

# Amostragem Multi-adaptativa de Tráfego de Rede

João Marco C. Silva, Solange Rito Lima  
Departamento de Informática  
Centro Algoritmi, Universidade do Minho  
Braga, Portugal  
Email: jotamarco@gmail.com, solange@di.uminho.pt

**Resumo**—A monitorização eficiente de redes e serviços tem um papel crucial no suporte à gestão e engenharia das redes de comunicações atuais. Obter uma visão realista do estado da rede sem, no entanto, interferir na sua operação normal é essencial no suporte à decisão dos operadores de redes e dos sistemas auto-configuráveis.

Neste sentido, as técnicas de amostragem de tráfego procuram reduzir o impacto da monitorização no funcionamento da rede, mantendo a acurácia na estimação de parâmetros referentes ao seu comportamento estatístico. Embora seja observada uma evolução das atuais técnicas na correta estimação do desempenho da rede, o *overhead* associado ao volume de dados envolvidos nos processos de amostragem de tráfego ainda é considerável. Neste contexto, este trabalho propõe uma nova técnica de amostragem multi-adaptativa de tráfego, baseada em predição linear, que tem por objetivo reduzir significativamente o volume total de dados envolvidos na monitorização do estado da rede, sem comprometer a representatividade das amostras na estimação dos parâmetros em análise.

A avaliação desta proposta, com base em coletas de tráfego real, demonstra que é possível reduzir significativamente o volume de dados analisado e mesmo assim alcançar, em vários parâmetros de caracterização, resultados mais precisos que as técnicas classicamente utilizadas.

## I. INTRODUÇÃO

A necessidade de gerir redes de computadores de elevada dimensão e heterogeneidade tem exigido o estudo e a proposta de técnicas eficazes e de custos aceitáveis. Atividades como a gestão e auditoria de serviços, a engenharia de tráfego e a autoconfiguração de redes, dependem de uma análise realista do estado da rede e dos seus serviços. Embora com requisitos muito próprios, estas atividades têm em comum a necessidade de monitorização do tráfego de rede.

A tarefa inicial da monitorização de rede é a sua medição, esta por sua vez decomposta nos processos de captura e filtragem de tráfego. A primeira consiste em coletar o tráfego a ser analisado num ponto de medição (*Measurement Point*), que pode ser interno ou externo ao nó onde ocorre a captura. Já a filtragem consiste em selecionar um subconjunto de pacotes com propriedades comuns, que podem derivar diretamente do conteúdo do pacote ou depender do tratamento dado ao pacote pelo nó de comutação [1].

Os procedimentos de medição podem adotar metodologias ativas ou metodologias passivas. As metodologias ativas recorrem ao uso de tráfego intrusivo através da injeção de pacotes de prova (*probes*) na rede com propósito de medição. Embora versáteis, estes métodos impõem cuidados com o *overhead*

do tráfego de prova, já que pode interferir no funcionamento normal da rede.

As metodologias passivas de medição adotam técnicas não intrusivas, considerando apenas o tráfego real da rede em estudo. Estas técnicas geralmente utilizam dispositivos específicos ou mecanismos embebidos nos dispositivos da rede. Uma das dificuldades associadas ao seu uso é o volume de tráfego envolvido, resultando numa necessidade elevada de recursos para o processamento, armazenamento e transmissão [1].

Uma forma de atenuar os impactos provocados pela medição é através de amostragem (*Sampling*) do tráfego. O objetivo principal da amostragem é selecionar um subconjunto dos pacotes de tráfego que permita efetuar estimativas, com compatíveis graus de exatidão, sobre todo o tráfego da rede, evitando processá-lo completamente [2].

As várias técnicas de amostragem distinguem-se de acordo com o processo adotado de seleção dos pacotes que compõem a amostra de tráfego. Na *amostragem sistemática*, a seleção respeita uma função determinística baseada na posição (*count-based*), no tempo de chegada do pacote ao ponto de medição (*time-based*) ou no seu conteúdo (*content-based*). Contudo, o padrão de comportamento do tráfego associado ao determinismo da abordagem traz riscos de sobrecarga do ponto de medição e produção de amostras tendenciosas.

As técnicas de *amostragem aleatória* selecionam pacotes de maneira a evitar os riscos de amostras tendenciosas, inerentes aos processos sistemáticos. Para isso a seleção dos pacotes é definida por uma função aleatória [1]. Embora solucione o problema de amostras tendenciosas, esta técnica apresenta-se inadequada para a estimativa de métricas que necessitam de medição multiponto (*edge-to-edge*), pois, como não há correlação entre os processos em cada nó, não há garantias de que as diferentes amostras sejam formadas pelos mesmos pacotes [3].

Na *amostragem adaptativa* o processo de seleção dos pacotes considera o valor de uma variável de interesse observada durante a medição. Desta maneira, dota-se o processo de flexibilidade para mudar o plano de seleção dos pacotes durante o curso da monitorização.

As atuais técnicas de amostragem adaptativa têm apresentado grandes evoluções na capacidade de representar o comportamento estatístico do tráfego de rede. Contudo, estas técnicas não objetivam a redução do impacto do *overhead* associado ao volume de dados envolvidos nos processos de medição. Este aspecto interfere diretamente nos custos da

monitorização de rede.

