

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
ESCOLA DE INFORMÁTICA APLICADA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

UMA FERRAMENTA AMIGÁVEL PARA COLETA E
ANÁLISE DE TRÁFEGO MULTIMÍDIA

Anderson Pereira
João Alberto Neves dos Santos Filho

Morganna Carmem Diniz

Julho/2005

UMA FERRAMENTA AMIGÁVEL PARA COLETA E
ANÁLISE DE TRÁFEGO MULTIMÍDIA

Projeto de Graduação apresentado à Escola
de Informática Aplicada da Universidade
Federal do Estado do Rio de Janeiro
(UNIRIO) para obtenção do título de
Bacharel em Sistemas de Informação

Anderson Pereira
João Alberto Neves dos Santos Filho

Morganna Carmem Diniz

ANDERSON PEREIRA - AGRADECIMENTOS

Gostaria de mostrar minha gratidão à todas as pessoas que nesses quatro anos torceram, sofreram, lutaram e comemoraram comigo neste longo e marcante estágio da minha vida.

Agradeço à toda minha família por ser um exemplo de amor e apoio mesmo nas horas difíceis; Nenhuma educação que já recebi ou receberei se comparará àquela que vocês me ensinaram e que com orgulho absorvi. Amo vocês.

À minha noiva quase esposa, por transbordar amor e compreensão desde quando nos conhecemos na faculdade há três anos atrás. Tenho certeza que nunca seria tão completo e não chegaria onde cheguei sem você ao meu lado. Você é o maior exemplo de garra e perseverança que eu já conheci, pois nunca me deixou desistir. Eu te amo do fundo do meu coração.

Aos meus amigos de verdade, que mesmo presentes ou ausentes, sempre mostrando cada vez mais que amizades fortes atravessam anos e mesmo assim continuam inabaláveis. Quero agradecer especialmente a João Alberto por ser um incrível amigo desde o início da faculdade até os últimos dias, que nossa amizade nunca se acabe.

Aos nossos professores, que, mesmo os mais rígidos, por todo este tempo vêm se esforçando para despertar em nós um interesse incessante de aprender e crescer, já querendo nos preparar para o mundo que enfrentaremos lá fora. Em especial gostaria de agradecer à dois professores: Prof.a Morganna por, há pelo menos um ano, nos orientar tão bem nessa fase de transição que estamos passando, e agradecer a quem considere meu maior e mais importante tutor, e que com maior orgulho segui e me interessei em sua área de conhecimento: Prof. Maurício Ferreira. Infelizmente não pôde estar agora ao meu lado nesta fase final mas mesmo assim ainda não sei como agradecer por tudo que fez e estou aqui mostrando toda minha gratidão.

E finalmente a Deus por tornar isto tudo possível.

Anderson Pereira

JOÃO ALBERTO NEVES DOS SANTOS FILHO – AGRADECIMENTOS

Em primeiro lugar, quero agradecer à Deus, pois Ele é o responsável por tudo o que consegui conquistar em toda a minha vida.

Minha família constitui toda a base para encarar todas as situações presentes no cotidiano. A começar pelo meu irmão, tem uma personalidade muito diferente da minha, mas é uma pessoa companheira e atenciosa, para qualquer hora, eu o admiro muito. Minha mãe, sempre preocupada com até os mínimos detalhes, é um exemplo de mãe que me apoiou em muitas decisões difíceis, e me fez compreender que, mesmo as situações mais difíceis, tudo acontece para o próprio amadurecimento. Meu pai, o mais responsável e trabalhador, é meu ídolo. Possui um conhecimento inacreditável e, ao mesmo tempo, é um ótimo pai.

À minha namorada, Bruna. Graças à ela, mudei radicalmente a minha personalidade, tornando-me uma pessoa que sempre busca o equilíbrio entre o emocional e o intelectual. É uma pessoa extraordinária, muito cativante, carinhosa e atenciosa, me ajuda em todos os momentos.

Aos meus amigos, por todos os momentos que compartilhamos. Especialmente ao Anderson, possui personalidade forte, mas sempre me deu forças a nunca desistir de minhas metas e revelou-se o melhor amigo que pude conquistar na faculdade.

Finalmente, aos professores da UNIRIO, por todo o conhecimento que pude adquirir através dos mesmos. No ambiente da universidade, aprendi a dar importância aos estudos e, ao passar dos períodos, fui amadurecendo para a vida. Em especial, quero agradecer ao professor Márcio Barros, por sua simplicidade, disposição em ajudar e carisma. Também não poderia me esquecer da professora Morganna, por ser uma pessoa muito atenciosa e acreditar no meu trabalho durante o desenvolvimento do projeto de graduação. Acreditou tanto, que me tornou bolsista de iniciação científica no CNPq e me alocou trabalhos em parceria com o IME, tudo em relação ao projeto.

