ser vistas as diferenças entre o *baseline*/DSNS e o movimento real que ocorreu neste dia. As linhas pontilhadas se referem à média ao centro, média mais duas vezes o desvio padrão e média menos duas vezes o desvio padrão. Os pontos que se situam fora dos limites estabelecidos por $\overline{d} \pm 2 * s$ são considerados erros, neste exemplo ilustrado pela figura 3.49, observa-se 12 erros que se apresentam sina sinalizados pelos pontos situados a partir das linhas pontilhas indicadas por média + 2 *s* e média – 2 *s*.

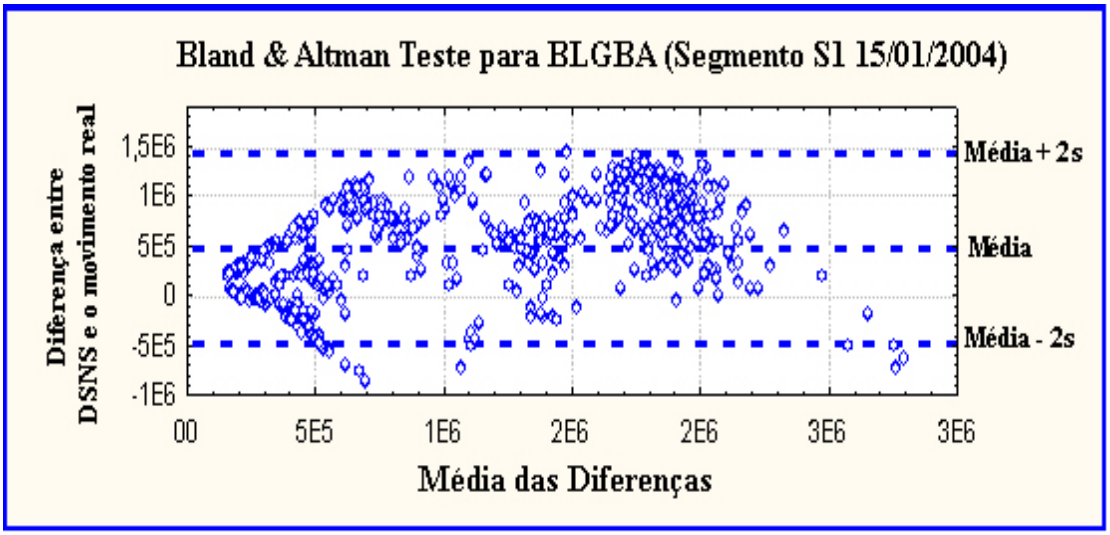


Figura 3.49 - Exemplo de gráfico utilizado para avaliação no teste de Bland e Altman.

A seguir serão apresentados os resultados do teste de Bland e Altman para os segmentos estudados, referentes a seis meses consecutivos, de julho a dezembro de 2004, exceto para o segmento S_6. No caso do segmento S_6, somente são apresentados os

resultados de agosto a dezembro de 2004, período que foi possível realizar teste neste segmento.

De modo geral, como é demonstrado nos gráficos resultantes do teste de Bland e Altman, os resultados são inteiramente satisfatórios, mostrando o bom ajuste entre o *baseline* gerado e o movimento real ocorrido no segmento analisado. Alguns dias ocorreram problemas e os resultados apresentados não são de acordo com o limite de erros aceitáveis. Após análise de todos estes dias constatou-se que se tratava de alterações no tráfego, não prevista no *baseline* e que de fato deveriam ser sinalizadas, para que o administrador da rede pudesse tomar ciência das mesmas.

Figura **3.51**, Figura **3.52** e Figura **3.53**: ilustram os resultados do teste de Bland e Altman para os meses de julho a dezembro de 2004, do segmento S_1, utilizando os objetos *ifInOctets*, *ifOutOctets* e *ipInReceives*. Este segmento se refere ao servidor de firewall da rede da UEL, que é responsável por interligar sua rede local à Internet. Os resultados neste segmento foram ótimos, conforme podem ser observados nas figuras, demonstrando, segundo os critérios do teste, que o modelo BLGBA realmente produziu uma caracterização adequada a este tipo de segmento.

Figuras 3.54, 3.55, 3.56 e 3.57: demonstram os resultados do teste de Bland e Altman para os meses de julho a dezembro de 2004, do segmento S_2, utilizando os objetos *ifInOctets*, *ifOutOctets*, *ipInReceives* e *tcpInSegs*. Este segmento se refere ao servidor de Web da rede da UEL. Os resultados obtidos neste segmento foram ótimos conforme é mostrado nas figuras, comprovando assim, segundo os critérios do teste proposto por Bland e Altman, o bom ajuste do *baseline* gerado em relação ao movimento que realmente ocorreu. Alguns problemas são observados nos meses de novembro e dezembro. Eles ocorreram em função de uma alteração no horário do backup que era realizado neste servidor. A Figura 3.50 demonstra exemplos reais do movimento e respectivo *baseline*/DSNS referentes aos dias 29/11/2004 a 03/12/2004. Pode ser constatada claramente nesta figura a alteração no horário do *backup* que era realizado no período de 02:00 e 04:00 horas. Estas alterações afetaram tanto os resultados deste teste como os de regressão apresentados anteriormente.

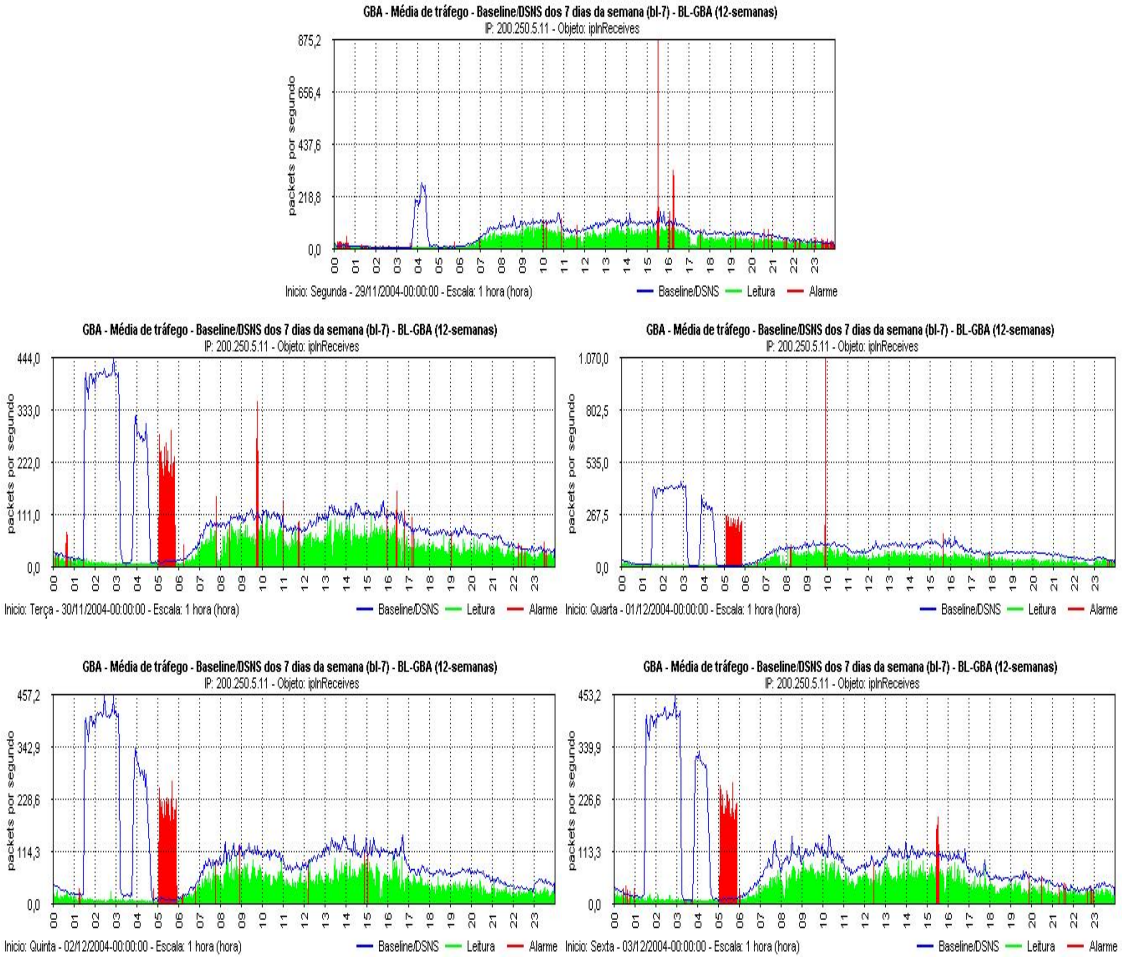


Figura 3.50 - Exemplo de alteração no horário de backup do segmento S_2, ocorrida em novembro e dezembro de 2004.

Figuras 3.58, 3.59, 3.60 e 3.61: demonstram os resultados do teste de Bland e Altman referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets*, *ifOutOctets*, *ipInReceives*, e *tcpInSegs* utilizados no monitoramento do segmento $\boldsymbol{S_3}$ que interliga ao *Proxy* da rede da UEL. Todas as estações da rede da UEL utilizam este servidor como *proxy* para acesso a Web. Em algumas situações foram constatados alguns valores altos referentes à quantidade de erros, pertencentes ao objeto *IfInOctets*. Estas situações, também constatadas no cálculo de regressão linear para o mesmo período e objeto, deve-se ao fato do servidor de *proxy* realizar o balanceamento de carga entre os *links* de acesso externo à Internet. Os demais resultados do teste referentes a segmento foram satisfatórios demonstrando o bom ajuste do modelo em relação ao movimento real.

Figuras 3.62, 3.63 e 3.64: demonstram os resultados do teste de Bland e Altman referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets*, *ifOutOctets* e *ipInReceives*, utilizados no monitoramento do segmento S_4. Os resultados obtidos referente ao teste para este segmento, que é responsável por interligar a rede da UEL a seu roteador principal, foram muito bons, exceto em algumas situações onde a quantidade de erros apresentados foram superiores ao limite estabelecido pelo teste, que ocorreram devido a períodos de feriados ou recessos ocorridos durante os meses analisados. A ocorrência de um número de erros maior do que o normal nos dias de feriados e em recessos, se deve ao fato da diminuição do tráfego normal não estar prevista no *baseline*.

Figuras 3.65 e 3.66: ilustram os resultados do teste de Bland e Altman referente aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets* e *ifOutOctets*, utilizados no monitoramento do segmento S_5. Os resultados obtidos neste segmento, que é composto por poucas estações e pouco tráfego agregado, foram notadamente inferiores aos outros segmentos. O grande volume de erros encontrados reafirmou o que já havíamos constatado através de avaliações visuais entre o *baseline*/DSNS e o movimento real, ou seja, o modelo BLGBA apresenta um desempenho inferior, em segmentos com baixo volume de tráfego agregado.

Figuras 3.67 e 3.68: demonstram os resultados do teste de Bland e Altman referentes aos meses de julho a dezembro de 2004, respectivamente, para os objetos *IfInOctets* e *ifOutOctets*, utilizados no monitoramento do segmento S_6. Os resultados obtidos referentes ao teste realizado para este segmento, que interliga a rede da UNICAMP à Internet foram excelentes, apresentando problemas somente em feriados ou recessos ocorridos.

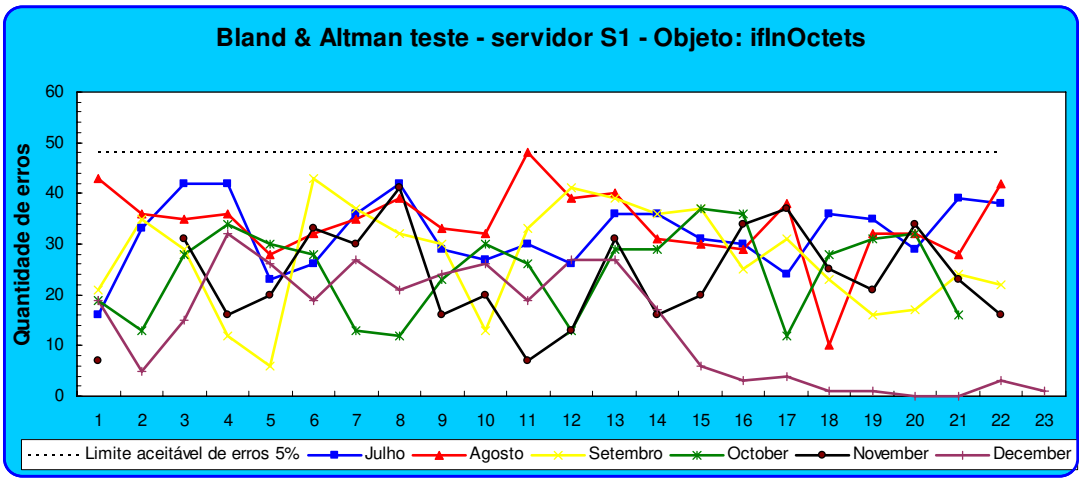


Figura 3.51 - Bland & Altman teste para o segmento S_1 de julho a dezembro de 2004, objeto *ifInOctets*.

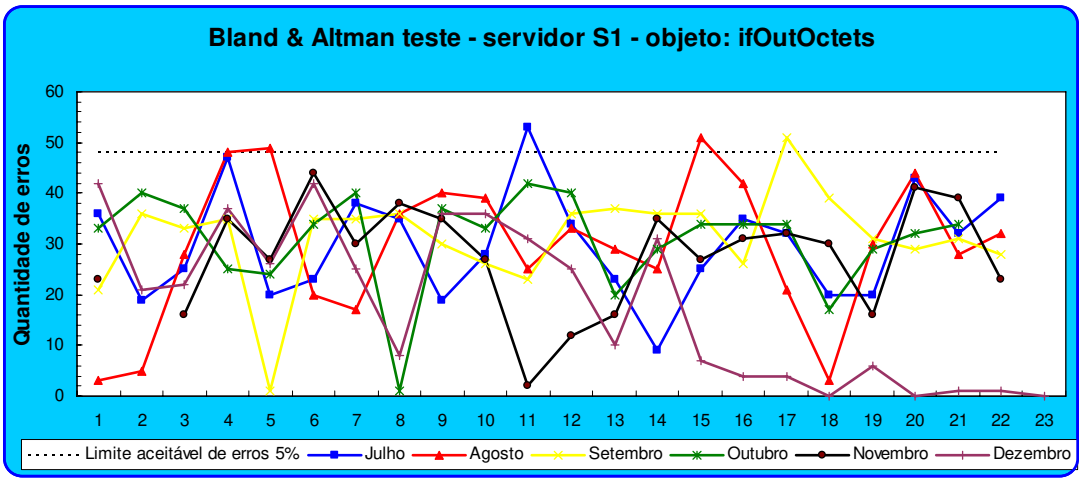


Figura 3.52 - Bland & Altman teste para S_1 de julho a dezembro de 2004, objeto *ifOutOctets*.

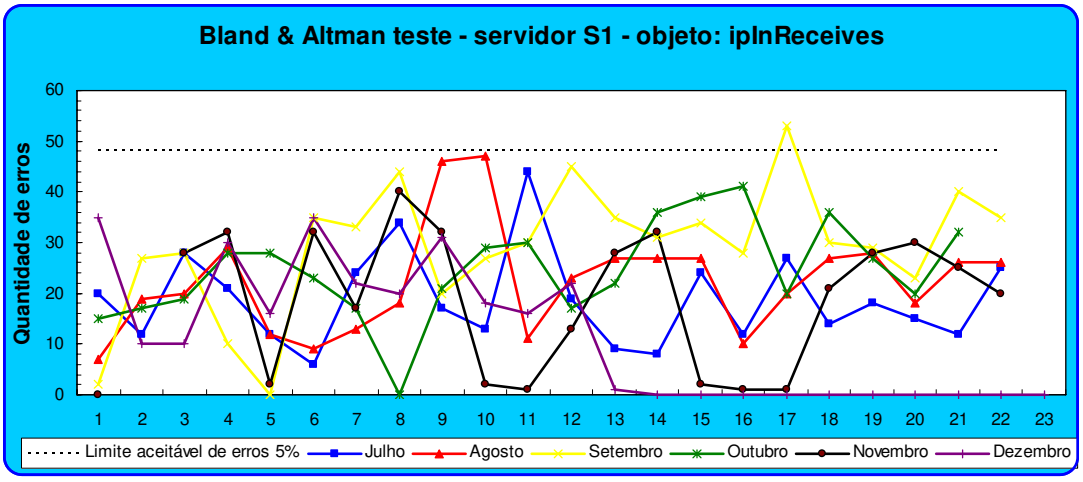


Figura 3.53 - Bland & Altman teste para S_1 de julho a dezembro de 2004, objeto *ipInReceives*.

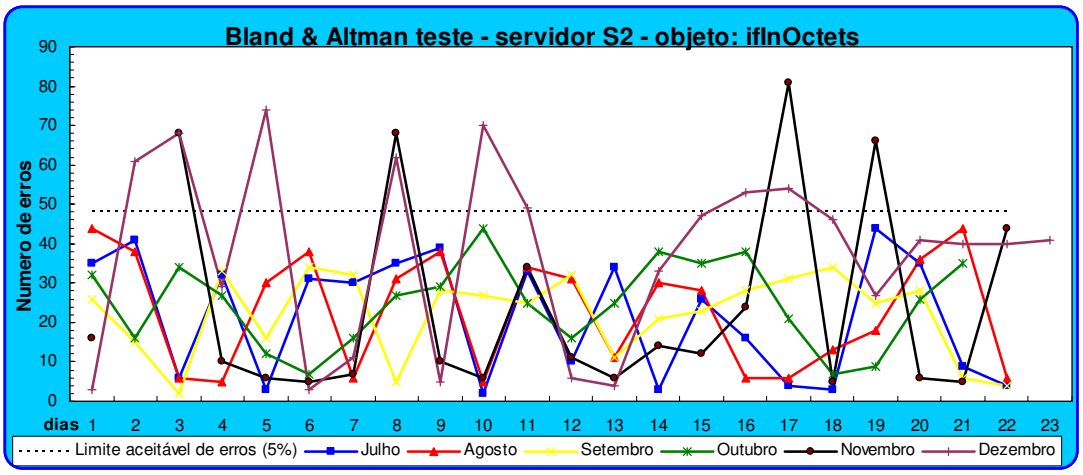


Figura 3.54- Bland & Altman teste para o segmento S_2 de julho a dezembro de 2004, objeto *ifInOctets*.