Neste contexto, este trabalho apresenta uma nova técnica de amostragem multi-adaptativa de tráfego, baseada em predição linear, que tem por objetivo reduzir o volume total de dados envolvidos na monitorização do estado da rede. Com esta técnica pretende-se reduzir o impacto do *overhead* associado aos processos de medição, garantindo, no entanto, que o tráfego amostrado representa as características estatísticas do tráfego total da rede em análise. Para tal, o processo de captura de tráfego é balanceado em função do nível de atividade da rede, de forma a reduzir o impacto da medição nos instantes de maior atividade. O caráter multi-adaptativo é alcançado considerando-se o tamanho da amostra, além do intervalo entre amostras, um fator adaptativo.

O restante do documento apresenta na Secção II, os trabalhos existentes nesta área e a motivação para a proposta de uma nova técnica. Na Secção III são apresentados alguns conceitos importantes utilizados ao longo do artigo e é descrita a técnica desenvolvida. Na Secção IV são descritos o cenário e os parâmetros usados nos testes de desempenho da técnica. Na Secção V discutem-se os resultados obtidos nos testes efetuados. Por fim, na Secção VI, apresentam-se as conclusões finais e as propostas de trabalho futuro.

## II. TRABALHOS RELACIONADOS

As técnicas de medição adaptativas geralmente são desenvolvidas para a estimação de um parâmetro específico, como a perda [4] [5] ou o atraso e sua variação [6]. Um conjunto maior de propostas são destinadas à caracterização de tráfego e monitorização do cumprimento de SLAs [7] [8] [9] [15].

As propostas mais generalistas costumam basear-se em lógica *Fuzzy* e em Predição Linear. Nas técnicas de amostragem adaptativa baseadas em lógica *Fuzzy* [10] [11], um controlador ajusta a taxa de amostragem baseado em experiências de situações passadas semelhantes, determinando a ação mais adequada para uma determinada condição do tráfego [12]. Esta abordagem exige a existência de uma base de dados de longo prazo, que pode ser interna ou externa ao ponto de medição, com possível impacto no desempenho da técnica ou do nó que implementa o ponto de medição.

As técnicas baseadas em Predição Linear [13] [14], tentam prever o comportamento futuro da rede com base num parâmetro observado em amostras passadas. Assim, se uma previsão se confirma, a taxa de amostragem pode ser reduzida, enquanto que previsões incorretas indicam uma mudança no comportamento da rede e requerem um aumento na taxa de amostragem para determinar o novo padrão [12].

Nesta abordagem o número de amostras consideradas na predição é fixo e estas podem ser armazenadas por um curto intervalo de tempo no próprio ponto de medição, diminuindo assim a necessidade de recursos adicionais. No entanto, a maior frequência de amostragem, provocada pelo aumento na atividade da rede, pode exigir mais recursos do ponto de medição, exatamente nos momentos mais críticos de funcionamento da rede.

A técnica desenvolvida neste trabalho procura diminuir a necessidade de recursos relacionados com o processamento, armazenamento e transmissão dos pacotes capturados, nos períodos de maior atividade da rede, mantendo a acurácia das estimativas efetuadas sobre o seu comportamento estatístico. Para isso, dota-se a técnica da capacidade de alterar dinamicamente o tamanho da amostra de maneira a reduzir a sobrecarga provocada pela maior frequência de amostragem.

## III. DESCRIÇÃO DA PROPOSTA

### A. Conceitos

As técnicas de medição de redes baseadas em amostras partilham um conjunto de conceitos que nem sempre são claros e, por vezes, apresentam-se de forma ambígua. Para evitar tais inconsistências, este trabalho utiliza os termos mais comuns de acordo com as seguintes definições:

**Amostras:** Subconjuntos dos pacotes da rede analisada que são selecionados pelo ponto de medição e considerados na estimação dos parâmetros escolhidos. Estes pacotes são também utilizados pelos algoritmos de medição adaptativa como entrada para estimação do *parâmetro de referência*.

**Tamanho da amostra:** Intervalo de tempo no qual todos os pacotes que chegam ao ponto de medição são selecionados e capturados para compor uma unidade de *amostra*.

**Intervalo entre amostras:** Período de tempo em que todos os pacotes que chegam ao ponto de medição são ignorados para fins de medição. Neste intervalo, todo o comportamento da rede é desconsiderado na estimação dos parâmetros escolhidos e na definição do regime de amostragem.

**Parâmetro de referência:** Valor observado em cada amostra que serve como entrada para a decisão sobre a mudança no regime de amostragem adotado.

Estas definições são representadas na Figura 1.



Figura 1. Representação conceitual.

### B. Descrição da Técnica de Amostragem Multi-adaptativa

A técnica de medição por amostragem multi-adaptativa proposta neste trabalho baseia-se nas técnicas adaptativas apresentadas em [12] [15], adaptando-as e estendendo as suas propriedades de forma a possibilitar reatividade não só nos intervalos entre amostras, mas também nos tamanhos das amostras. Desta forma, desenvolve-se uma técnica de medição que atende os seguintes requisitos:

- 1) o caráter adaptativo da técnica é baseado em predição linear. Esta decisão é justificada pela sua simplicidade de implementação e baixo consumo de recursos;
- 2) à medida que se nota um aumento na atividade da rede, a frequência de amostragem precisa ser aumentada para que o novo padrão de comportamento seja identificado.