João Alberto Neves dos Santos Filho

EPÍGRAFE

"É melhor tentar e falhar,
que se preocupar e ver a vida passar;
é melhor tentar, ainda que em vão,
que se sentar fazendo nada até o final.
Eu prefiro na chuva caminhar,
que em dias tristes me esconder.
Prefiro ser feliz, embora louco,
que em conformidade viver ..."

Martin Luther King

SUMÁRIO

CAPÍTULO 1 – INTRODUÇÃO	1
1.1 Visão geral do projeto	1
1.2 Objetivos	3
1.2.1 Medição do tráfego multimídia	3
1.2.2 Desenvolvimento da ferramenta	4
1.3 Organização do trabalho	5
CAPÍTULO 2 – IMPLEMENTAÇÃO DA FERRAMENTA	6
2.1 Coleta de dados	6
2.1.1 Estrutura de rede da UNIRIO	6
2.1.2 Ethereal	7
2.1.2.1 Agendamento por Crontab.....	8
2.1.2.2 Scripts de coleta	8
2.1.2.3 Amostra de coleta	9
2.2 Solver	9
2.3 Interface gráfica	9
2.3.1 Descrição conceitual	9
2.3.2 Diagrama de classes (modelo conceitual)	13
2.3.3 Diagrama de casos de uso	13
2.3.4 Descrição da interface gráfica – visão geral	14
2.3.4.1 Processing	16
2.3.4.1.1 Statistics	16
2.3.4.1.2 Mean/variance	17
2.3.4.1.3 Autocorrelation	17

2.3.4.1.4 Regression	17
2.3.4.1.5 Heavy tail distribution	18
2.3.4.1.6 pdf (probability of density function)	18
2.3.4.2 Estimators	18
2.3.4.2.1 Higuchi	19
2.3.4.2.2 VAR	19
2.3.4.2.3 R/S	19
2.3.4.2.4 Wavelets	20
2.3.4.3 Graphics	20
2.3.4.4 About	23
2.3.4.4.1 Help	23
2.3.4.4.2 About SET	24
2.3.4.5 Quit	24
CAPÍTULO 3 – ESTUDO DE CASOS	25
3.1 Banda de rede da UNIRIO	25
3.2 Gráficos de amostras	25
3.2.1 Autocorrelação	25
3.2.2 Regressão	26
3.2.3 Cauda Pesada	26
3.2.4 Densidade	27
3.2.5 Estimador Higuchi	27
3.2.6 Estimador R/S	28
3.2.7 Estimador VAR	28
3.2.8 Estimador Wavelets	29
3.3 Resultado do estudo de caso	29

CAPÍTULO 4 – CONCLUSÃO	31
REFERÊNCIAS	32
ANEXO I	34

RESUMO

Este projeto tem como finalidade o desenvolvimento de uma ferramenta gráfica multiplataforma que analise o tráfego de dados coletados em uma determinada rede, a fim de auxiliar o usuário a estudar o comportamento dos fluxos dos dados baseados em variados tipos de algoritmos matemáticos conhecidos e aplicados no processamento destes tráfegos. A fim de facilitar estas pesquisas, a ferramenta possui uma interface gráfica amigável e intuitiva que possibilita ao usuário traçar, com base nos dados coletados, gráficos em duas dimensões de cada algoritmo. O foco da ferramenta baseia-se em modelos não-markovianos, pois se tratam de modelos de estimação ainda não explorados neste campo de pesquisa. Foi comprovado que os modelos markovianos, tradicionalmente utilizados para estimar o desempenho de sistemas, não possibilitam apresentar efetivamente uma regularidade estatística, ou seja, não são confiáveis do ponto-de-vista matemático para servirem de fonte de pesquisas comportamentais de redes.

PALAVRAS-CHAVES: markoviano, fluxo, rede, análise.

ABSTRACT

This project has as purpose the development of a multiplatform graphical tool that analyzes the collected data traffic of a given network, in order to assist the users to study the behavior of the flows of data based on varied types of known and applied mathematical algorithms in the processing of these traffics. In order to facilitate this research the tool has a friendly and intuitive graphical interface that makes possible the users to trace, with this collected data, graphics in two dimensions of each cited algorithm above. The focus of the tool is based on non-markovian models, because these estimation models still not explored in this field of research. It was proven that the markovian models, traditionally used to estimate the performance of systems, do not make possible to present a regular and effective statistics, or either, they are not trustworthy of the mathematical point-of-sight to serve of source of manning research of networks.

KEYWORDS: markovian, flow, network, analysis.