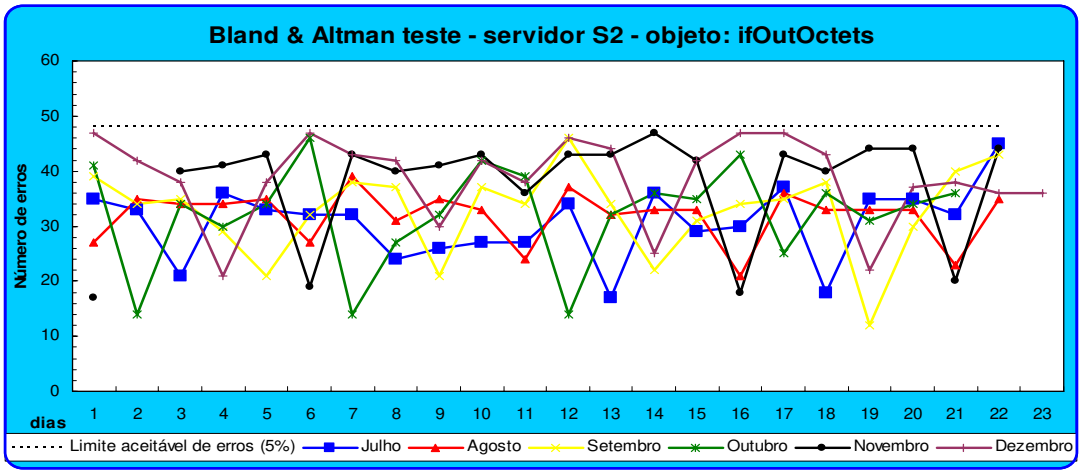


Figura 3.55 - Bland & Altman teste para o segmento S_2 de julho a dezembro de 2004, objeto *ifOutOctets*.

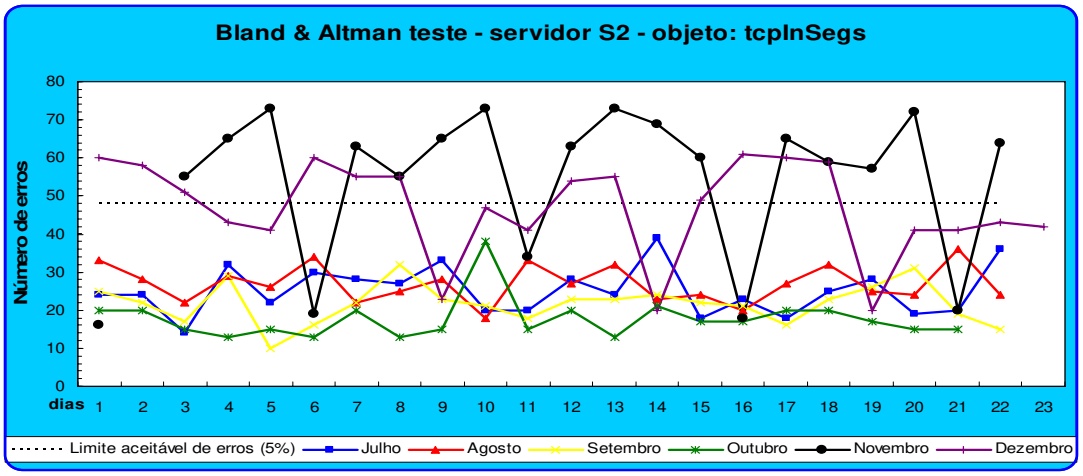


Figura 3.56 - Bland & Altman teste para o segmento S_2 de julho a dezembro de 2004, objeto *tcpInSegs*.

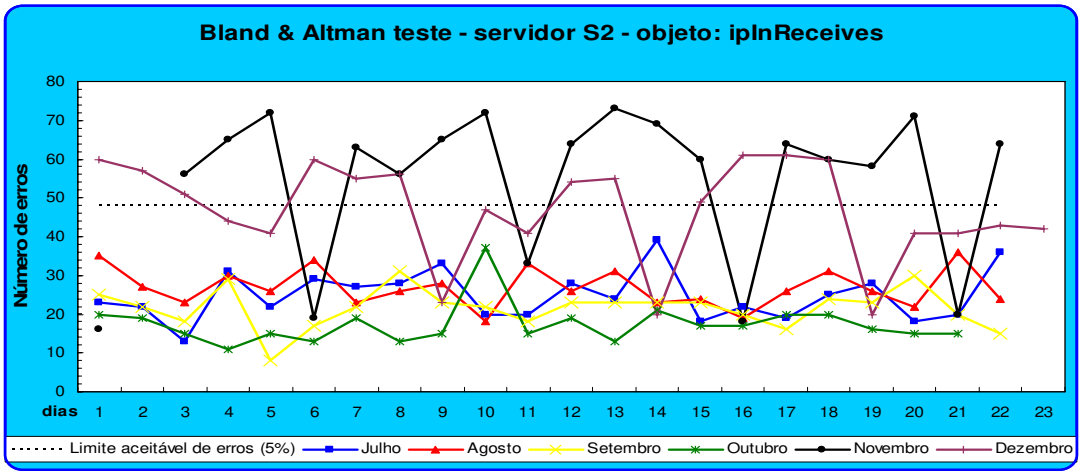


Figura 3.57 - Bland & Altman teste para S_2 de julho a dezembro de 2004, objeto *ipInReceives*.

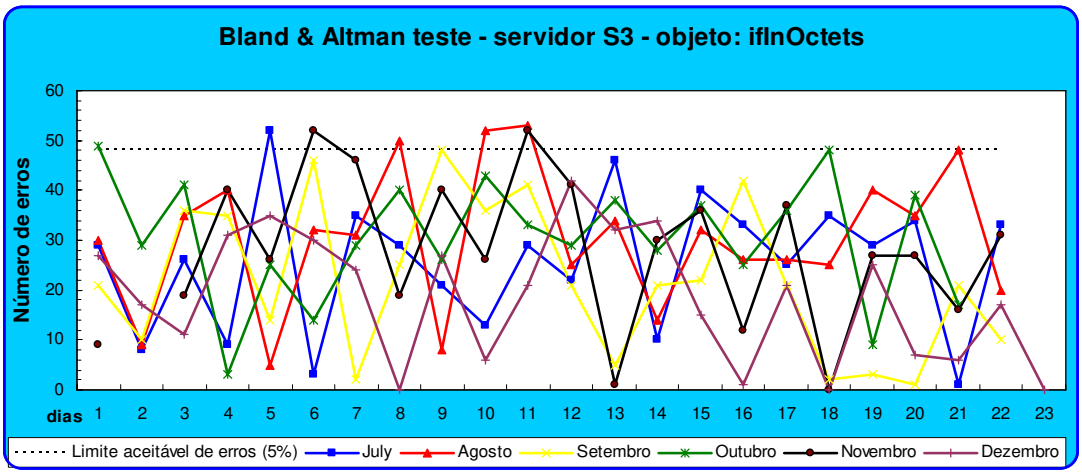


Figura 3.58 - Bland & Altman teste para o segmento S_3 de julho a dezembro de 2004, objeto *ifInOctets*.

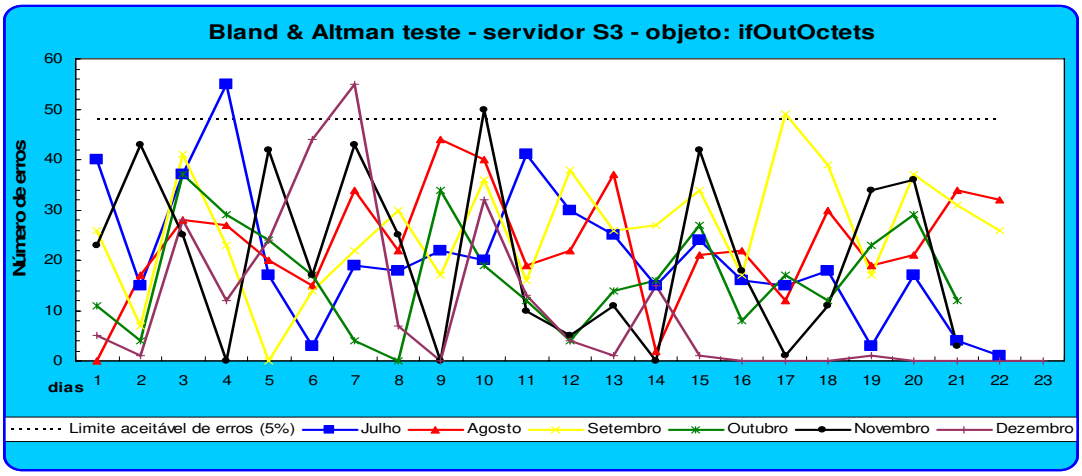


Figura 3.59 - Bland & Altman teste para o segmento S_3 de julho a dezembro de 2004, objeto *ifOutOctets*.

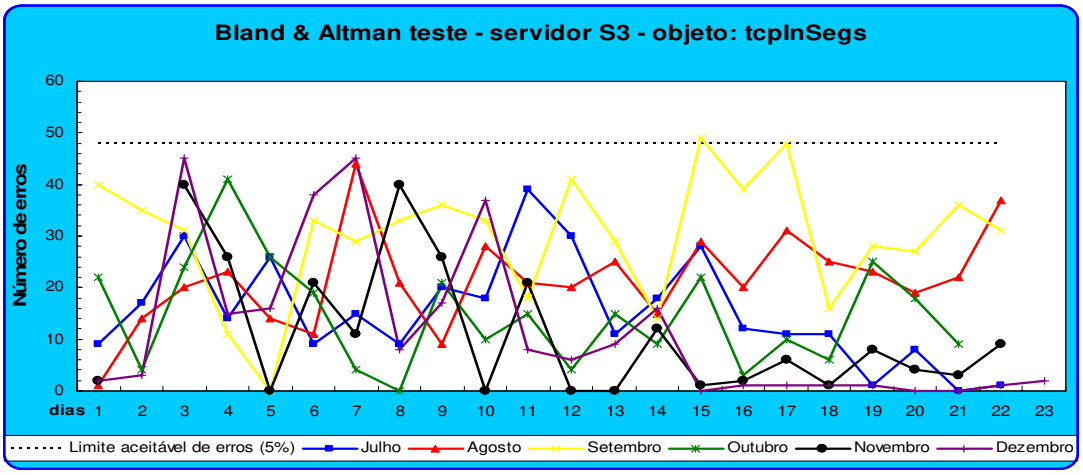


Figura 3.60 - Bland & Altman teste para o segmento S_3 de julho a dezembro de 2004, objeto *tcpInSegs*.

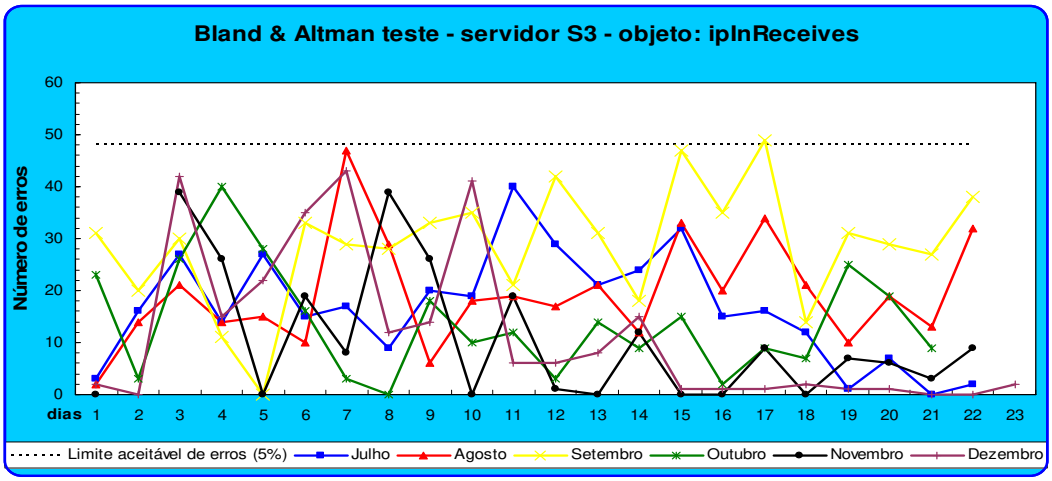


Figura 3.61 - Bland & Altman teste para S_3 de julho a dezembro de 2004, objeto *ipInReceives*.

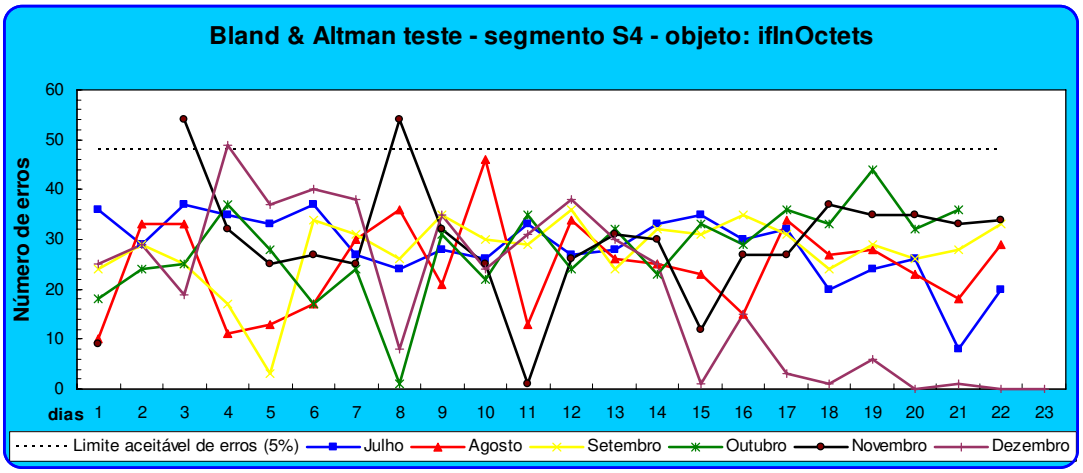


Figura 3.62 - Bland & Altman teste para o segmento S_4 de julho a dezembro de 2004, objeto *ifInOctets*.

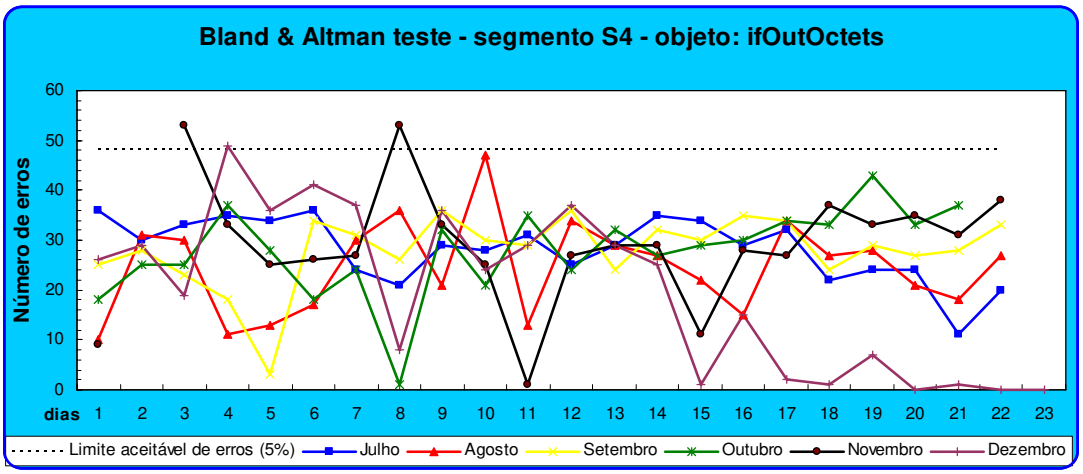


Figura 3.63 - Bland & Altman teste para o segmento S_4 de julho a dezembro de 2004, objeto *ifOutOctets*.

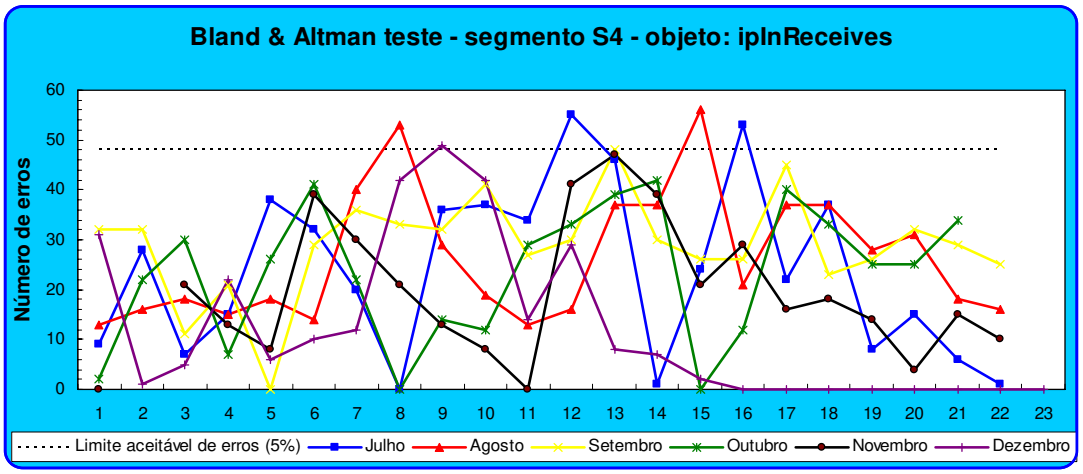


Figura 3.64- Bland & Altman teste para S_4 de julho a dezembro de 2004, objeto *ipInReceives*.