Este aumento é alcançado com base na diminuição do intervalo entre amostras;

- 3) como o aumento na frequência de amostragem implica um maior consumo de recursos (processamento e armazenamento), o tamanho das amostras deve sofrer alterações de maneira a atenuar o *overhead* mantendo a representatividade das amostras na estimação dos parâmetros analisados.

Para atender o requisito (1), utiliza-se a função de predição linear apresentada em [12]. No entanto, são considerados, para o cálculo do parâmetro de referência, apenas os pacotes pertencentes às amostras anteriormente coletadas, ao contrário do que é apresentado em [12], que considera todo o tráfego da rede estudada, independente se os pacotes pertencem ou não a uma amostra. A função preditiva adaptada é apresentada na Equação 1.

O requisito (2) é atendido integralmente pela técnica fornecida por [12]. O conjunto de cálculos e regras que determinam o fator de alteração no intervalo entre amostras são apresentados e discutidos na Equação 2 e na Tabela I.

Como nenhuma das técnicas até então publicadas, apresenta características que atendam o requisito (3), este trabalho adapta o índice de variação do intervalo entre amostras apresentado em [15] para permitir alterar o tamanho das amostras de acordo com nível de atividade da rede analisada. A variação no tamanho da amostra respeita os critérios apresentados na Tabela II.

A técnica apresentada neste trabalho considera as últimas  $N$  amostras coletadas para estimar o valor futuro do parâmetro de referência e assim determinar o próximo intervalo entre amostras e o tamanho da próxima amostra. Desta forma, para um amostrador de ordem  $N$ , o valor previsto  $Xp$  para o parâmetro de referência na próxima amostra coletada é definido pela Equação 1.

$$Xp = X[N] + \frac{\Delta T_{atual}}{N-1} \sum_{i=1}^{N-1} \left( \left| \frac{X[i+1] - X[i]}{\Delta T[i]} \right| \right) \quad (1)$$

Para a estimação do valor de  $Xp$  um vetor  $X$  armazena os valores do parâmetro de referência das últimas  $N$  amostras, onde  $X[N]$  se refere à amostra mais recente. Um segundo vetor  $T$  armazena os intervalos entre amostras, onde cada  $\Delta T$  corresponde ao tempo entre o fim da amostra  $X[i]$  e o início da amostra  $X[i+1]$ . Desta forma,  $\Delta T_{atual}$  é o intervalo entre as amostras  $X[N]$  e  $X[N-1]$ .

Assim que uma nova amostra  $S$  é coletada, o valor do parâmetro de referência de  $S$  é comparado com valor previsto  $Xp$ . Baseado na taxa de mudança, comparada entre  $Xp - X[N]$  e  $S - X[N]$ , é aplicado um conjunto de regras que definem o  $\Delta T_{seguinte}$ , que é usado para agendar o início da próxima amostra. Esta comparação define um fator  $m$  dado pela Equação 2.

$$m = \left| \frac{Xp - X[N]}{S - X[N]} \right| \quad (2)$$

A taxa de mudança fornecida pela Equação 2 apresenta resultado próximo de 1 quando o valor previsto  $Xp$  for próximo do valor atual  $S$ . O alcance dos valores que satisfazem esta condição é definido como:

$$m_{min} < 1 < m_{max}$$

Se  $m$  for menor que  $m_{min}$ , o parâmetro de referência está a mudar mais rapidamente que o previsto. Tal comportamento indica mais atividade que o esperado na rede. Desta forma, o intervalo entre amostras deve ser diminuído para alcançar valores mais precisos, que servem como base para futuras previsões.

Inversamente, se  $m$  for maior que  $m_{max}$ , o valor do parâmetro de referência está a mudar mais lentamente que o previsto, então o intervalo entre amostras pode ser aumentado, reduzindo o impacto da medição.

Se o valor de  $S$  não mudar desde a última amostra  $X[N]$ ,  $m$  assume um valor indefinido, o que indica que a rede está ociosa. Neste caso, o intervalo entre amostras é aumentado exponencialmente.

A Tabela I lista as regras usadas para gerar o  $\Delta T_{seguinte}$  a partir do valor atual de  $m$ .

Tabela I  
REGRAS PARA DEFINIR O PRÓXIMO INTERVALO ENTRE AMOSTRAS [12]

$m$ atual	$\Delta T_{seguinte}$
$m < m_{min}$	$\Delta T_{seguinte} = m * \Delta T_{atual}$
$m_{min} \leq m \leq m_{max}$	$\Delta T_{seguinte} = \Delta T_{atual}$
$m_{max} < m$	$\Delta T_{seguinte} = \Delta T_{atual} + 1seg$
$m_{indefinido}$	$\Delta T_{seguinte} = 2 * \Delta T_{atual}$

Os valores de  $m_{min}$  e  $m_{max}$  são atribuídos experimentalmente e, segundo [12], são definidos como  $m_{min} = 0.9$  e  $m_{max} = 1.1$  por apresentar bom desempenho para um grande número de tipos de tráfego.

Uma restrição adicional pode ser usada para limitar as possibilidades dos valores de  $\Delta T_{seguinte}$  de forma que este não cresça ilimitadamente, mantendo assim um regime mínimo de amostragem que garanta dados para novas previsões. Por outro lado, limita-se também a frequência máxima, de maneira que o intervalo entre amostras não tenda a zero, o que se aproximaria de uma captura total do tráfego. Estes limitadores, atribuídos experimentalmente de forma a ajustar as características de multi-adaptação, definem o intervalo mínimo entre amostras de 0.1s e o intervalo máximo entre amostras de 8s.

Para adaptação nos tamanhos das amostras considera-se também a indicação de atividade na rede fornecida pelo fator  $m$ . No entanto, estas alterações são lineares com ordem de mudança fixo igual a  $1/k$ . Este parâmetro é adotado para garantir que uma mudança brusca no comportamento da rede não produza amostras de tamanhos muito diferentes.

O parâmetro  $k$  é apresentado em [15], onde é atribuído experimentalmente e tem como objetivo alterar o intervalo entre amostras baseado num conjunto de estatísticas comparadas entre duas amostras vizinhas. Na presente técnica, o parâmetro foi adotado como fator de alteração no tamanho da amostra baseado num número  $N$  de amostras previamente coletadas.

A Tabela II apresenta as regras usadas para definir o tamanho da próxima amostra, onde  $\Delta S_{atual}$  representa o tamanho de amostra atual e  $\Delta S_{seguinte}$  representa o tamanho da próxima amostra a ser coletada.

Tabela II  
REGRAS PARA DEFINIR O TAMANHO DA PRÓXIMA AMOSTRA

Tamanho atual	Próximo tamanho
$m < m_{min}$	$\Delta S_{seguinte} = \Delta S_{atual} - (\Delta S_{atual}/k), k = 4$
$m_{min} \leq m \leq m_{max}$	$\Delta S_{seguinte} = \Delta S_{atual}$
$m_{max} < m$	$\Delta S_{seguinte} = \Delta S_{atual} + (\Delta S_{atual}/k), k = 10$
$m_{indefinido}$	$\Delta S_{seguinte} = \Delta S_{atual} + (\Delta S_{atual}/k), k = 10$

De acordo com a Tabela II, para momentos de menor atividade na rede, o tamanho da amostra é aumentado num fator de  $1/k$  com  $k = 10$ , sendo coletadas mais informações sobre a rede em períodos menos críticos de seu funcionamento. Já nos momentos de maior atividade, o tamanho da amostra é reduzido num fator de  $1/k$  com  $k = 4$ . Este maior fator de redução é compensado pela maior frequência no processo de amostragem, reduzindo assim o risco de *overhead* no ponto de medição, enquanto se mantém a representatividade das amostras.

Com o objetivo de limitar os tamanhos das amostras são incluídas restrições. Tais restrições evitam amostras demasiadamente pequenas, o que dificulta a estimação de parâmetros estatisticamente, assim como excessivamente grandes, que novamente se aproximaria de uma captura total do tráfego. Estes limitadores, atribuídos experimentalmente, definem que o tamanho mínimo da amostra é 0.1s e o tamanho máximo da amostras é igual a 2s.

#### IV. CENÁRIO DE TESTES

##### A. Metodologia

Os testes da técnica proposta são efetuados através do uso de *traces* de tráfegos capturados durante a realização da conferência SIGCOMM 2008.

A captura é composta por tráfego gerado pelos participantes da conferência conectados através dos diversos APs - *Access Points*, padrão IEEE 802.11a, disponibilizados no local. Este tráfego é bastante heterogêneo e abrange acessos a páginas da internet, serviços de e-mail, clientes de mensagens instantâneas, *upload* e *download* de ficheiros etc.

Para disponibilizá-lo, foi utilizada uma ferramenta de captura conectada a uma porta *Ethernet* do concentrador de todos os APs. Desta forma, todos os pacotes enviados a partir do NAT - *Network Address Translation* para os APs, assim como os pacotes enviados pelos APs com destino ao NAT foram capturados. O anonimato dos utilizadores da conexão foi garantido através do uso de uma versão modificada da

ferramenta *tcpkpub*. O *trace* é público e disponibilizado em [16].

A metodologia dos testes consiste na comparação estatística do tráfego resultante do processo de medição baseado na técnica de amostragem desenvolvida em comparação com o tráfego total (não amostrado). Além desta análise, o desempenho da técnica multi-adaptativa é também comparado ao desempenho de duas outras técnicas, nomeadamente a técnica sistemática baseada no tempo [1] e a técnica LP - *Linear Prediction* [12], usada como referência.

Em todos os testes efetuados, o parâmetro de referência adotado é o débito médio observado em cada amostra. Desta forma, para cada decisão sobre o próximo tamanho de amostra e o intervalo até o início da amostra seguinte, observa-se, em cada amostra, a razão entre o volume total de *bits* pelo respectivo tamanho nas últimas  $N$  amostras, de acordo com a *ordem* definida.

Salienta-se ainda que a metodologia de testes *offline* adotada não compromete o realismo da amostra final capturada face ao que ocorreria num ambiente de amostragem em tempo real. Na medida em que, mesmo utilizando tráfego controlado, nenhuma das técnicas avaliadas possui conhecimento prévio sobre as características desse tráfego, o processo de amostragem é configurado dinamicamente no decurso da amostragem à semelhança do que ocorreria em ambiente *online*.

##### B. Parâmetros estatísticos

Os parâmetros estatísticos comparativos têm por objetivo determinar o nível de equivalência entre o comportamento do tráfego total e o tráfego resultante dos processos de amostragem.

Embora os parâmetros estatísticos sejam os mesmos usados na maiorias das pesquisas na área, este trabalho diferencia-se em compará-los ao tráfego total, enquanto a abordagem comum é comparar a técnica adaptativa ao desempenho das técnicas sistemáticas.