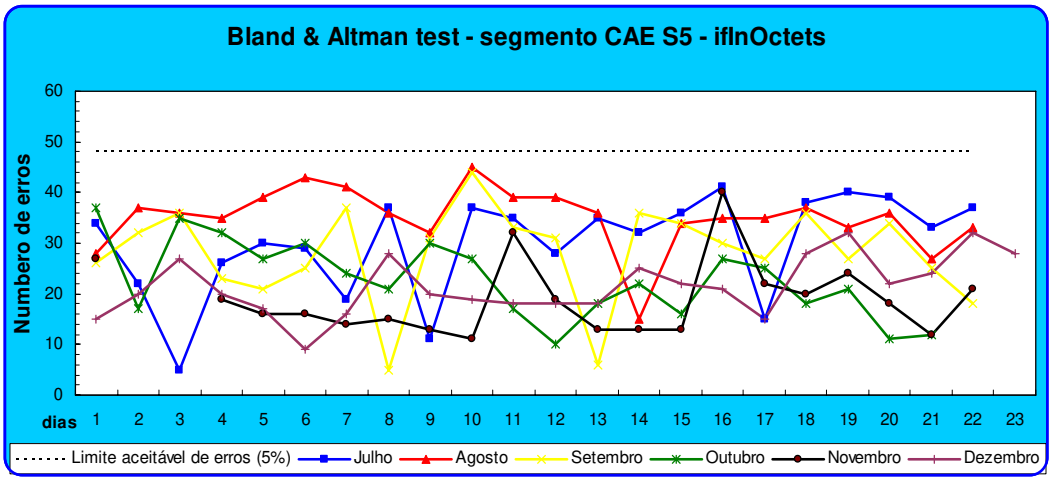


Figura 3.65 - Bland & Altman teste para o segmento S_5 de julho a dezembro de 2004, objeto *ifInOctets*.

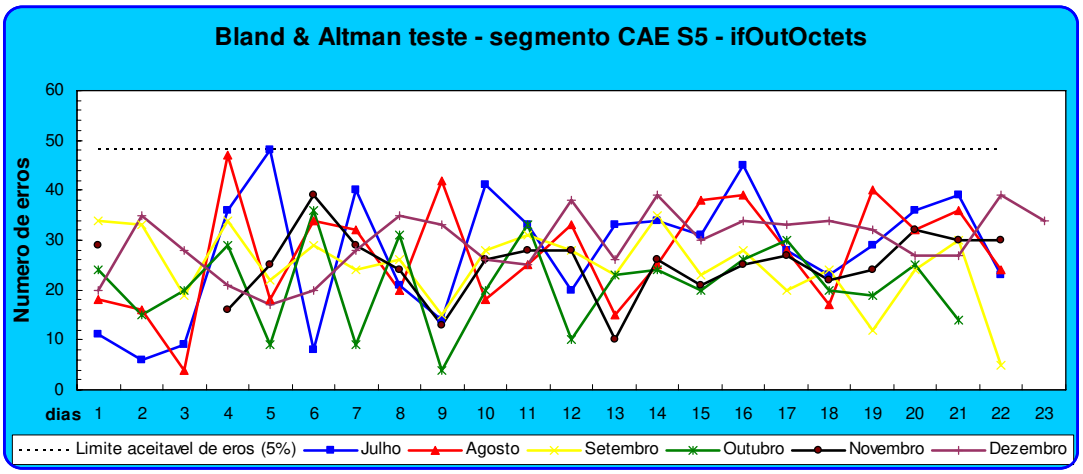


Figura 3.66 - Bland & Altman teste para o segmento S_5 de julho a dezembro de 2004, objeto *ifOutOctets*.

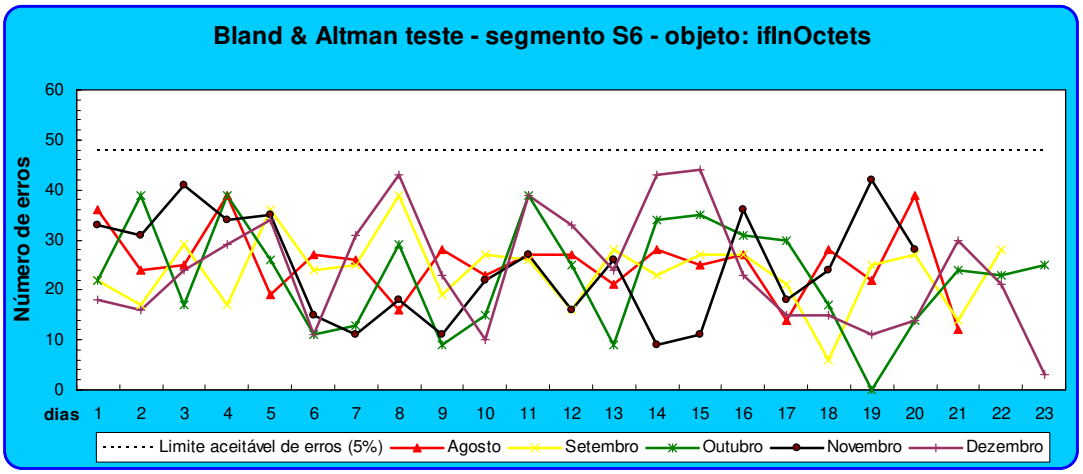


Figura 3.67 - Bland & Altman teste para o segmento S_6 de julho a dezembro de 2004, objeto *ifInOctets*.

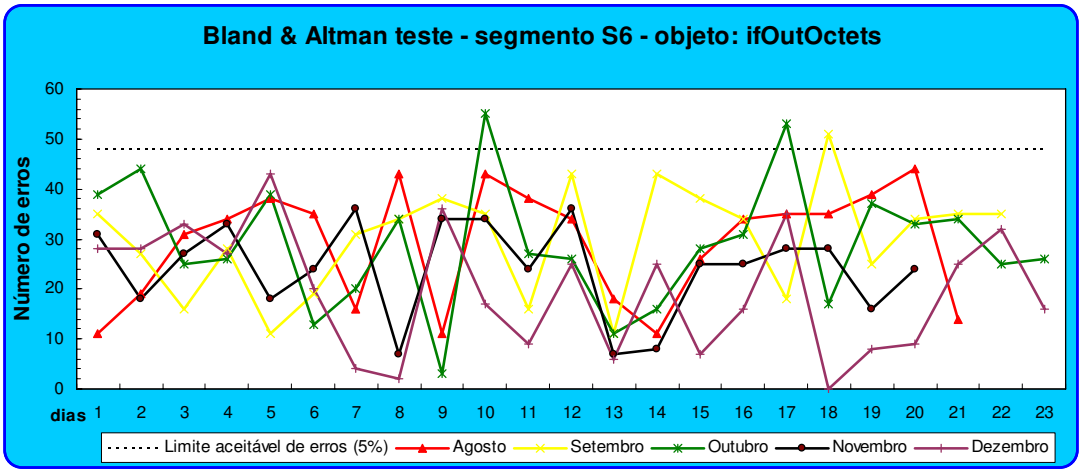


Figura 3.68 - Bland & Altman teste para o segmento S_6 de julho a dezembro de 2004, objeto *ifOutOctets*.

3.6 Resultados do Modelo BLGBA

Os testes visando a validação do modelo foram realizados em seis diferentes segmentos de rede, conforme descrito anteriormente neste trabalho. Procurou-se avaliar na prática o desempenho do modelo utilizando segmentos com características diferentes e com diferentes objetos SNMP. Após a consolidação do modelo BLGBA, que ocorreu em meados de 2003, os testes foram intensificados e têm sido realizados diuturnamente até o presente momento.

Durante o período de testes pôde-se observar a ocorrência de alguns fatores esporádicos, porém determinantes, para construção do *baseline,* que implicam diretamente no tráfego da rede e na qualidade do *baseline* gerado. Períodos como férias, greves, resultados de vestibular, recesso de fim de ano e mesmo feriados, causam alterações significativas no volume do tráfego, implicando diretamente na sua caracterização; por isso, optou-se pela desconsideração dos mesmos na realização de alguns testes. O problema ocorre porque o período de aprendizagem para formação do *baseline* leva no mínimo quatro e no máximo doze semanas e a duração destes fatores normalmente é menor do que este período. A ausência ou mesmo o excesso de tráfego resultante destes fatores sazonais pode interferir significativamente na caracterização, estabelecendo tendências no *baseline* que não correspondem à realidade.

Os resultados obtidos demonstraram, de forma objetiva e prática, a validade do modelo levando em conta a comparação de forma visual do movimento real com seu *baseline*/DSNS. A utilização do *baseline* se mostrou bastante útil para administração da rede nos segmentos estudados. No segmento S_4 ocorreu um fato onde pôde-se observar muitas situações nas quais o *baseline* apresentava alterações significativas e importantes que demandavam controle e monitoramento pelo pessoal da gerência da rede estudada. Todos os dias da semana, no período da madrugada, são realizados neste segmento backup dos servidores da rede da UEL. Neste caso o controle da realização e da duração podia claramente ser verificado através da visualização do movimento e de seu respectivo *baseline*.

A seguir será demonstrada uma semana de cada segmento estudado, onde pode ser observado o movimento real que aparece em verde e o *baseline*/DSNS representado por uma linha continua em azul. Gráficos ilustrativos sobre os segmentos estudados:

1. Figura 3.69: mostra a 1ª semana de março de 2005 do segmento S_1 (firewall da rede da UEL) referente ao objeto *ifInOctets*,
2. Figura 3.70: apresenta a 2ª semana de outubro de 2004 referente ao segmento S_2 (servidor Web) para o objeto *ipInReceives*,
3. Figura 3.71: apresenta a 2ª semana de outubro de 2004 referente ao segmento S_3 (servidor de Proxy da UEL) e o objeto *ifOutOctets*,
4. Figura 3.72: mostra a 2ª semana de outubro de 2004 referente ao segmento S_4 (roteador principal da rede da UEL) e ao objeto *ifinOctets*,
5. Figura 3.73: ilustra a 2ª semana de setembro de 2004, referente ao segmento S_5 (pró-reitoria de assuntos acadêmicos da UEL) e o objeto *ifInOctets*,
6. Figura 3.74: ilustra a 2ª semana de novembro de 2003, referente ao segmento S_6 (ligação da rede da UNICAMP à Internet) e o objeto *ifInOctets*.

Algumas conclusões sobre os resultados demonstrados nas figuras:

1. Claramente pode-se identificar picos no movimento real e no *baseline*/DSNS que se referem a *backups* realizados no período da madrugada nos servidores de rede da UEL. A Figura 3.69 e a Figura 3.70 apresentam estes picos iniciando as 02:30 horas e terminando as 04:15 horas e novamente das 04:50 as 04:45 horas. No sábado e domingo somente é realizado o 2º backup nos servidores de rede que se inicia as 04:45 horas, conforme é retratado pelo pico de tráfego neste período;
2. O *baseline*/DSNS, sofre nos casos estudados, influência de fatores temporais que, neste caso, está relacionado com o horário do expediente da Universidade, que se inicia as 8:00 horas e termina aproximadamente as 22:00 horas. Esta característica pode ser observada nos exemplos

apresentados nas figuras Figura 3.69, Figura 3.70, Figura 3.71, Figura 3.72, Figura 3.73 e Figura 3.74;

3. Pode-se observar nas figuras a seguir, períodos onde o tráfego normal ou movimento real ao longo do dia se torna maior que o *baseline*/DSNS. Neste caso, sua cor é alterada de verde ou cinza claro (caso este trabalho esteja sendo lido em preto e branco) para vermelha ou cinza escuro. O objetivo desta alteração é chamar a atenção para picos superiores ao *baseline*/DSNS. Nas figuras aparecem o indicativo de alarmes, porém, isto é colocado somente como um indicativo de alerta, podendo ou não ser interpretado como alarme dependendo exclusivamente do administrador;

4. Em todos os casos demonstrados nas figuras a seguir, se observa que nos finais de semana o volume de tráfego diminui sensivelmente. Estas características vêm a confirmar a necessidade de se criar *baselines*/DSNS distintos para cada dia da semana;

5. Os resultados apresentados nesta seção somente utilizaram *baselines* gerados pelo modelo ***bl-7*** que estabelece um *baseline*/DSNS para cada dia da semana. Contudo, o modelo ***bl-3*** que gera um único *baseline*/DSNS para os dias úteis também poderia ter sido utilizado. Em algumas situações a utilização tanto do *baseline* gerado para o ***bl-7*** como para o ***bl-3*** apresentam resultados semelhantes, podendo ser utilizado qualquer um, que o resultado seria o mesmo. Isto pode ser obtido em casos onde não exista nenhum tipo de característica singular presente durante os dias da semana. No exemplo apresentado na Figura 3.70, existe uma situação onde a utilização do ***bl-3*** não seria adequada. Na segunda feira, dia 04/10/2004, pode-se observar que não ocorre à realização do *backup* no horário das 02:30 as 04:15 horas. Neste caso, se fosse utilizado o *baseline*/DSNS gerado pelo modelo ***bl-3***, o *baseline* gerado seria como o apresentado na Figura 3.75, onde é apresentado o mesmo dia só que com o *baseline* gerado pelo modelo ***bl-3***. Esta figura demonstra claramente a previsão de um volume de tráfego que na

verdade não ocorre neste horário e dia pelo modelo ***bl-3***. Neste exemplo, verifica-se uma situação real onde a utilização do modelo ***bl-3*** não é adequada para o segmento;

6. As imagens apresentadas referentes aos seis segmentos analisados, demonstram obviamente de forma visual, que o modelo BLGBA para criação de *baseline*/DSNS atinge seu objetivo no tocante a criar uma caracterização particular para o segmento analisado.

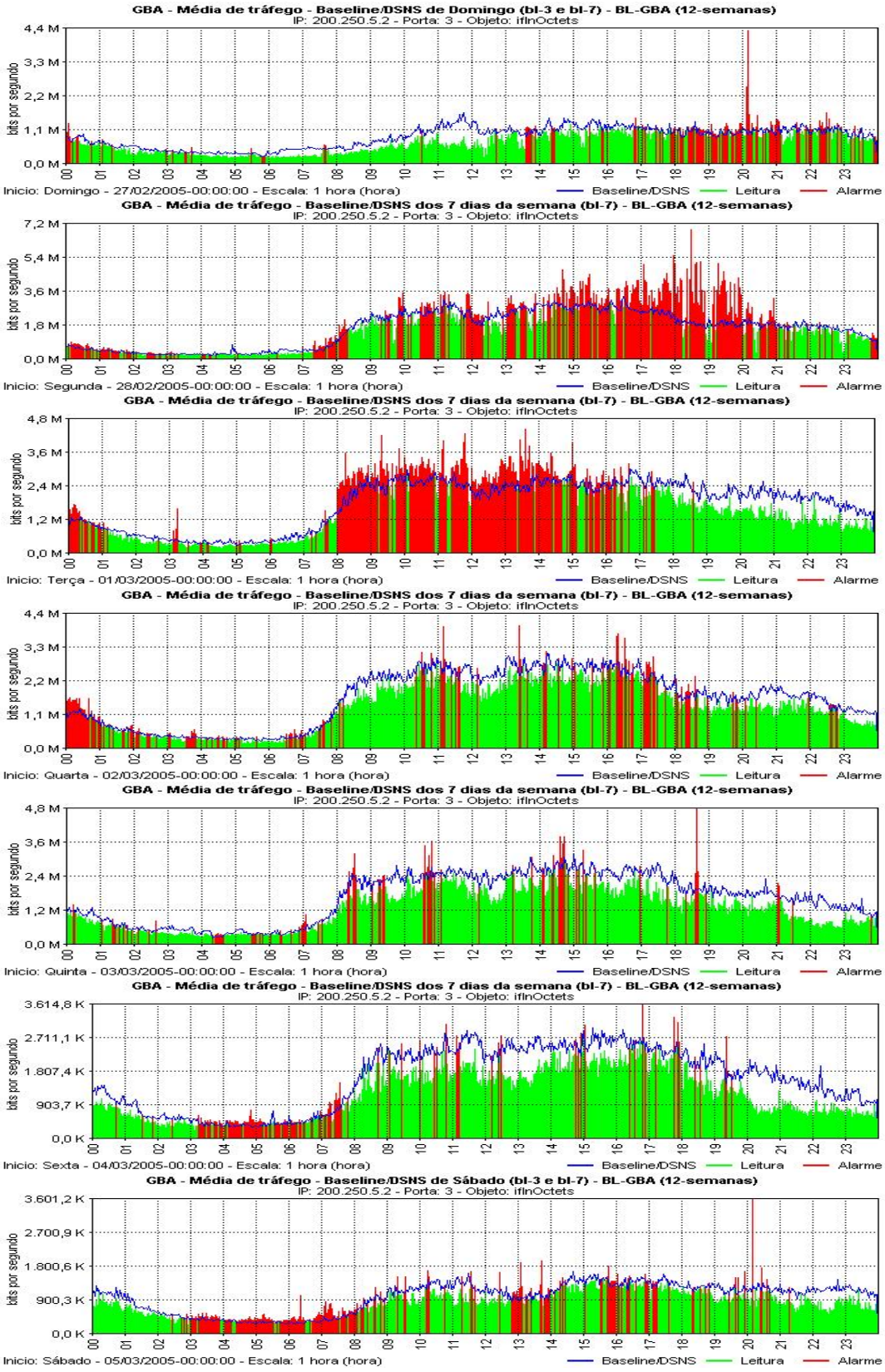


Figura 3.69 - 1ª semana de março de 2005 do segmento S_1 objeto *ifInOctets*.