Os parâmetros estatísticos utilizados na comparação são:

1) *Overhead*: O objetivo principal da técnica proposta é manter a representatividade do tráfego amostrado, em relação ao tráfego total, reduzindo o *overhead* associado ao volume de pacotes processados, armazenados e transmitidos pelo ponto de medição. Para isso, comparam-se as técnicas através dos seguintes parâmetros:

*Número de Pacotes*: Corresponde ao número total de pacotes capturados durante o processo de amostragem para cada uma das técnicas usadas nos testes.

*Volume de Dados*: Soma do volume, em *Mbytes*, de todos os pacotes coletados em cada uma das técnicas de amostragem utilizadas nos testes. Para esta estimativa, usa-se o campo referente ao tamanho total do pacote no cabeçalho IP.

*Número de Amostras*: Numero total de amostras capturadas durante o processo de amostragem. Este parâmetro é relevante no sentido em que, para cada nova amostra, são necessários cálculos para definir o novo tamanho da amostra e intervalo de tempo até o seu início. Desta maneira, quanto maior o número

de amostras, maior o número de cálculos efetuados pelo ponto de medição.

2) *Débito Estimado*: O parâmetro básico de observação do débito estimado na rede monitorizada é o próprio parâmetro de referência definido, ou seja, o débito médio observado em cada amostra. Através dele é possível obter informações em tempo real sobre o comportamento da rede através de gráficos onde cada instância do parâmetro representa o *débito instantâneo* observado.

Outros parâmetros que permitem o conhecimento do comportamento do tráfego são a *explosividade* e o *coeficiente de variação* (CV). Estas métricas pertencem à estatística descritiva e medem o nível de variabilidade de uma série. Na monitorização de tráfego de rede, elas ajudam a identificar e caracterizar a impulsividade do tráfego.

A explosividade é obtida pela razão entre o valor máximo dos débitos instantâneos pela média do mesmo parâmetro num intervalo de observação. Já o coeficiente de variação é obtido pelo quociente da divisão do desvio padrão pela média dos débitos instantâneos.

Calcula-se também a *correlação* entre o tráfego amostrado e o tráfego total. Como para o cálculo da correlação é preciso que as duas séries tenham tamanhos iguais, nestes testes considera-se apenas os intervalos de um segundo em que seja capturado, pelo menos, um pacote. Esta métrica indica, estatisticamente, o quanto o tráfego amostrado se assemelha ao tráfego total. O coeficiente é obtido através do método de Pearson [17], onde:

$0.7 > \text{coeficiente} \geq 1$ , indica forte correlação

$0.3 > \text{coeficiente} \geq 0.7$ , indica correlação moderada

$0 \geq \text{coeficiente} \geq 0.3$ , indica fraca correlação

3) *Erro*: Um critério largamente utilizado para avaliar o desempenho de uma técnica de amostragem é o *erro relativo*. Esta métrica é utilizada em [15] e tem por objetivo avaliar a acurácia da técnica quanto ao processo de amostragem de tráfego. A sua equação é definida como:

$$ERM = \frac{|M_{total} - M_{estimado}|}{M_{total}} \quad (3)$$

Onde,  $M_{total}$  é o débito médio do tráfego total e  $M_{estimado}$  é o débito médio do tráfego final amostrado.

## V. RESULTADOS

Nos testes realizados aplicou-se o algoritmo de amostragem multi-adaptativa com definições de ordem entre 1 e 5. Assim como definido para a técnica LP [12], os melhores resultados foram alcançados para ordem de predição igual a 2. Isto identifica que, ao usar uma ordem mais baixa, o amostrador torna-se mais reativo aos comportamentos passados mais recentes, adequando-se melhor à natureza do tráfego de rede em rajada.

Em todos os testes comparativos, o amostrador multi-adaptativo atingiu índices de redução de *overhead* altamente significativos. A Figura 2 apresenta visualmente a redução alcançada nos critérios de quantidade de pacotes, volume

de dados capturados e número de amostras, respectivamente. Estes resultados são discutidos, quantitativamente, na Tabela III e, correspondem a um período de aproximadamente quatro horas de medição.

De acordo com a Tabela III, foram necessários apenas 5.3% do total de pacotes do tráfego para caracterizá-lo. Esta fração dos pacotes equivale a apenas 5.4% de todo o volume de dados em trânsito na rede analisada.

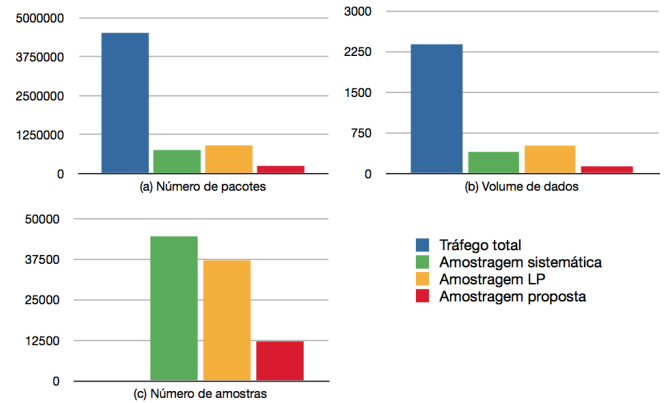


Figura 2. Redução do *overhead*

Tabela III  
REDUÇÃO DO *overhead*

Parâmetro	Total	Sistemática	LP	Multi-adaptativa
Número de pacotes	4513615	749977	899145	240159
Volume de dados (MB)	2382.92	395.97	511.93	128.78
Número de amostras	-	44511	37104	12152

Ainda de acordo com a Tabela III, quando comparada com a técnica sistemática, mais amplamente utilizada atualmente, a técnica multi-adaptativa, reduziu em 67.97% o número de pacotes necessários para caracterizar, estatisticamente, o tráfego. A redução no volume de dados capturados é de 67.47% e o número de amostras é reduzido em 72.69%.