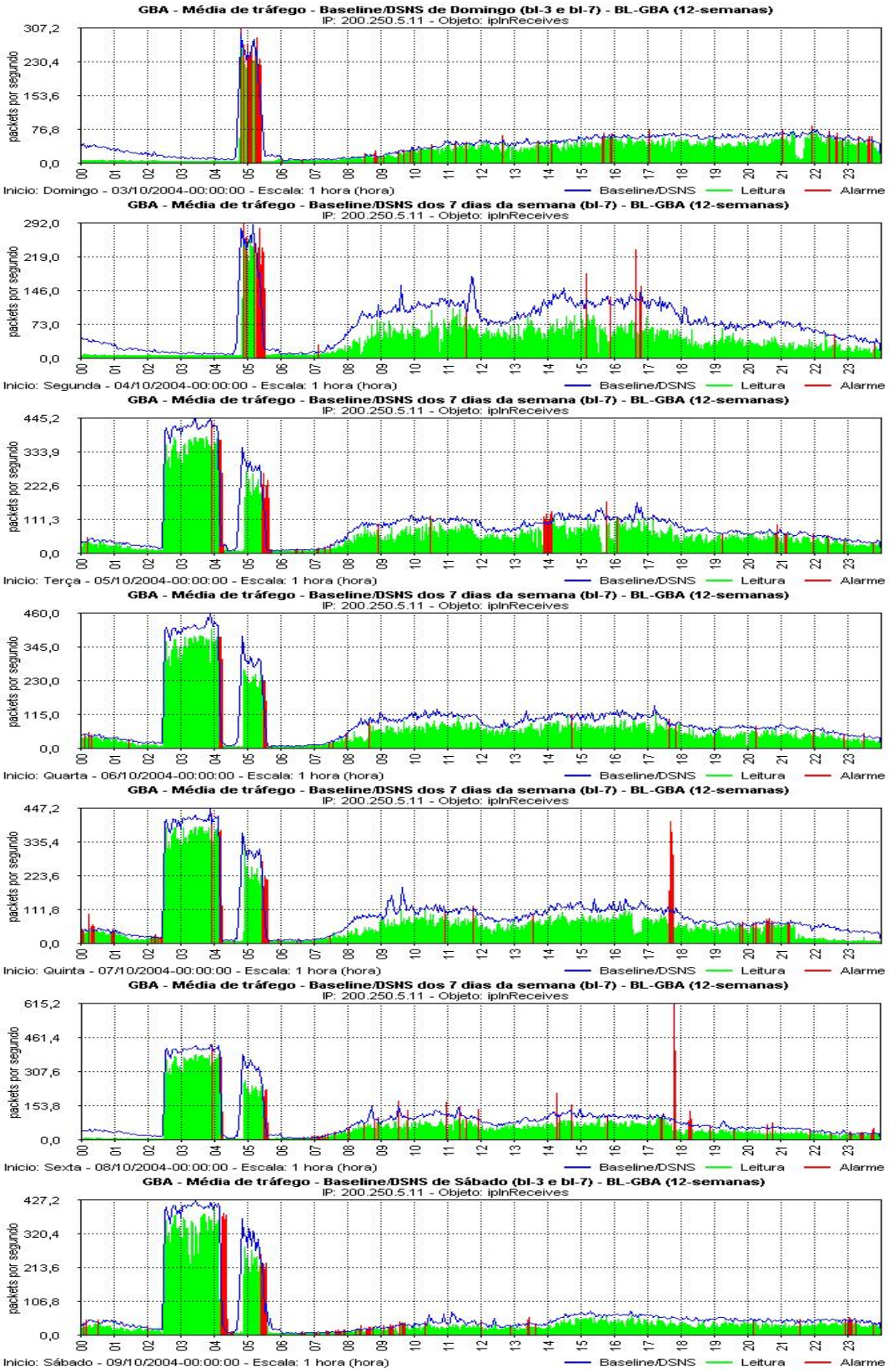


Figura 3.70 - 2ª semana de outubro de 2004 referente ao segmento S_2 objeto *ipInReceives*.

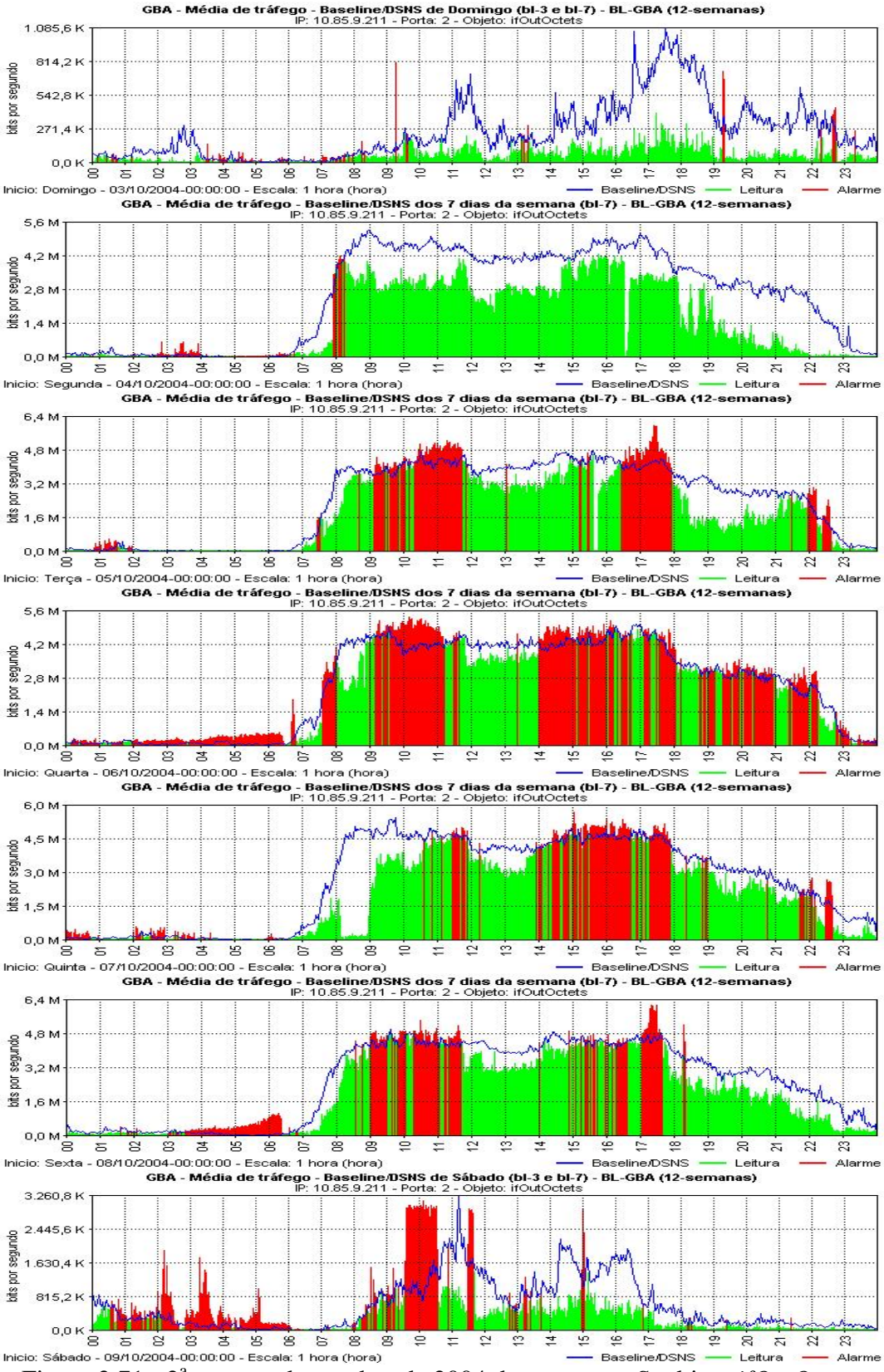


Figura 3.71 - 2ª semana de outubro de 2004 do segmento S_3 objeto *ifOutOctets*.

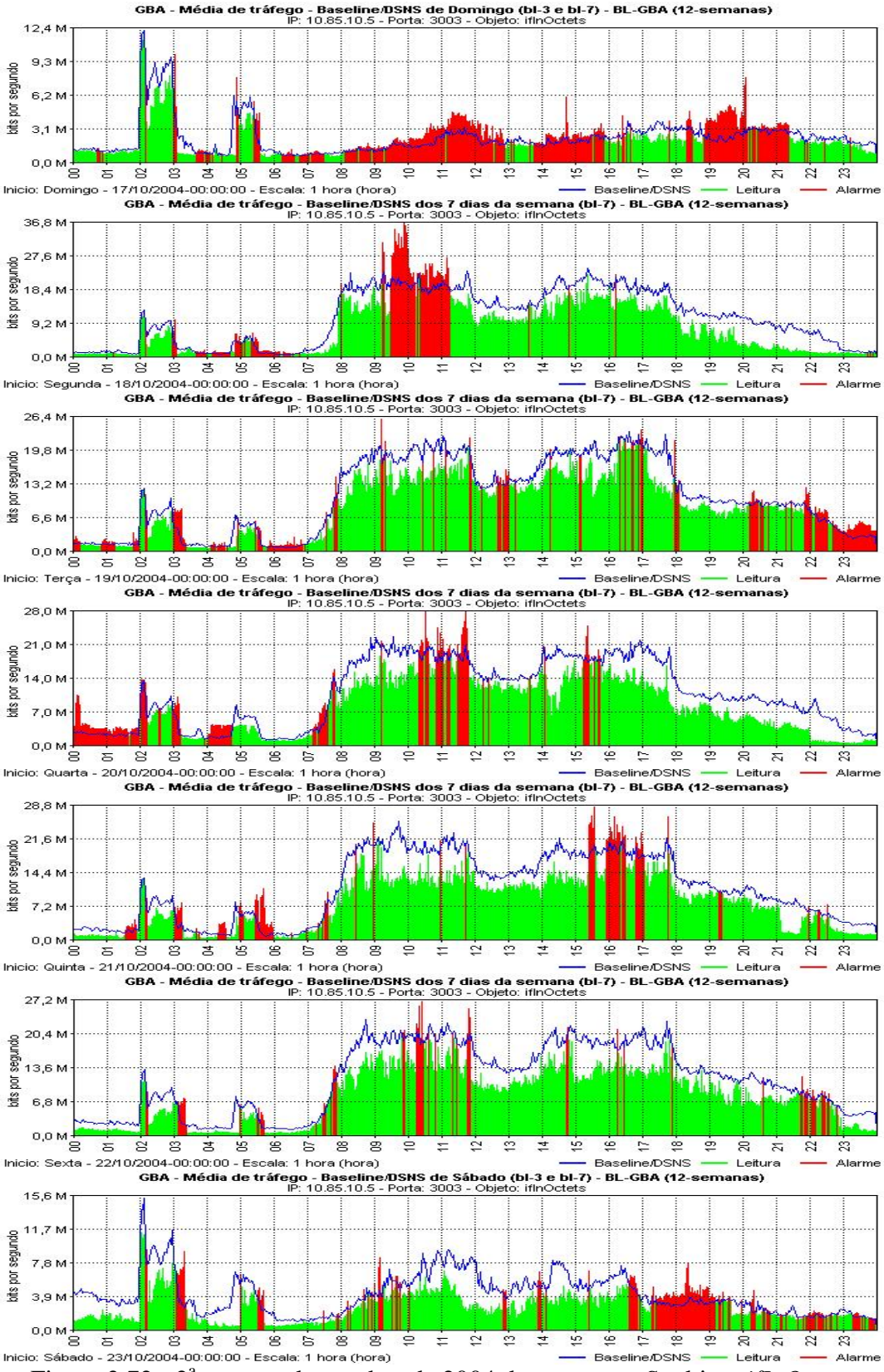


Figura 3.72 - 2ª semana de outubro de 2004 do segmento S_4 objeto *ifInOctets*.

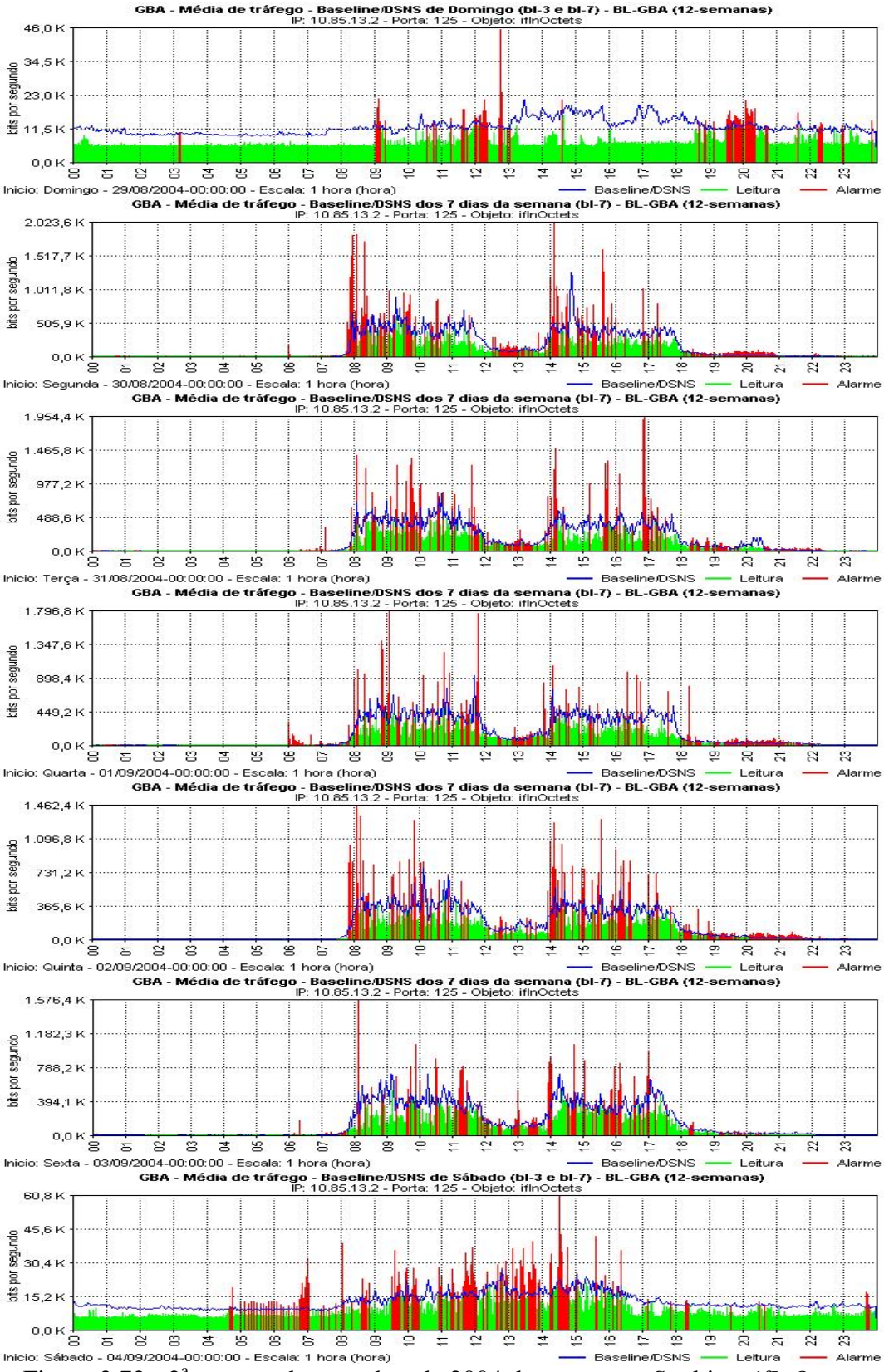


Figura 3.73 - 2ª semana de setembro de 2004 do segmento S_5 objeto *ifInOctets*.

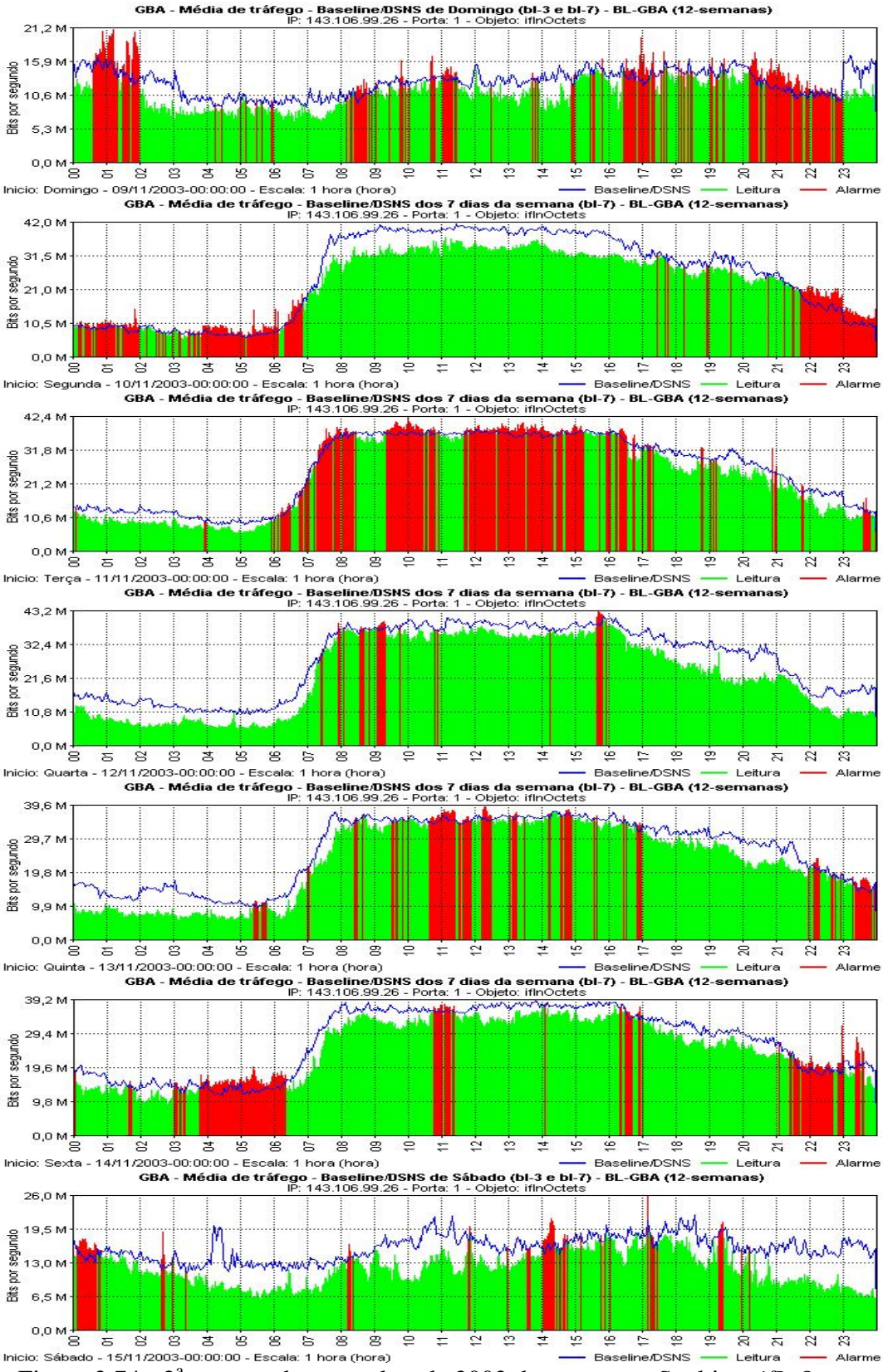


Figura 3.74 - 2ª semana de novembro de 2003 do segmento S_6 objeto *ifInOctets*.

Figura 3.75 - *baseline bl-3* e movimento do segmento S_2 em 04/10/2004.

Durante a realização dos testes, surgiram vários exemplos de situações definidas como diferenciadas, por não estarem de acordo com a normalidade esperada pelo *baseline* e que foram identificadas justamente pela utilização do mesmo. Figura 3.76 demonstra o movimento e seu respectivo *baseline* referente à semana de 26/09/2004 a 01/10/2004 monitoradas do segmento S_2 e S_4. Nestas figuras se observa a ocorrência de *backup* no período das 02:30 horas e 05:00 horas da madrugada. Pontos que devem ser salientados referentes a esta semana:

1. Dia 27/09, Figura 3.76 (b) e (i), segunda feira, nota-se que a partir das 21:30 horas, o movimento teve uma alteração significativa comparado a seu *baseline* indicando redução do tráfego. Já no dia 02/10, sexta feira, a partir das 20:00 horas, também houve uma diminuição no tráfego no segmento S_2 Figura 3.76 (g), porém, esta alteração não pode ser percebida no segmento $S_{4;}$

2. Dia 01/10, o *backup* que é realizado as 05:00 horas não ocorreu. A ausência de tráfego neste período sinaliza este fato, que aconteceu em conseqüência de uma falha humana e pode ser observado na Figura 3.76 (f). A Figura 3.76 (m) ilustra o reflexo da mesma situação acontecendo no segmento S_4 com monitoramente do objeto *ifInOctets* e no mesmo horário. Outra situação que ocorreu no mesmo dia e que também pode ser notado nesta figura é a diminuição no volume de tráfego a partir das 22:00 horas;

3. Para os outros dias da semana, observa-se que tanto o *baseline* como o movimento real apresentaram as mesmas características demonstrando um bom ajuste do modelo.

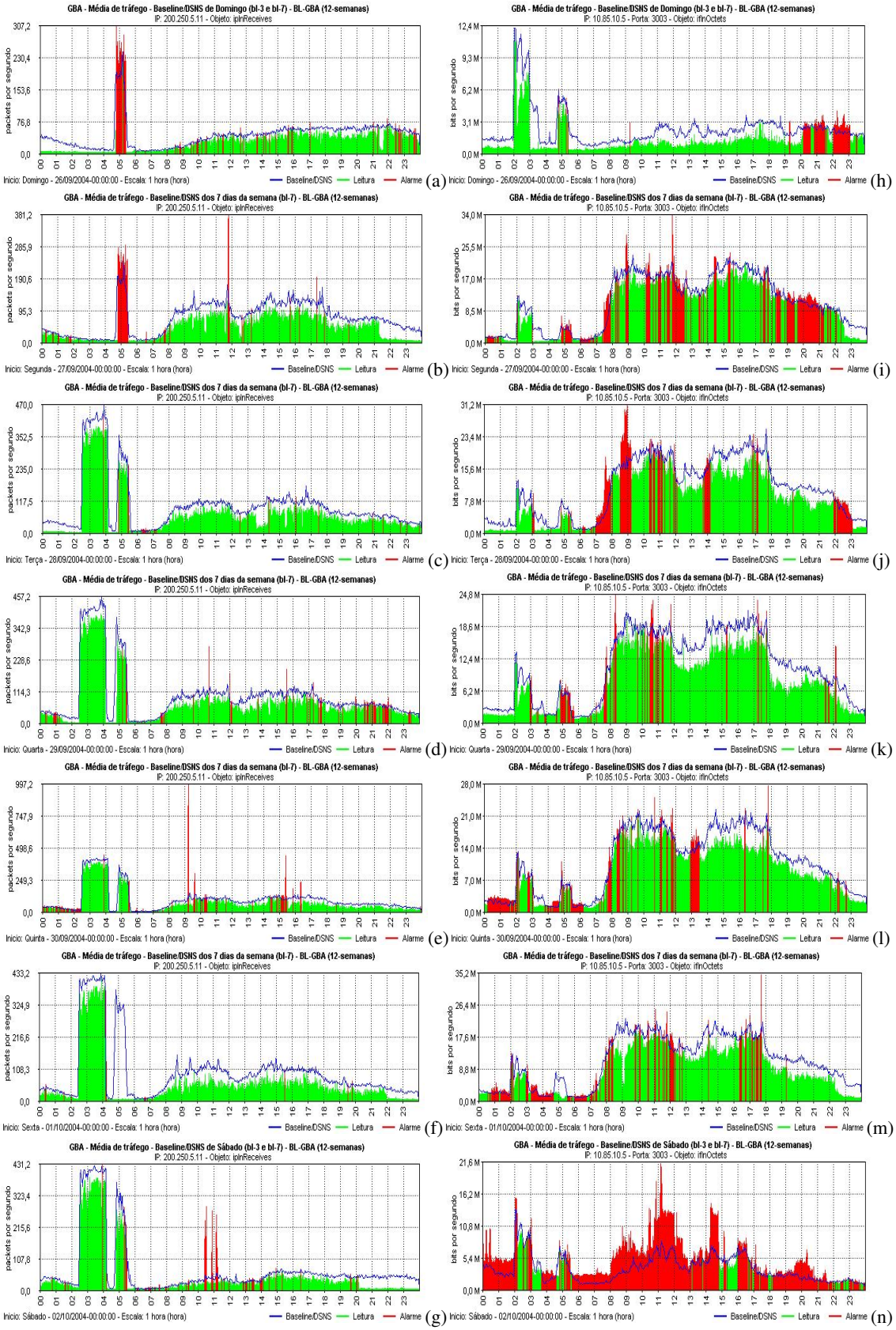


Figura 3.76 - Semana de 26/09/2004 a 02/11/2004 do segmento S_2 e do S_4.

Uma outra situação interessante é apresentada na Figura 3.77 (a), (b), (c), (d), (e), (f) e (g), onde podem ser constatados, picos de trafego que iniciam as 04:45 horas e terminam as 05:30 horas. Estes picos se referem à realização de um *backup*, neste segmento, sinalizada no movimento real, porém, ainda não retratada no *baseline*. Este *backup* foi iniciado pelos administradores justamente nesta semana, por isso ainda não tinha sido "aprendido" pelo *baseline*. Já na Figura 3.77 (h), (i), (j), (k), (l), (m) e (n), é apresentado novamente o mesmo segmento só que no período de 05/09/2004 a 11/09/2004. Neste novo período observa-se que o *backup* já foi incorporado ao *baseline*. Nota-se também que no domingo e segunda-feira, Figura 3.77 (h) e (i) o *backup* foi iniciado em horário diferente do previsto, isto ocorreu por falha no setor de operação, constatada posteriormente. Outro ponto importante apresentado nesta semana é o feriado de 07 de setembro, onde observa-se o *baseline* e a ausência de movimento durante todo o dia e inclusive do *backup* que se inicia as 04:45. Isto também ocorreu por outra falha no setor de operação.

Os casos citados nesta seção foram utilizados para exemplificar de forma prática e com situações reais, vantagens na utilização do *baseline*. Inúmeros outros casos semelhantes, ocorreram durante a realização deste trabalho, que reafirmam vantagens em se ter um *baseline* específico para cada segmento de rede. Situações que envolvem, por exemplo, falhas de hardware ou software, que provocaram ausência de volume de tráfego nos segmentos monitorados, ou excesso de tráfego tendo em vista falhas ou transferência de dados não autorizados, foram identificadas de forma mais rápida, justamente pela facilidade obtida pela comparação do *baseline* com o movimento real.

Obviamente que resultados melhores foram obtidos com a automação destes monitoramentos, através de um sistema de alarmes integrados com a coleta de informações que é realizada pela ferramenta GBA. O próximo capítulo apresenta um sistema de alarmes e de detecção de anomalia que também foi desenvolvido durante a realização deste trabalho.

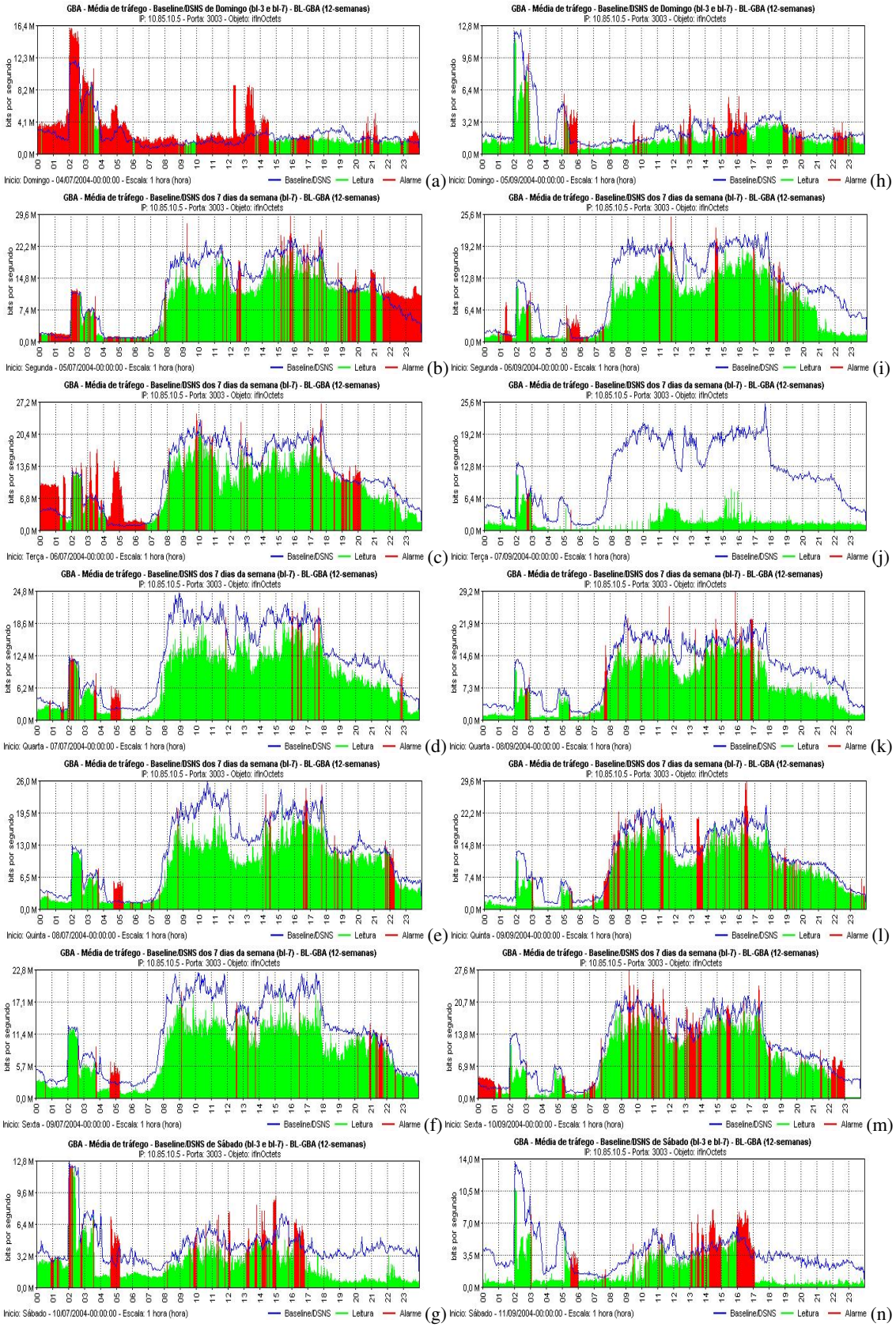


Figura 3.77 – Exemplo de semanas do S_1 referente aos meses 07 e 10 de 2004.

4 Alarmes e Anomalias

O desenvolvimento do modelo BLGBA para criação de *baseline*/DSNS, baseado em análises estatísticas realizadas com dados históricos referentes ao movimento real do segmento analisado, possibilitou a solução de outro importante problema relacionado à gerência de redes. Ele se refere à construção de um sistema de alarmes preciso e adaptável às características do segmento monitorado.

Os modelos existentes atualmente para geração de alarmes, disponíveis nos sistemas de gerenciamento de redes (NMS), se baseiam somente em limites lineares configurados pelo administrador da rede (Barford *et al.,* 2002). Este tipo de sistema de alarmes, na prática, não atende as demandas existentes para um controle preciso, efetivo e confiável relacionados à gerência de performance e segurança.

Outro fator muito importante e deficiente atualmente com relação de *softwares* de gerência de redes, que pode ser resolvido com a implantação de um sistema de alarmes baseado no *baseline,* é a possibilidade de se automatizar o monitoramento de segmentos, baseados somente em controle visual dos gráficos gerados a partir de ferramentas como o MRTG (MRTG, 2005) e a ferramenta GBA. Esta prática é muito comum atualmente nos centros de operações de rede (*Network Operations Center* - NOC). Porém, acaba sendo prejudicada ou mesmo inviabilizada na medida em que se têm muitos gráficos para serem monitorados. A definição do que é ou não normal, não é fornecida por estas ferramentas. Esta definição de normalidade ou *baseline* do segmento, o administrador da rede deve fazer com base em seus conhecimentos empíricos adquiridos sobre o funcionamento da rede. Na verdade, o que se observa é a disponibilização de muitos gráficos, com diversas informações sobre os segmentos monitorados, porém, sem nenhum valor agregado que possa contribuir para sua análise automatizada. Esta automação será possível com a utilização do *baseline* em conjunto com um sistema de alarmes.

A meta que buscamos no desenvolvimento do sistema de alarmes e anomalias é informar ao administrador de rede o exato momento em que um evento anômalo estiver

acontecendo. Para tanto, desenvolvemos o sistema de detecção de anomalias ADGBA, que atua em tempo real com o monitoramento realizado pela ferramenta GBA. Este sistema analisa informações lidas da MIB dos equipamentos de rede e compara com seu respectivo *baseline*/DSNS.

Neste trabalho definiu-se como anomalia, o conjunto de eventos que indiquem uma variação significativa do movimento real analisado em relação ao que foi predito no seu *baseline*/DSNS. Esta variação pode ocorrer por problemas físicos em algum componente ou equipamento da rede, excesso de carga na rede, ou mesmo algum tipo de problema relacionado à segurança como ataques ou intrusões (Thottan, 2003) (Xinzhou *et al.,* 2002) (Lakhina *et al.,* 2004).

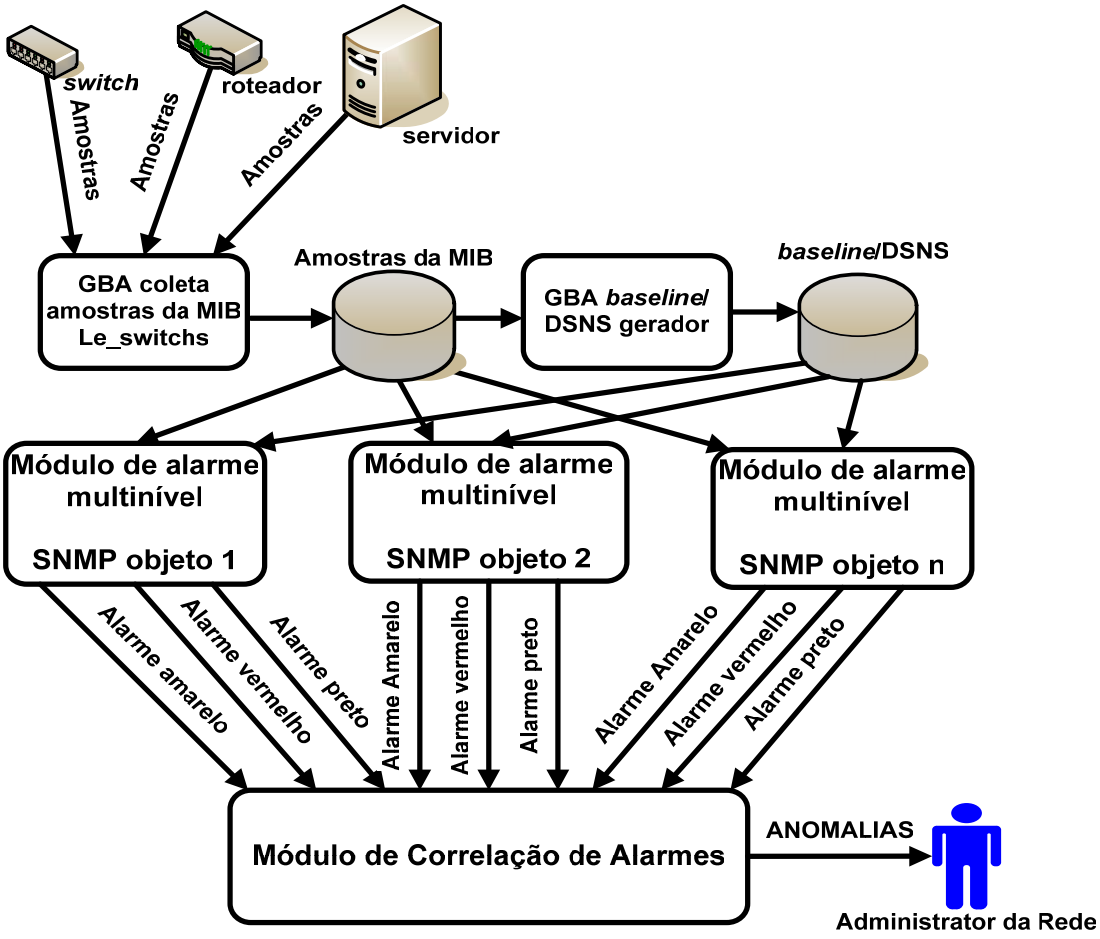


Figura 4.1 - Modelo de referência do Sistema de Detecção de Anomalias (ADGBA).