Em comparação com a técnica LP, a amostragem multi-adaptativa alcança uma redução de 73.29% no número de pacotes, 74.84% no volume de dados total e 67.24% no número de amostras necessárias para a caracterização. Esta redução demonstra a potencialidade de evolução nas técnicas baseadas em predição linear, quando aplicadas considerando o tamanho da amostra como fator adaptativo.

Para demonstrar a capacidade de representar o tráfego, através de um volume de dados significativamente inferior, foram também adotadas representações visuais e estatísticas que a seguir se apresentam.

A maioria das ferramentas para gestão de rede dispõem de recursos para a representação gráfica do comportamento da rede em tempo real. Geralmente, este gráfico permite visualizar o volume de dados em kilobits por intervalos de

tempo de um segundo. A Figura 3 apresenta esta representação para cada técnica de amostragem usada nos testes.

Nela, a Figura 3(a) corresponde ao tráfego total não amostrado, ou seja, este é o comportamento real da rede, analisado em um período de aproximadamente quatro horas. A Figura 3(b), corresponde ao comportamento visual do tráfego resultante do processo de amostragem sistemática. Na Figura 3(c), é representado o tráfego resultante do processo de amostragem LP. Já a Figura 3(d), corresponde ao tráfego amostrado através da técnica multi-adaptativa.

Todas as representações do tráfego amostrado, apresentadas na Figura 3, correspondem, visualmente, ao tráfego total não amostrado. Isto significa que, mesmo usando aproximadamente 5% do volume total de dados em trânsito, a técnica de amostragem multi-adaptativa, é capaz de reproduzir adequadamente o comportamento da rede analisada.

A Tabela IV apresenta a capacidade de representação estatística do tráfego. Nela, observa-se que, o débito médio, estimado através da técnica de amostragem multi-adaptativa, apresenta o valor mais próximo, entre as técnicas de amostragem, do valor real do parâmetro. É importante observar que, neste caso, há uma notável evolução quando comparado com a técnica LP, usada como referência.

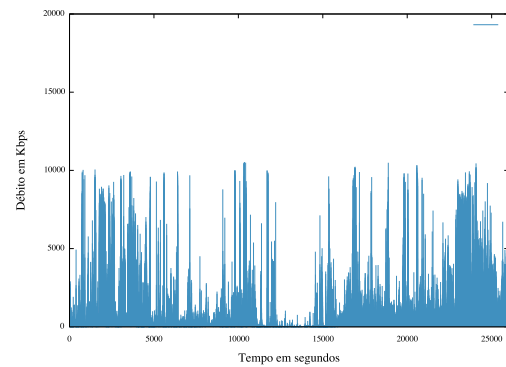
O melhor desempenho alcançado pela amostragem multi-adaptativa é ratificado através do erro relativo médio, apresentado na Tabela IV. Através desta estimativa, observa-se que o erro relativo associado a todo o processo de amostragem na técnica proposta é da ordem de 0.1%. A amostragem sistemática apresenta erro ligeiramente superior, na ordem de 0.2%. Já a amostragem LP, apresenta erro relativo na ordem de 54%. Este elevado índice de erro é resultado de uma significativa sobrestimação do parâmetro de referência, como apresentado na Tabela IV.

Tabela IV  
ESTIMATIVAS DE DÉBITO

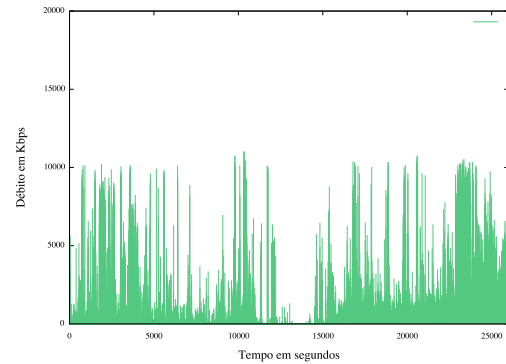
Parâmetro	Total	Sistemática	LP	Multi-adaptativa
Débito médio (Kbps)	730.95	728.76	1130.27	732.00
Explosividade	14.39	15.16	10.05	15.42
CV	1.90	2.07	1.56	2.07
Correlação	-	0.91	0.91	0.89
Erro relativo	-	0.002	0.54	0.001

Nos índices referentes à explosividade e ao coeficiente de variação, que permitem quantificar o comportamento em rajada do tráfego da rede, a técnica multi-adaptativa atinge, novamente, resultados significativamente melhores que a técnica LP, e semelhantes aos resultados da amostragem sistemática para a explosividade. Para o coeficiente de variação, os resultados obtidos são idênticos.