A Figura 4.1 apresenta o modelo de referência do sistema ADGBA para detecção de anomalia apresentado neste trabalho. O modelo é composto por dois módulos. O primeiro tem a função de geração de alarmes com três níveis de sensibilidade (*amarelo*, *vermelho* e *preto*), e o segundo é responsável pela correlação dos alarmes gerados para múltiplos objetos SNMP, com a finalidade de detectar a ocorrência de anomalias.

O sistema ADGBA analisa as amostras coletadas pela ferramenta GBA, segundo a segundo, e informa ao administrador da rede se alguma anomalia for detectada. Para evitar falsos alarmes, adicionamos ao processo de detecção de anomalias, a correlação entre eventos que sinalizem desvio do comportamento real do tráfego em relação ao esperado pelo *baseline*/DSNS, e que ocorram em mais de um objeto SNMP ao mesmo tempo. Estes eventos serão detectados pelo sistema de alarmes multinível que será apresentado a seguir.

A construção do sistema de alarmes foi baseada em limites estabelecidos pelo *baseline*/DSNS e um mecanismo de histerese. A idéia central era que o administrador somente iria receber alarmes se ocorressem desvios do comportamento normal que justificasse sua atenção. O simples desvio do movimento real em relação ao *baseline* poderia disparar um alarme, portanto, constatou-se a geração de um número muito grande de alarmes que inviabilizaria a utilização do sistema.

Durante os testes comprovou-se que a simples utilização de alarmes baseados somente no *baseline*/DSNS em relação do movimento real, causava um número muito grande de alarmes, da ordem de 15% do movimento, ou seja, 12.960 alarmes seriam gerados por dia. Isto nos permitiu constatar, na prática, uma das características do tráfego de redes *ethernet* que é extremamente aleatório e não comportado, constantemente formado por rajadas. Estas características nos levaram a constatar que a utilização de alarmes baseados simplesmente no *baseline*/DSNS seria inviável em razão da grande quantidade de alarmes gerados. Estes testes foram realizados para o limite superior estabelecido pelo *baseline*.

Para obter maior precisão na detecção de anomalias e reduzir o número de falsos alarmes, foi adicionado ao sistema de alarmes um conjunto de regras que o tornou baseado em três níveis de sensibilidade. O primeiro, chamado *amarelo*, correspondente ao nível mais sensível a variações do tráfego em relação ao *baseline*/DSNS, devendo ser utilizado

em casos que se requer alta sensibilidade às variações do movimento real em relação ao seu *baseline*. O segundo, chamado de *vermelho*, é um nível intermediário, que pode ser usado em situações que não requerem um alto nível de sensibilidade, mesmo sem a perda da confiabilidade. O último, chamado de *preto*, é o nível de monitoramento mais forte, onde o administrador da rede somente será informado se houver um desvio muito significativo do tráfego em relação ao DSNS.

O mecanismo criado para detecção de anomalias estabelece uma janela de tempo para detecção, a qual foi chamada de janela de histerese. Nesta janela, os desvios do comportamento normal em relação ao previsto pelo *baseline*/DSNS são analisados e correlacionados com objetivo de se detectar eventos anômalos. O objetivo da janela de histerese é a redução do número de falsos alarmes que normalmente são gerados devido a uma ou mais rajadas ocasionais no tráfego. O tamanho da janela é expresso em segundos e dado pela variável ***t*.**

Com a criação de três níveis de alarmes, estabeleceu-se três tamanhos para janelas de histerese, uma para cada nível de alarme. Para alarmes *amarelos* $\boldsymbol{t}$ = 300 segundos, alarmes *vermelhos* $\boldsymbol{t}$ = 600 segundos e alarmes *pretos* $\boldsymbol{t}$ = 900 segundos.

A janela de histerese estabelecida para redução de falsos alarmes, não se mostrou suficiente para o estabelecimento de um mecanismo eficiente e seguro para detecção de eventos anômalos, por isso, foi criado também um conjunto de regras que devem ser consideradas para geração de alarmes no intervalo $\boldsymbol{t}$. Somente se as três regras abaixo forem quebradas o alarme será gerado. A expressão (4.1) ilustra este conjunto de regras criadas como gatilho para a geração de alarme.

$$(\forall x)(P(x, y) \wedge Q(x,t) \wedge P(x, v) \wedge P(z, \delta)) \rightarrow x \in A \qquad \textbf{(4.1)}$$

Onde A = alarme gerado; $P(x, y)$ = verdade, se amostra (x) é maior que *baseline* (y); $P(x, v)$ = verdade, se amostra (x) é maior anterior que ultrapassou *baseline* (v); $Q(x, t)$ = amostra (x) esta no intervalo $\boldsymbol{t}$; $P(z., \delta)$ = Amostras que violaram as regras 1 e 2 e são maiores que δ.

A descrição textual destas regras é apresentada a seguir:

- Regra número 1: A amostra analisada é maior do que o limite superior ou inferior estabelecidos pelo DSNS. Em (4.1) a amostra lida é representada por x e o seu respectivo *baseline*/DSNS é representado por y.

- Regra número 2: é maior que a última amostra que violou a regra número 1 dentro do mesmo intervalo t. Em (4.1) a amostra anterior é representada por *v*.
- Regra número 3: A quantidade de amostras que violaram as regras número 1 e 2 são maiores do que δ. Em (4.1) esta quantidade é representada por *z*.

A regra número 3 foi incluída com objetivo de minimizar o número excessivo de alarmes gerados pelo mesmo evento, tipicamente rajadas ocasionais que são muito comuns no tráfego de redes *ethernet*. Os estudos e testes demonstraram ser inversamente proporcional à quantidade de alarmes gerados em relação a δ. Com isso, temos que quanto maior δ, menor seria a quantidade de alarmes gerados. Após vários testes e análises práticas nos segmentos estudados, conclui-se que para se obter uma boa relação entre a quantidade de alarmes que realmente indicariam problemas, teríamos o δ com valores de 130, 260 e 390, associados aos intervalos de histereses de 5, 10 e 15 minutos, respectivamente, para alarmes *amarelo*, *vermelho* e *preto*. A Figura 4.2 apresenta o autômato do algoritmo implementado para avaliação das três regras no sistema de alarmes usado pelo ADGBA.

Após a implementação do mecanismo de histerese em conjunto com o uso do acumulador de violações δ, constatou-se um número muito pequeno de alarmes durante o dia nos segmentos analisados, sinalizando efetivamente quando ocorriam mudanças significativas do movimento real, em relação a seu *baseline*/DSNS, que deveriam ser analisadas pelo administrador da rede. A utilização do acumulador δ foi a solução encontrada para viabilizar a utilização dos alarmes, baseados em análises segundo a segundo. Este fator, que introduziu um redutor ao número de alarmes gerados dentro de um mesmo intervalo de histerese, se mostrou eficiente e eficaz na agregação de eventos correlatos. Notou-se que neste nível de abstração, a agregação de evento deveria ser introduzida com objetivo de reduzir a quantidade de alarmes repetidos para mesma situação. As constantes apresentadas para o intervalo de histerese foram escolhidas com base na relação evento e necessidade de sinalização. Esta relação foi criada com base em uma simulação, realizada em função da quantidade de alarmes gerados por intervalo de tempo que demarca a janela de histerese.

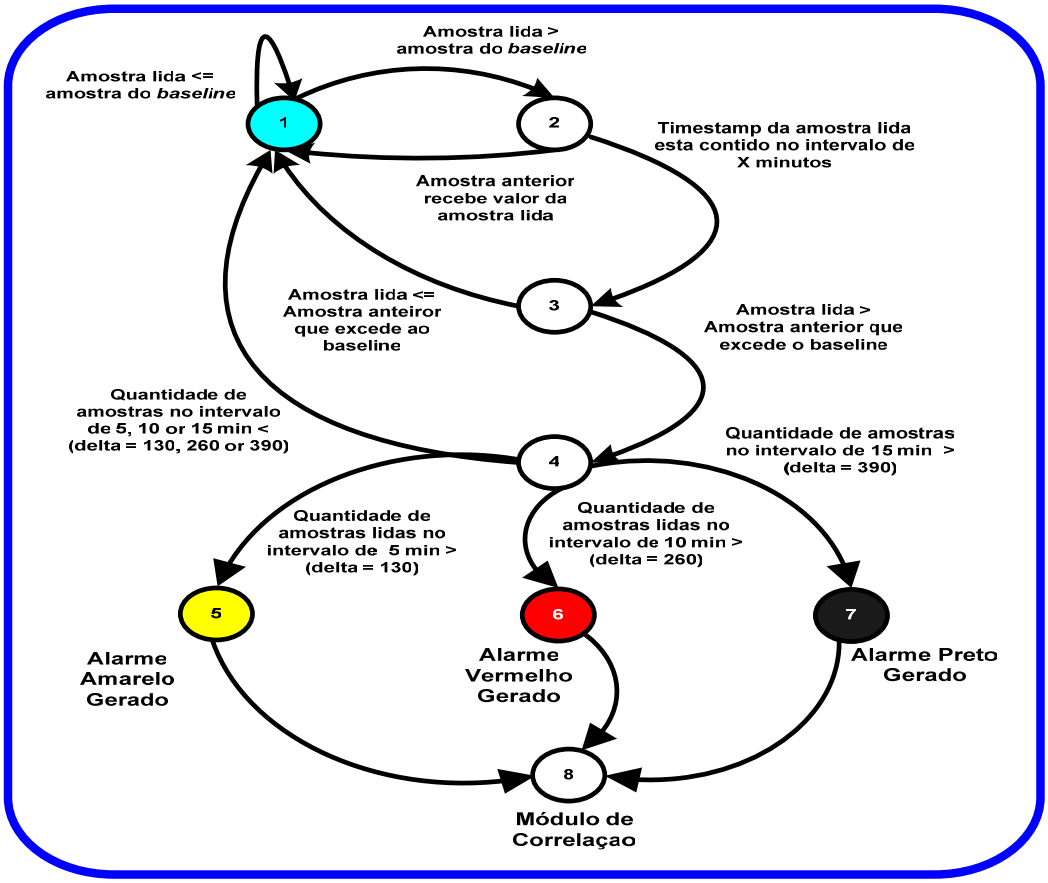


Figura 4.2 - autômato para geração de alarmes multinível do sistema ADGBA.

O módulo de correlação recebe informações do módulo de alarmes multinível. Seu objetivo é fazer a correlação de alarmes que ocorreram no mesmo intervalo de histerese e em diferentes objetos SNMP. Uma anomalia será detectada e notificada se o total de alarmes gerados para diferentes objetos, no mesmo intervalo de histerese, seja maior ou igual a ***n***, onde ***n*** é igual ao total de objetos analisados. A Figura 4.3 apresenta o autômato responsável pelo módulo de correlação usado pelo sistema ADGBA.

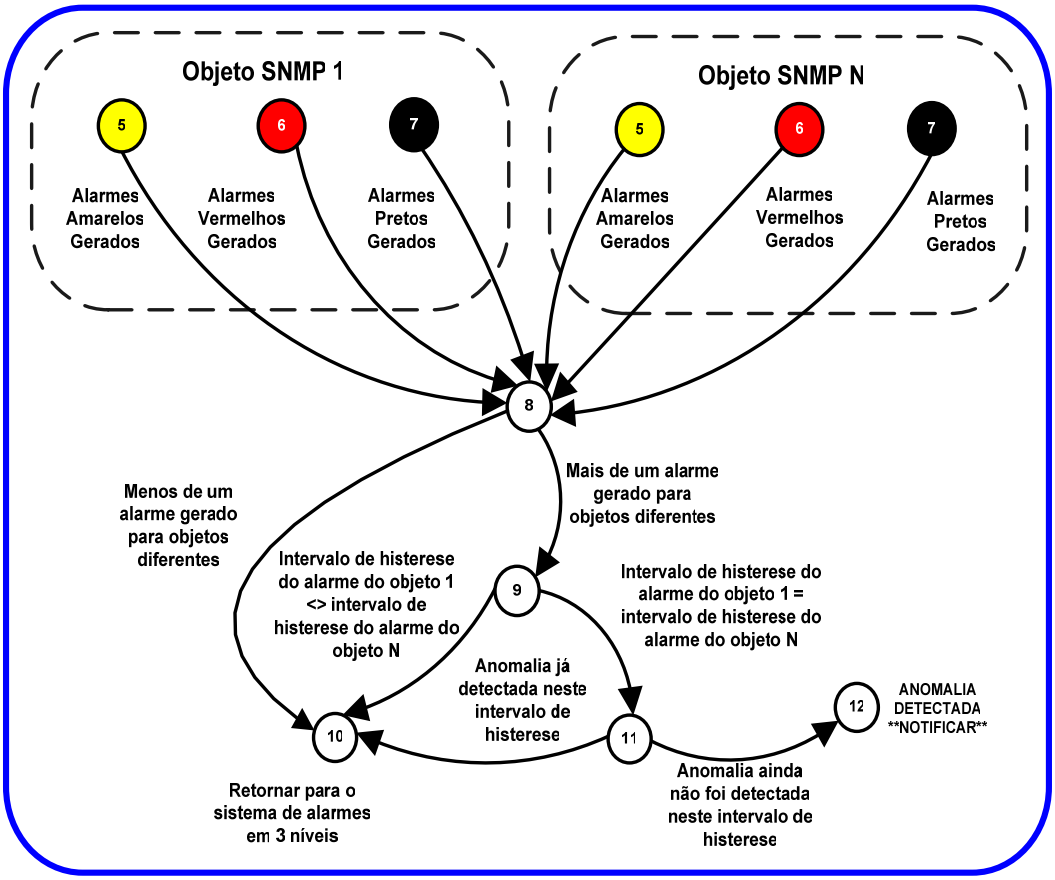


Figura 4.3 - Autômato para correlação de alarmes do ADGBA.

Para tratar dos limites inferiores ao *baseline*, foi adotada uma variação percentual fixa de 40 % em relação ao sugerido pelo *baseline*, ou seja para se definir o limite inferior foi utilizado o *baseline* – 40 % de seu valor. Este percentual aplicado foi utilizado com base nas necessidades apresentadas pelo centro de operações da rede da UEL, para ser usado durante os dias úteis, ou seja, de segunda a sexta feira, porém notou-se que em períodos de ausência de tráfego como finais de semana ou feriados, este valor não é adequado, pois as variações do tráfego nestes períodos, pode chegar a 100 % em relação ao *baseline*. Isto se deve principalmente a ausência de regularidade imposta por estes períodos, onde a utilização da rede por uma ou mais pessoas pode facilmente descaracterizar o perfil de normalidade apresentado pelo *baseline*.

A Figura 4.4 apresenta em forma de histograma a média mensal de alarmes que ocorreram de julho a dezembro de 2004, para o servidor WEB S_2, gerados pelo sistema de alarmes multinível para os objetos *ifInOctets, ipInReceives* e *tcpInSegs*. A Figura 4.5 apresenta também a média mensal de alarmes para o mesmo período, gerados para os objetos *ipInReceives* e *tcpInSegs* referentes ao servidor de Proxy da rede da UEL, chamado também de S_3. Nestas figuras, podem ser observados de forma isolada a média dos alarmes *amarelos*, *vermelhos* e *pretos* para os dois exemplos.

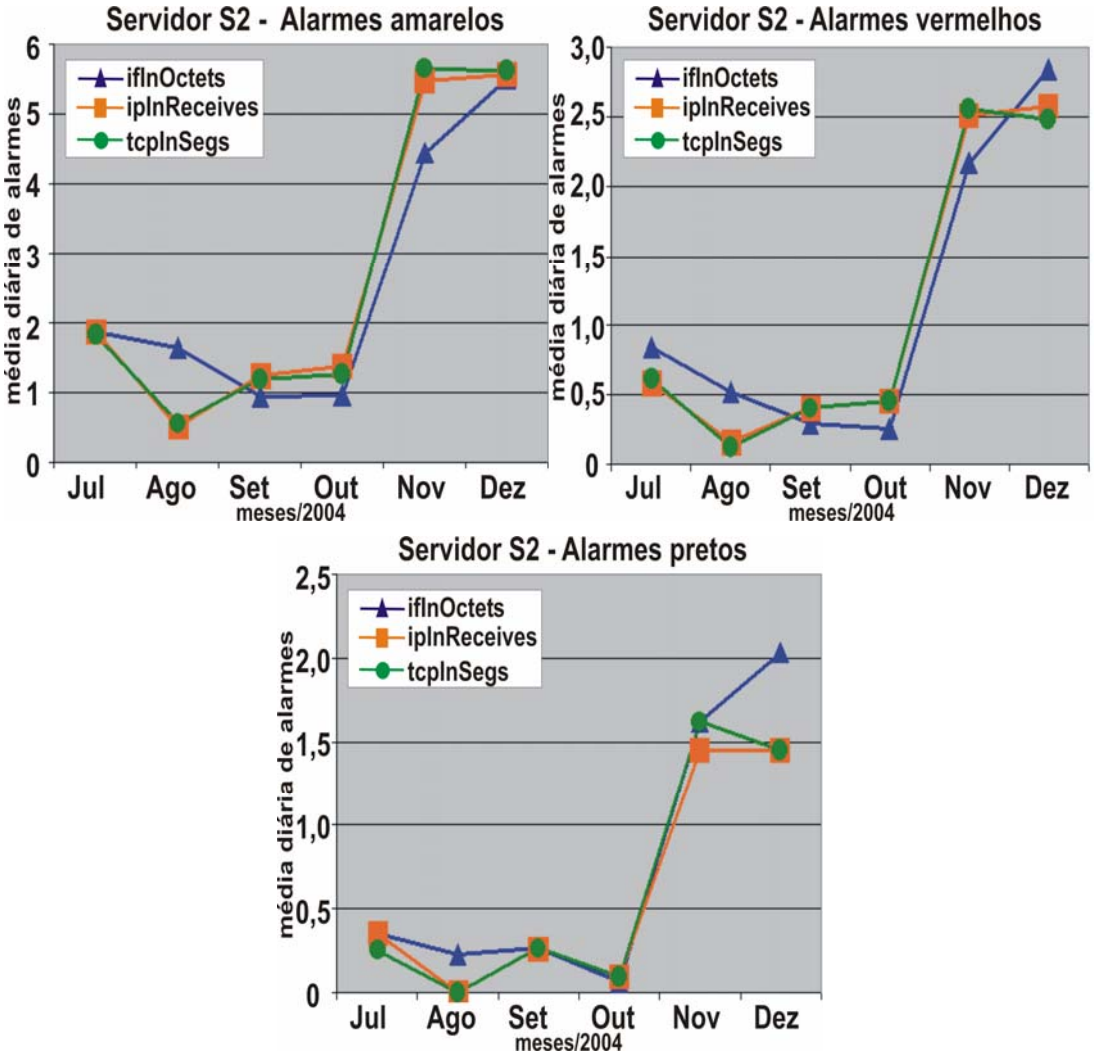


Figura 4.4 - Média de alarmes para o servidor S_2 de julho a dezembro de 2004.

Outra conclusão obtida neste trabalho, e que pode ser observada nas Figuras 4.4 e 4.5, é uma proximidade na média dos alarmes para os diferentes objetos. O que se observou nestes casos é que esta proximidade indicava que os alarmes foram gerados para os mesmos eventos, porém, em objetos diferentes, porém em intervalos de histerese similares.

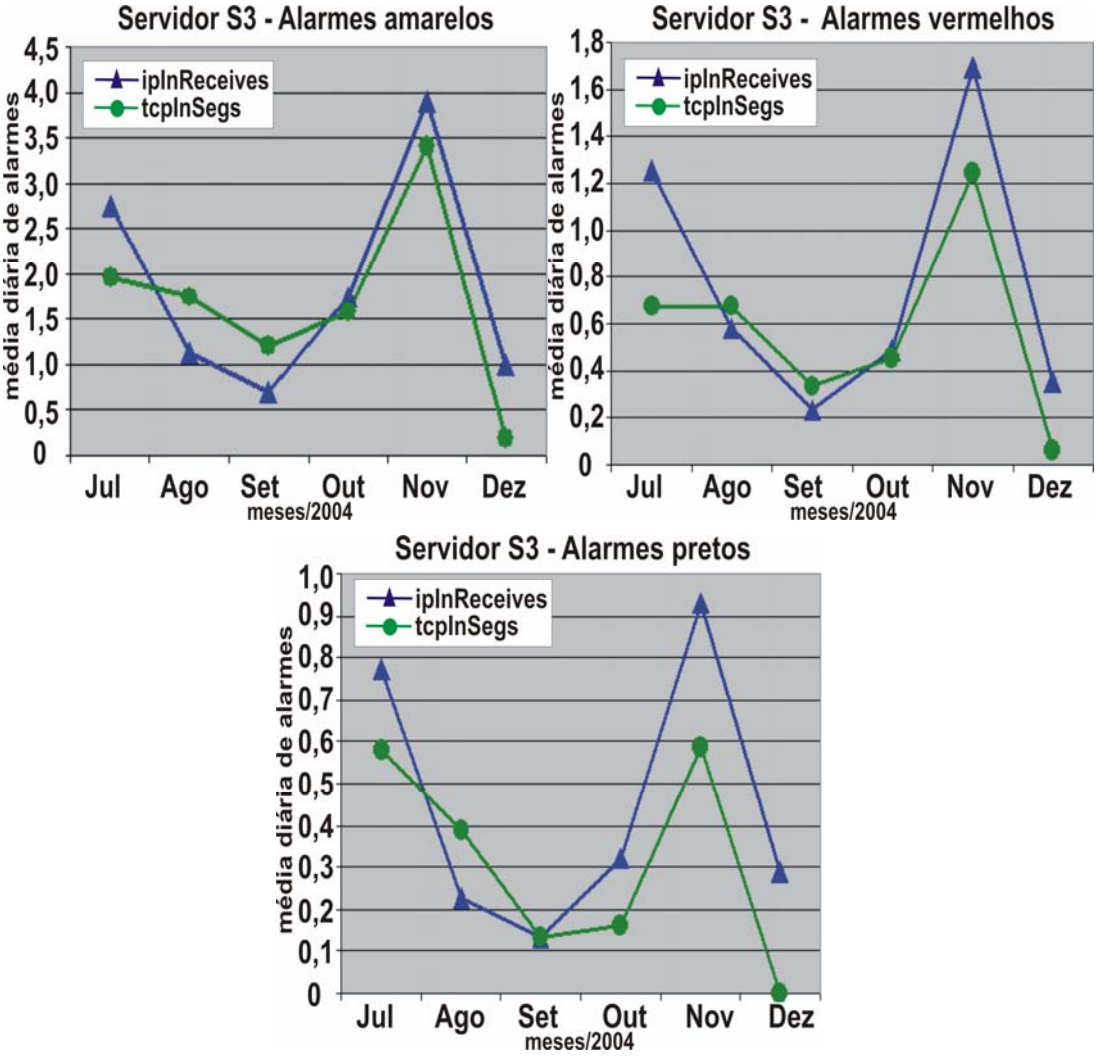


Figura 4.5 - Média de alarmes para o servidor S_2 de julho a dezembro de 2004.

Tabela 4.1 - Média de anomalias ocorridas de julho a dezembro/2004 no servidor S_1.

Firewall S_1							
	Amarelo		Vermelho		Preto		Anomalias
	ifIO	ipIR	ifIO	ipIR	ifIO	ipIR	
Jul	2.84	1.45	1.39	0.81	0.77	0.42	0.06
Ago	1.35	1.23	0.42	0.45	0.13	0.39	0.03
Set	8.07	1.40	3.50	0.67	2.33	0.47	0.07
Out	7.48	0.71	3.26	0.26	2.00	0.13	0.03
Nov	3.90	0.31	1.72	0.10	1.07	0.00	0.10
Dez	2.61	1.32	1.23	0.52	0.87	0.29	0.06

Tabela 4.2 – Média de anomalias ocorridas de julho a dezembro/2004 no servidor S_2.

Servidor Web S_2										
	Amarelo			Vermelho			Preto			Anomalias
	ifIO	ipIR	tcpIS	ifIO	ipIR	tcpIS	ifIO	ipIR	tcpIS	
Jul	1.84	1.87	1.84	0.87	0.58	0.61	0.35	0.35	0.26	0.35
Ago	1.65	0.52	0.55	0.52	0.16	0.13	0.23	0.00	0.00	0.26
Set	0.93	1.23	1.20	0.30	0.40	0.40	0.27	0.27	0.27	0.27
Out	0.97	1.39	1.26	0.26	0.45	0.45	0.06	0.10	0.10	0.52
Nov	4.45	5.48	5.66	2.17	2.52	2.55	1.62	1.45	1.62	0.63
Dez	5.52	5.58	5.65	2.84	2.58	2.48	2.03	1.45	1.45	0.77

Tabela 4.3 - Média de anomalias ocorridas de julho a dezembro/2004 no servidor S_3.

Proxy S_3							
	Amarelo		Vermelho		Preto		Anomalias
	ipIR	tcpIS	ipIR	tcpIS	ipIR	tcpIS	
Jul	2.74	1.97	1.26	0.68	0.77	0.58	0.19
Ago	1.13	1.74	0.58	0.68	0.23	0.39	0.35
Set	0.70	1.20	0.23	0.33	0.13	0.13	0.23
Out	1.74	1.58	0.48	0.45	0.32	0.16	0.35
Nov	3.90	3.41	1.69	1.24	0.93	0.59	0.23
Dez	1.00	0.19	0.35	0.06	0.29	0.00	0.06

Tabela 4.4 - Média de anomalias ocorridas de julho a dezembro/2004 no servidor S_4.

Roteador S_4							
	Amarelo		Vermelho		Preto		Anomalias
	ifIO	ipIR	ifIO	ipIR	ifIO	ipIR	
Jul	7.45	14.61	3.45	6.77	2.16	4.35	0.39
Ago	1.55	17.23	0.61	8.29	0.42	5.68	0.16
Set	3.83	7.73	1.67	3.70	1.00	2.47	0.10
Out	6.71	8.42	2.68	4.16	1.48	2.58	0.19
Nov	13.14	11.03	6.00	5.45	3.48	3.52	0.27
Dez	7.23	16.81	3.06	7.87	1.87	5.26	0.32

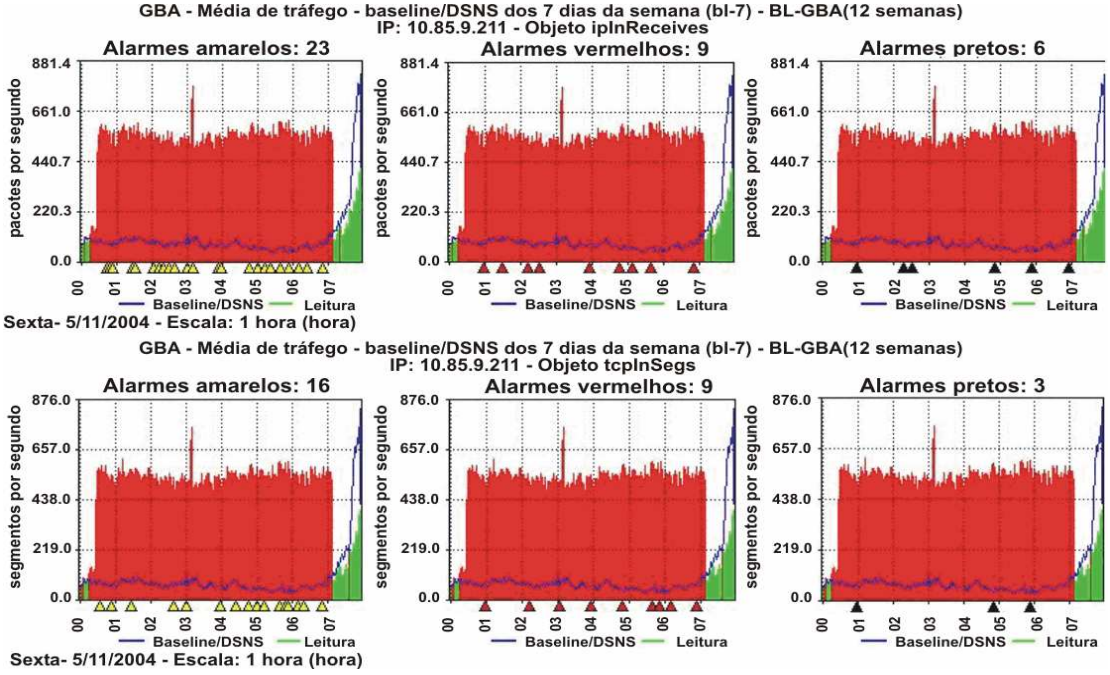


Figura 4.6 - Situação anômala relatada pelo sistema de alarmes para o servidor Proxy S_3.

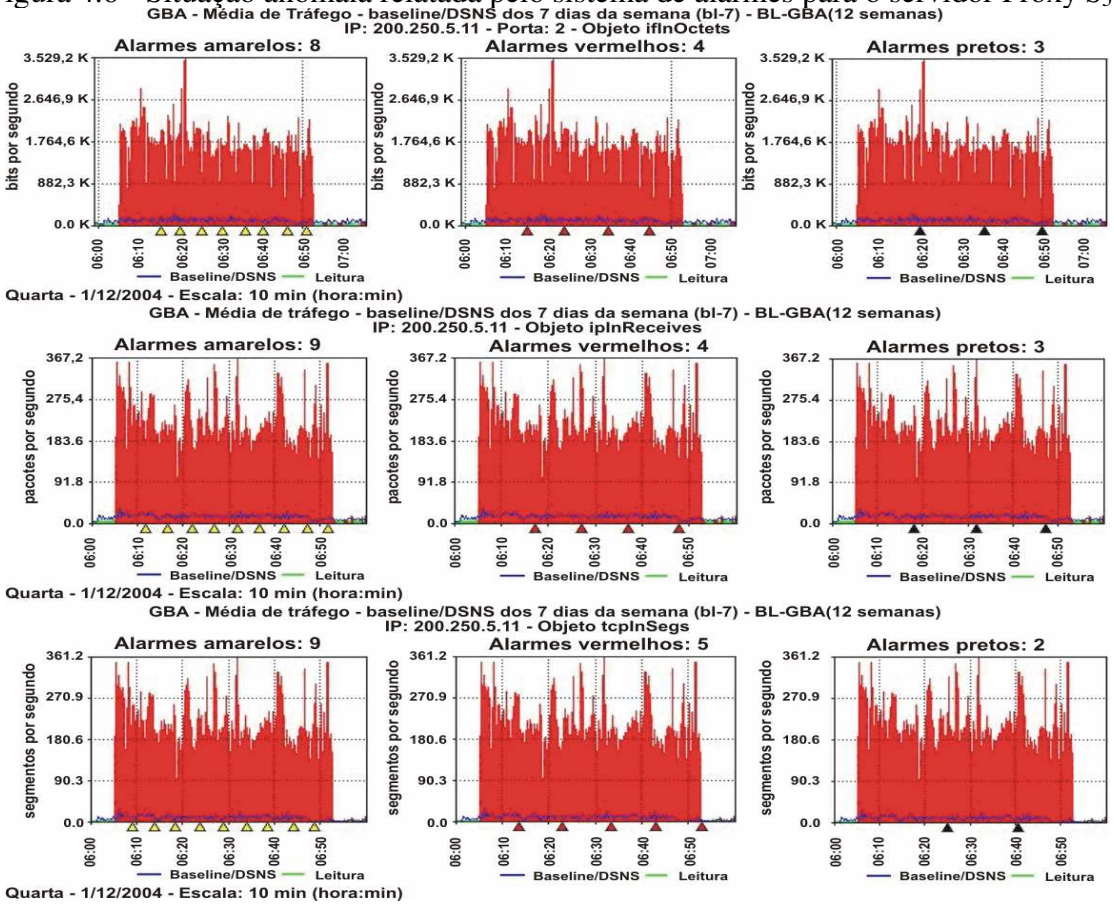


Figura 4.7 - Situação anômala relatada pelo sistema de alarmes para o servidor Web S_2.

Na Figura 4.6 é ilustrado um exemplo ocorrido no servidor Proxy da UEL S_3 em 11/05/2004, onde ocorreram 23 alarmes amarelos, 9 vermelhos e 6 pretos para o objeto *ipInReceives* e, 16 alarmes amarelos, 9 vermelhos e 3 pretos para o objeto *tcpInSegs.* Neste exemplo, pode ser observada uma grande diferença entre o previsto pelo *baseline*/DSNS e o movimento real, o que claramente indica uma anomalia. Estas diferenças são superiores a 100 % e ocorreram simultaneamente em mais de um objeto monitorado. Neste caso, justifica-se plenamente a comunicação através de um alarme para o administrador da rede. O diagnóstico encontrado para esta anomalia detectada no período da madrugada, se deve a uma grande transferência de dados realizada a partir de um segmento administrativo da Universidade durante este período, na verdade isto ocorreu devido a um usuário que permaneceu em seu local de trabalho para realização de um *download*.

Na Figura 4.7 é demonstrado um exemplo de situação no servidor Web S_2 em 12/01/2004, onde podemos verificar a ocorrência de alarmes amarelos, vermelhos e pretos para os objetos *ifInOctets*, *ipInReceives* e *tcpInSegs,* que demonstram uma situação fora do normal esperado pelo *baseline*. Observa-se que a quantidade de alarmes não são iguais para todos os objetos e nível de sensibilidade, porém, eles ocorrem no mesmo intervalo de histerese e indicam simultaneamente a ocorrência de uma anormalidade nos objetos analisados. Esta anomalia ocorreu devido à liberação de um resultado de concurso durante este período.

As tabelas 4.1, 4.2, 4.3 e 4.4, demonstram o resumo médio diário dos últimos 6 meses referentes às quantidades de alarmes amarelos, vermelhos, pretos e anomalias dos servidores analisados neste trabalho. A coluna de anomalias somente foi contabilizada quando ocorreu correlação de alarmes em mais de um objeto analisado. Podemos notar que ocorreu uma baixa quantidade de anomalias, em relação a média de diária de alarmes.