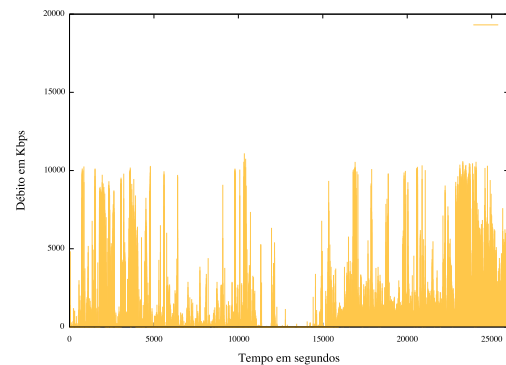
Já na correlação, embora com um coeficiente ligeiramente inferior, a técnica desenvolvida continua a amostrar tráfego resultante fortemente correlacionado com o tráfego total da rede. A diferença obtida é justificada pela significativa redução do volume de dados utilizado na estimativa do parâmetro. No entanto, o menor coeficiente de correlação não gera distorções significativas na caracterização do tráfego.



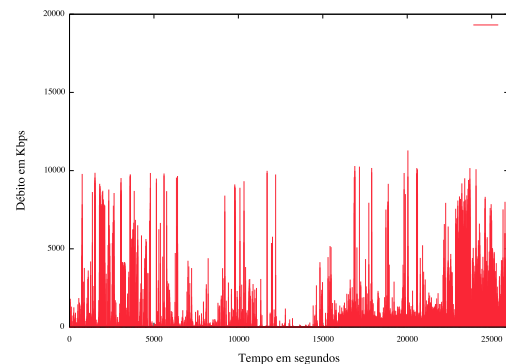
(a) Tráfego total



(b) Amostragem Sistemática



(c) Amostragem LP



(d) Amostragem Multi-adaptativa

Figura 3. Tráfego instantâneo

Para determinar o padrão de incidência dos erros de estimação, utilizam-se gráficos de dispersão, como apresentados na Figura 4. Neles, para cada intervalo de um segundo, onde ocorreu a captura de, pelo menos, um pacote, compara-se a estimativa do débito com o débito real do tráfego não amostrado no mesmo intervalo. Quanto mais próxima a estimativa estiver da linha de referência, mais correta é a estimativa do débito no respectivo intervalo.

Os gráficos de dispersão, apresentados na Figura 4, indicam que todas as técnicas de amostragem se mostram mais eficazes em momentos de baixo débito na rede. Isto caracteriza-se pela distribuição dos desvios na estimação, que se concentram mais próximos da linha de referência nestes instantes de observação.

## VI. CONCLUSÕES

Neste artigo foi proposta uma técnica de amostragem multi-adaptativa de tráfego baseada em predição linear que visa reduzir o *overhead* associado aos processos de monitorização mantendo a acurácia na estimação do comportamento da rede. Esta redução foi alcançada considerando o tamanho da amostra como um segundo fator adaptativo, além do intervalo entre amostras.

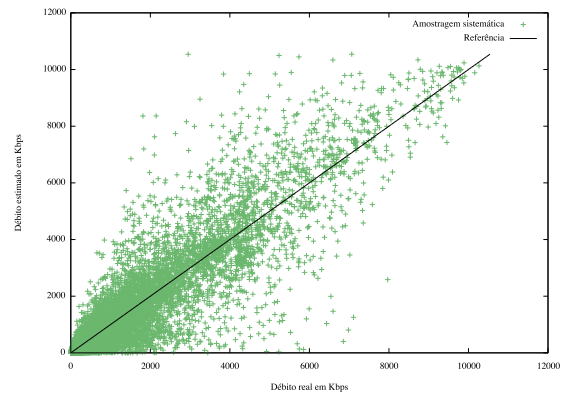
Os testes aplicados à técnica de amostragem multi-adaptativa de tráfego demonstram que é possível reduzir, significativamente, o *overhead* associado ao volume de dados envolvidos no processo de monitorização de rede. Mesmo com tal redução, a técnica desenvolvida é capaz de caracterizar, estatisticamente e visualmente, o tráfego analisado, alcançando em vários parâmetros resultados mais precisos que as técnicas classicamente utilizadas.

Atualmente, estão a ser efetuados testes com um conjunto mais heterogêneo de tipos de tráfego, como tráfego de *backbone*, vídeo *streaming* e VoIP - *Voice over IP*. Estes novos testes procuram identificar que aspectos contribuem para uma melhoria tão significativa nos resultados obtidos e se tal desempenho é alcançado de forma generalista.

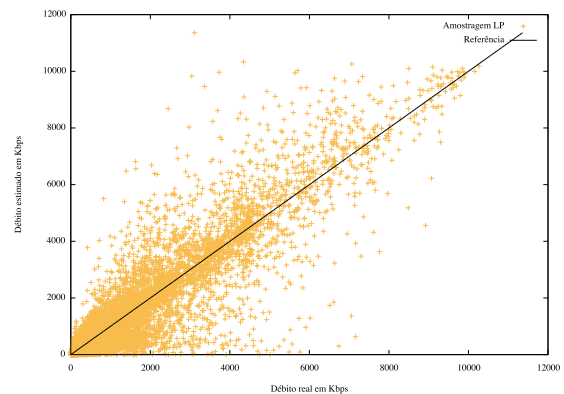
A sequência deste trabalho consistirá na evolução do algoritmo de forma a permitir que novos parâmetros sejam estimados através da técnica proposta, nomeadamente os parâmetros multiponto como atraso, variação de atraso e perda.