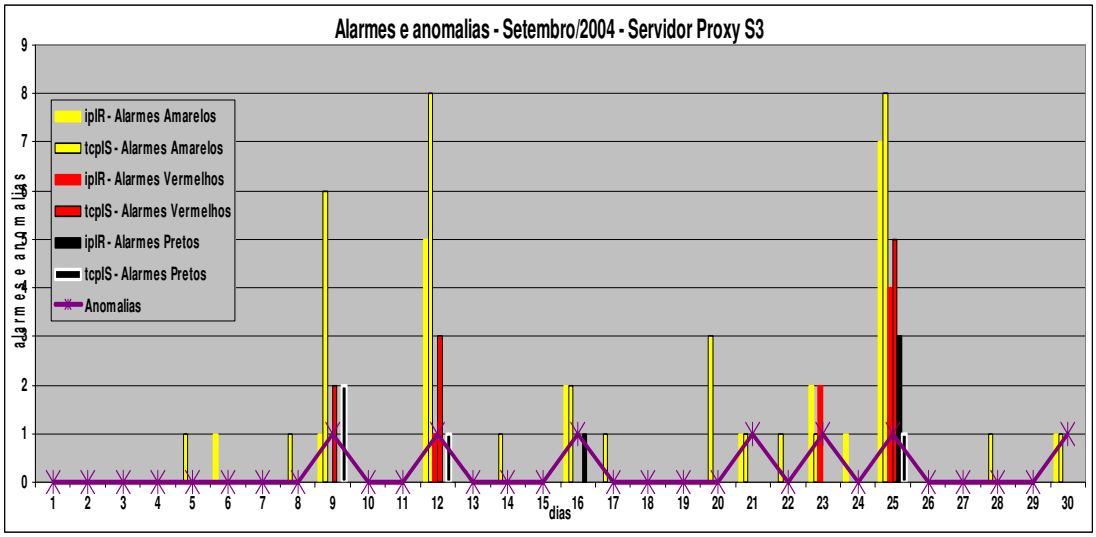


Figura 4.8 – alarmes e anomalias que ocorreram em Setembro de 2004 para o servidor S_3.

A Figura 4.8 ilustra um gráfico em forma de histograma com todas as ocorrências de anomalias, e de alarmes amarelo, vermelho e preto para os objetos *ipInReceives* e *tcpInSegs* durante o mês de setembro de 2004. Nesta figura, podemos verificar a ocorrência de alarmes para diferentes objetos monitorados e somente sete anomalias que ocorreram nos dias 9, 12, 16, 21, 23, 26 e 30 no decorrer do mês.

Constatou-se que os mesmos objetos em diferentes servidores apresentam comportamentos distintos em relação a eventos anômalos. Por exemplo, o objeto *ifInOctets* apresentou grande sensibilidade a variações de tráfego em relação ao DSNS no servidor Firewall S_1, enquanto no Roteador S_4 a sensibilidade notada foi menor. Já com o objeto *ipInReceives* ocorreu o contrário com os mesmos servidores. Em relação aos servidores que estabelecem conexão TCP, com os clientes de rede como o S_2 e o S_3 o comportamento destas variáveis foi semelhante.

Os resultados obtidos para os objetos *ipInReceives* e *tcpInSegs* foram bastante semelhantes entre si nos servidores S_2 e S_3. Conclui-se que este fato ocorreu em função do tipo de serviço prestado por estes servidores, que requerem o estabelecimento de conexões TCP.

5 Conclusão

O objetivo principal deste trabalho foi a construção e implementação de um modelo destinado a caracterização de tráfego de rede, visando à criação de mecanismos automáticos para auxiliar na gerência da rede. Estes mecanismos poderiam ser desde simples sistemas de alarmes baseados nos limiares estabelecidos pelo *baseline*, até sofisticados sistemas de detecção de anomalias que deverão utilizar, além do *baseline*, a correlação de eventos em diferentes objetos e segmentos analisados. Como principais contribuições resultantes deste trabalho, devem ser salientadas:

1. A criação do modelo BLGBA para geração de *baseline*/DSNS;
2. O sistema de alarmes multinível em conjunto com o sistema de correlação de alarmes que formam o sistema de detecção de anomalias ADGBA;
3. Os testes analíticos realizados e os resultados práticos obtidos, que tiveram como objetivo validar a proposta principal deste trabalho;
4. Os dados coletados que se constituem em um banco de informações históricas, com dados reais dos segmentos estudados nos últimos quatro anos e que continuam disponíveis para futuros trabalhos;
5. A pesquisa realizada que demonstra a importância da caracterização de tráfego para a área de gerência de redes.

O modelo BLGBA atingiu seu objetivo no que se refere à caracterização do tráfego de segmentos de rede. A singularidade obtida para cada segmento monitorado é retratada no *baseline* gerado pelo modelo BLGBA. Esta é uma característica fundamental, que pode ser definida como a principal qualidade do resultado gerado, pela aplicação do modelo sobre dados históricos obtidos através de objetos SNMP nestes segmentos.

Os testes analíticos utilizados mostraram, através de seus resultados, o grau de ajuste e a coerência entre o *baseline* gerado pelo modelo BLGBA e o movimento real. A

aplicação destes testes serve como indicador de sua capacidade para avaliar outros modelos similares ao apresentado neste trabalho.

Os testes práticos que proporcionaram inúmeros resultados reais permitiram, através de análises visuais, a confirmação de uma forma bastante simples do ajuste obtido pelo modelo, em relação ao movimento real, além de reafirmarem os resultados dos testes analíticos. Este tipo de análise também é útil em função da riqueza de informações proporcionadas.

Uma característica importante do modelo BLGBA para geração de *baselines* e que adicionou eficiência ao mesmo, foi a possibilidade de serem criados *baselines* independentes para cara dia da semana. A implementação desta funcionalidade se mostrou acertada, considerando a preservação das características singulares de cada dia da semana.

Em relação aos segmentos de rede que apresentam um grande volume de tráfego agregado, provenientes de muitas estações ligadas à rede, como por exemplo, o S_4 que liga a rede da UEL ao seu roteador, e o S_6 que agrega o tráfego de toda a rede da UNICAMP para acesso à Internet, o modelo BLGBA se ajusta de forma bastante satisfatória, apresentando alto grau de acerto no previsto pelo *baseline*. Uma característica fundamental que se destaca neste tipo de segmento e que de forma direta acaba por explicar a eficiência do modelo BLGBA, se relaciona a forte tendência do estabelecimento de padrões cíclicos e com dependência temporal, que nos casos estudados, estavam relacionadas ao horário do expediente das Universidades. Como o BLGBA funciona fundamentalmente baseado na análise estatística de dados históricos, a existência de padrões facilita seu bom desempenho neste caso.

Em segmentos com pequeno volume de tráfego agregado, que normalmente estão relacionados à pequena quantidade de estações ligadas à rede, como o segmento estudado o S_5, o modelo BLGBA apresentou resultados satisfatórios. Exceto em algumas situações onde a caracterização não apresentou bons resultados. Isto se deve a extrema volatilidade do tráfego que torna difícil o estabelecimento de um padrão. Mais testes e estudos devem ser realizados com o modelo BLGBA, em segmentos de rede que apresentam poucas estações ou mesmo pouco tráfego agregado, com intuito de realizar ajustes para melhor caracterização do tráfego neste tipo de segmentos, caso necessário.

A construção do sistema de alarmes foi realizada como objetivo secundário deste trabalho, contudo este sistema pôde efetivamente demonstrar na prática vantagens na utilização do *baseline*, que ocorreram justamente pelo estabelecimento de limiares adaptáveis de forma temporal às características de cada segmento. O sistema de alarmes multinível foi uma conseqüência natural da criação do *baseline*, tendo em vista a redução da quantidade de alarmes gerados e da necessidade de somente avisar ao administrador, quando um evento mereça sua atenção. Os parâmetros que foram propostos neste trabalho para os três níveis de alarmes (amarelo, vermelho e preto), foram designados segundo um estudo empírico sobre as necessidades estabelecidas no centro de operações da rede da Universidade Estadual de Londrina. Não obstante, eles podem ser alterados em função de outras necessidades que exijam controles diferentes dos que foram pré-estabelecidos pela constante δ.

O sistema ADGBA apresentado neste trabalho, na verdade é uma proposta, na qual se faz a correlação dos alarmes gerados para mais de um objeto SNMP ao mesmo tempo e no mesmo intervalo de histerese, com objetivo de detectar anomalias. Esta proposta de correlação pode ser definida como um avanço significativo aos sistemas de alarmes, disponíveis atualmente nas ferramentas tradicionais de gerência, pois na verdade elas somente oferecem alarmes com limites constantes que não se adaptam ao movimento do tráfego ao longo do dia. Como trabalho futuro mais testes, estudos e análises da aplicabilidade dos resultados devem ser realizados com objetivo de se consolidar esta proposta.

A detecção de anomalias ainda se apresenta como uma questão não totalmente resolvida, principalmente devido a problemas relacionados a detecção e a localização da anomalia. As dificuldades em se determinar o que venha ser uma anomalia em meio ao tráfego de rede, que naturalmente é composto por rajadas ocasionais, tendem a confundir e dificultar sua descoberta. Após a descoberta surge um outro problema que é a localização da mesma.

Um outro trabalho futuro em decorrência natural desta tese, está relacionado justamente com a continuidade e o aprofundamento do sistema ADGBA. Nesta tese, somente foi utilizado como regra de correlação a ocorrência do alarme em mais de um

objeto. Portanto, o que se deve buscar é o aprofundamento do modelo de correlação apresentado, prevendo a utilização combinada de particularidades pertencentes a cada objeto, com a finalidade de aprimorar as regras de correlação. Estudos devem ser realizados nos objetos pertencentes aos agentes SNMP, com objetivo de identificar características que possam ser combinadas e correlacionadas em conjunto com seus respectivos *baselines*/DSNS para identificar anomalias de rede.

Atualmente, a gerência tradicional realizada em sua maioria através de ferramentas que não agregam valor na descoberta e solução de problemas, justamente por não disponibilizar ao administrador o *baseline* dos segmentos, deve ser melhorada no sentido de proporcionar condições para automação de tarefas que são fundamentais na administração de redes. Inúmeros problemas podem acontecer em uma rede, desde a falhas em equipamentos, aumento do tráfego desencadeando congestionamentos, redução de capacidade em segmentos, ataques ou mesmo invasões. Nos *backbones* atuais, que são compostos por vários segmentos e com altas taxas, a demora para reação em relação a um problema pode significar sérias implicações. A utilização do *baseline*/DSNS em todos os segmentos do *backbone* de uma rede deve ser uma prática obrigatória, tendo em vista os benefícios que podem ser obtidos para administração. Sua utilização para múltiplos objetos SNMP em conjunto com técnicas de mapeamento da topologia da rede (Breitbart *et at.,* 2004), tornará possível avaliar a origem de um problema e mesmo o seu reflexo nos demais segmentos da rede. Este é um ponto fundamental que deve ser considerado para o aprimoramento do sistema ADGBA, pois, de forma geral, uma anomalia pode ser um reflexo de um problema que está ocorrendo em outro segmento da rede (Lakhina *et al.,* 2004) (Roughan *et al.,* 2004).

A aplicabilidade do *baseline*/DSNS, gerado pelo modelo BLGBA, vai além das aplicações apresentadas aqui, ou seja, ela não se restringe somente ao monitoramento de número de bits que entram ou saem, número de datagramas recebidos ou ainda número de segmentos TCP recebidos por uma interface ou segmento monitorado. Sua aplicabilidade pode se estender a outras situações ou problemas que tenham seu comportamento retratados através de contadores SNMP. Foram realizados outros testes além dos apresentados nesta tese, onde obteu-se bons resultados, semelhantes aos apresentados no capítulo 3 deste trabalho. Nestes testes foram utilizados objetos SNMP pertencentes a uma MIB residente

em um servidor de banco de dados, da qual foram utilizados objetos que destinavam a retratar quantidade de memória utilizada pelo servidor, quantidade de processos sendo executados, quantidade de transações por segundo e a temperatura do servidor.

A utilização de agentes do tipo *Remote Network Monitoring* (RMON) e RMON2 irá possibilitar a geração de *baselines* mais significativos, tendo em vista o valor agregado disponível nos objetos SNMP residentes nas MIBs pertencentes ao RMON e ao RMON2 (RFC 2021, 1997) (RFC 2074, 1997). Além de *baselines* sobre características básicas do tráfego como já apresentado neste trabalho outros *baselines* poderão ser criados com informações sobre protocolos, serviços e aplicações utilizadas que tornarão a gerência mais completa e rica com informações sobre os usuários da rede.

6 Bibliografia

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

Barford Paul, 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.

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

Breitbart, Y., Garofalakis, M., Jai, B., Matin, C., Rastogi, R., Silberschatz, A.., **"Topology Discovery in Heterogeneous IP Networks: the *NetInventory* System"**. IEEE Transaction on Networking, v. 12, n. 3, p. 401-414, jun 2004.

Bussab, Wilton O.; Morettin Pedro A.; **Estatística Básica**, Editora Saraiva, 5a edição. 2003 ISBN 85-02-03497-9.

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.

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. Grossglauser

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.

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.

GBA - **Gerenciamento de *Backbone* Automatizado**, disponível via Web em http://proenca.uel.br/gba/ (21/02/2005).

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.

Ho, L.L.; Cavuto, D.J.; Papavassiliou, S.; Zawadzki, A.G.; **Adaptive and automated detection of service anomalies in transaction-oriented WANs: network analysis, algorithms, implementation, and deployment** Selected Areas in Communications, IEEE Journal on, Volume: 18, Issue: 5, May 2000; Pages: 744 – 757.

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

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

Jagerman, David L., Melamed, Benjamin, Willinger, Walter. **Stochastic Modeling of Traffic Processes**. In J. Dshalalow, editor, Frontiers in Queueing: Models, Methods and Problems. CRC Press, 1996.

Jain Raj, Hassan Mahbub; **High Performance TCP/IP Networking,** Concepts, Issues and Solutions**,** Pearson Prentice Hall, 2004, ISBN 0-13-06434-2.

Jain, Raj **The Art of Computer Systems Performance Analysis,** Techniques for experimental design, measurement, simulation and modeling**,** Willey Computing, 1991 ISBN 0-471-50336-3.

Jiao, J.; Naqvi, S.; Raz, D.; Sugla, B.*;* **Toward efficient monitoring;** Selected Areas in Communications, IEEE Journal on, Volume: 18, Issue: 5, May 2000 Pages: 723 – 732.

Jun Jiang; Papavassiliou, S.*;* **A network fault diagnostic approach based on a statistical traffic normality prediction algorithm;** Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE, Volume: 5, 1-5 Dec. 2003, Pages: 2918 – 2922.

Krishnamurthy Balachander, 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.

Lakhina A., Crovella M., Diot C.; **Characterization of Network-Wide Anomalies in Traffic Flows,** IMC'04 **-** Sicily, Italy, October 2004, ISBN: 1-58113-821-0.

Lakhina A., Crovella M., Diot C.; **Diagnosing network-wide traffic anomalies**, SIGCOMM'04 - Portland, Oregon, USA, 2004, ISBN:1-58113-862-8.

Leland, Willie E. , Taqqu, Murad S., Willinger, Walter, Wilson, Daniel V. **On the Self-Similar Nature of Ethernet Traffic (extended version).** IEEE/ACM Trans. On Networking, v.2, n.1, pp. 1-15, Feb. 1994.

Mandelbrot B. **Self-Similar Error Clusters in Communication Systems and the Concept of Conditional Stationarity**. Communications, IEEE Transactions on, Volume: 13 , Issue: 1 , Mar 1965, ISSN: 0096-2244.

Maxion Roy, Feather Frank, Siewiorek Dan, **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.

MRTG - **The Multi Router Traffic Grapher**, disponivel via Web no endereço http://www.mrtg.com/ (21/02/2005).

NET-SNMP – Net-SNMP disponivel via WEB no endereço: http://www.net-snmp.org/ em 01/03/2005.

Northcutt, Stephen, NOVAK Judy. **Network Intrusion Detection**, Third Edition, New Riders, 2002.

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.

Proença, Mario Lemes, Jr.; Sakuray Fabio; Mendes, Leonardo; "**Uma Experiência de Gerenciamento de Rede com *Backbone* ATM através da Ferramenta GBA**", artigo publicado no , XIX Simpósio Brasileiro de Telecomunicações – SBrT 2001, Fortaleza de 03-06 de Setembro 2001.

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

RFC 2021 - INTERNET ENGINEERING TASK FORCE (IETF). **Remote Network Monitoring Management Information Base Version 2**, RFC 2021, jan.1997.

RFC 2074 - INTERNET ENGINEERING TASK FORCE (IETF). **Remote Network Monitoring MIB Protocol Identifiers**, RFC 2074, jan.1997.

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

Roughan Matthew, Tim Griffin, Z. Morley Mao, Albert Greenberg, Brian Freeman. **IP forwarding anomalies and improving their detection using multiple data sources**. Proceedings of the ACM SIGCOMM workshop on Network troubleshooting: research, theory and operations practice meet malfunctioning reality. Portland, Oregon, USA, 2004, ISBN 1-58113-942-9.

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.

Sekar R., Gupta A., Frullo J., Shanbhag T., Tiwari A., Yang H., Zhou S. **Specification-based Anomaly Detection: A New Approach for Detecting Network Intrusions**; CCS´02 Proceedings of the 9th ACM conference on Computer and communications security, Washington 2002, ISBN:1-58113-612-9.

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.

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

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

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.

Wu, N., Zhang, J. **Factor analysis based anomaly detection** Information Assurance Workshop, 2003. IEEE Systems, Man and Cybernetics Society, June 2003, p. 108-115.

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.

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.