## REFERÊNCIAS

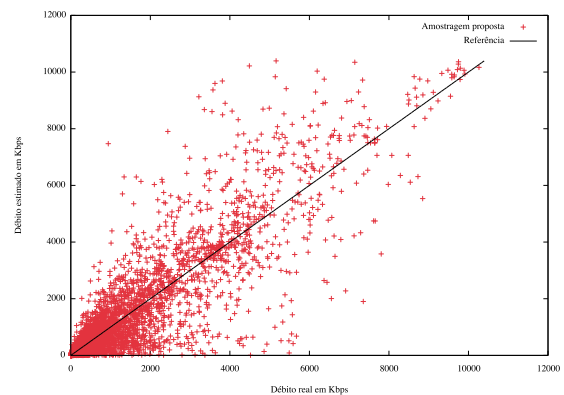
- [1] T. Zseby, M. Molina, and N. Duffield, "Sampling and Filtering Techniques for IP Packet Selection," RFC 5475, Internet Engineering Task Force, Mar. 2009.
- [2] S. K. Thompson, *Sampling*, 2nd ed. Wiley-Interscience, 2002.
- [3] C. Henke, C. Schmoll, and T. Zseby, "Empirical evaluation of hash functions for packetid generation in sampled multipoint measurements," in *Passive and Active Network Measurement*, ser. Lecture Notes in Computer Science, S. Moon, R. Teixeira, and S. Uhlig, Eds. Springer Berlin / Heidelberg, 2009, vol. 5448, pp. 197–206.
- [4] R. Serral-Gracia, A. Cabellos-Aparicio, and J. Domingo-Pascual, "Packet loss estimation using distributed adaptive sampling," in *Network Operations and Management Symposium Workshops, 2008. NOMS Workshops 2008. IEEE*, 2008, pp. 124–131.
- [5] J. Sommers, P. Barford, N. Duffield, and A. Ron, "Improving accuracy in end-to-end packet loss measurement," in *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '05. New York, NY, USA: ACM, 2005, pp. 157–168.



(a) Amostragem Sistemática



(b) Amostragem LP



(c) Amostragem Multi-adaptativa

Figura 4. Gráficos de dispersão

- [6] Z.-G. Hu, D.-L. Zhang, C.-P. Hou, and J.-S. Zhang, "Adaptive sampling algorithm of network round-trip time," in *Journal of Computer Applications*, J. Yingyong, Ed., vol. 30, no. 2, feb 2010, pp. 319–322.
- [7] T. Zseby, T. Hirsch, and B. Claise, "Packet sampling for flow accounting: Challenges and limitations," in *Passive and Active Network Measurement*, ser. Lecture Notes in Computer Science, M. Claypool and S. Uhlig, Eds. Springer Berlin / Heidelberg, 2008, vol. 4979, pp. 61–71.
- [8] C. Hu, S. Wang, J. Tian, B. Liu, Y. Cheng, and Y. Chen, "Accurate and efficient traffic monitoring using adaptive non-linear sampling method," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008, pp. 26–30.
- [9] A. N. Mahmood, J. Hu, Z. Tari, and C. Leckie, "Critical infrastructure protection: Resource efficient sampling to improve detection of less frequent patterns in network traffic," *Journal of Network and Computer*

*Applications*, vol. 33, no. 4, pp. 491 – 502, 2010.

- [10] J. Giertl, J. Baca, F. Jakab, and R. Andoga, "Adaptive sampling in measuring traffic parameters in a computer network using a fuzzy regulator and a neural network," *Cybernetics and Systems Analysis*, vol. 44, pp. 348–356, 2008, 10.1007/s10559-008-9005-0.
- [11] Q. Xin, L. Hong, and L. Fang, "A modified flc adaptive sampling method," in *Communications and Mobile Computing, 2009. CMC '09. WRI International Conference on*, vol. 2, 2009, pp. 515 –520.
- [12] E. A. Hernandez, M. C. Chidester, and A. D. George, "Adaptive sampling for network management," *Journal of Network and Systems Management*, vol. 9, pp. 409–434, 2001, 10.1023/A:1012980307500. [Online]. Available: <http://dx.doi.org/10.1023/A:1012980307500>
- [13] Y. Lu and C. He, "Resource allocation using adaptive linear prediction in wdm/tdm epons," *AEU - International Journal of Electronics and Communications*, vol. 64, no. 2, pp. 173 – 176, 2010.
- [14] Y. Wei, J. Wang, and C. Wang, "A traffic prediction based bandwidth management algorithm of a future internet architecture," *Intelligent Networks and Intelligent Systems, International Workshop on*, vol. 0, pp. 560–563, 2010.
- [15] A. Dogman, R. Saatchi, and S. Al-Khayatt, "An adaptive statistical sampling technique for computer network traffic," in *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, 2010, pp. 479 –483.
- [16] A. Schulman, D. Levin, and N. Spring, "CRAWDAD trace umd/sigcomm2008/pcap/ethernet (v. 2009-03-02)," Downloaded from <http://crawdad.cs.dartmouth.edu/umd/sigcomm2008/pcap/Ethernet>, Mar. 2009.
- [17] K. Pearson, *On a New Method of Determining Correlation Between a Measured Character A, and a Character B, of which Only the Percentage of Cases Wherein B Exceeds (or Falls Short of) a Given Intensity is Recorded for Each Grade of A*, 1st ed. *Biometrika*, 1909, vol. 7, pp. 96–105.