SUMÁRIO DE ILUSTRAÇÕES

CAPÍTULO 2

2.1	O ambiente usado na coleta de dados da rede UNIRIOTEC	6
2.2	Script de agendamento do Crontab	8
2.3	Conteúdo do script RUN_IN	8
2.4	Conteúdo do script RUN_OUT	9
2.5	Amostra de uma coleta	9
2.6	Fluxo do pacote JFreeChart	11
2.7	Demonstração da utilização do JFreeChart	12
2.8	Diagrama de classes do SET	13
2.9	Diagrama de casos de uso do SET	14
2.10	Tela inicial do SET	14
2.11	Console de mensagens	15
2.12	Seleção de arquivos	16
2.13	Menus da ferramenta	16
2.14	Janela de geração de estatísticas	17
2.15	Janela de cálculo de média/variância	17
2.16	Janela de autocorrelação	18
2.17	Janela de regressão	18
2.18	Janela de cauda pesada	18
2.19	Janela de pdf	19
2.20	Janela do estimador Higuchi	19
2.21	Janela do estimador VAR	20
2.22	Janela do estimador R/S	20
2.23	Janela do estimador Wavelets	20
2.24	Janela de adição dos gráficos	21
2.25	Janela de seleção dos arquivos/gráficos	22
2.26	Janela dos gráficos	22
2.27	Zoom de um gráfico	23
2.28	Help da ferramenta	24
2.29	Sobre os autores	24
2.30	Janela de saída	25

CAPÍTULO 3

3.1	Gráfico de autocorrelação	26
3.2	Gráfico de regressão	27
3.3	Gráfico de cauda pesada	27
3.4	Gráfico de densidade	28
3.5	Gráfico de Higuchi	28
3.6	Gráfico de R/S	29
3.7	Gráfico de VAR	29
3.8	Gráfico de Wavelets	30
3.9	Gráfico de médias kbytes/s da rede UNIRIOTEC (horas)	31
3.10	Gráfico de médias kbytes/s da rede UNIRIOTEC (minutos)	31

Capítulo 1 - Introdução

1.1 Visão geral do projeto

Com o advento da Internet, a comunicação entre países tornou-se mais fácil e rápida. Hoje, uma notícia de qualquer parte do mundo leva apenas alguns minutos para chegar ao Brasil. Basta que alguém se prontifique a mandar um e-mail ou a escrever uma página em um site ou em um blog. Meios de comunicação tradicionais, como telefone, também não ficaram imunes à Internet. Por exemplo, é possível baratear ligações telefônicas usando na Internet protocolos como VoIP [1]. A Internet abriu um leque de serviços e novas possibilidades de comunicação. Entretanto, também trouxe centenas de problemas a serem solucionados.

O problema principal está na própria origem da Internet, que nada mais é que uma coleção de redes de comunicação independentes. Não existe uma administração centralizada e global na Internet. Cada administrador de rede pode escolher os equipamentos de comunicação que melhor lhe convém e pode definir os protocolos de comunicação que lhe são mais apropriados. Uma comunicação entre dois equipamentos quaisquer na Internet pode ter de passar por dezenas de redes diferentes com diferentes qualidades de serviço. É por isso que a Internet oferece o serviço *best-effort*. Isto significa que pacotes de dados transmitidos podem chegar com atraso ao seu destino ou mesmo podem não chegar. É da responsabilidade dos sistemas finais (transmissor e receptor) verificar as distorções e corrigir os erros na comunicação quando necessário. Portanto, caracterizar o tráfego a ser transmitido entre os equipamentos de redes é um passo importante no desenvolvimento de novos aplicativos e na melhoria da qualidade de serviço a ser oferecida.

A partir de uma caracterização do tráfego mais precisa, é possível desenvolver modelos matemáticos que permitam estudar/avaliar o comportamento do sistema de comunicação em diferentes condições. Atualmente, os modelos de tráfego podem ser divididos em dois grandes grupos: modelos markovianos [2] e modelos não-markovianos (ou fractais)[3].

Basicamente, o primeiro grupo tem como característica a dependência de curto alcance (SRD – *short range dependence*). Isto significa que o próximo estado (comportamento) do sistema sendo modelado depende apenas do estado atual do sistema.

O segundo grupo argumenta que isto não é necessariamente verdade: o próximo estado do sistema pode estar relacionado a um estado mais antigo e não obrigatoriamente ao estado atual em que o sistema se encontra. Isto significa que existe um padrão de comportamento no tráfego multimídia que se repete a um certo período de tempo, e portanto, os modelos matemáticos devem refletir esta dependência temporal ou dependência de longo alcance (*long range dependence*).

Na última década, vários estudos comprovaram a presença de dependência temporal em diversos tipos de tráfego [5] como, por exemplo, tráfego WWW [6], WAN [7], ATM [8] e vídeo VBR [9]. Segundo esses (e muitos outros) trabalhos, as características do tráfego de LAN/WAN não são adequadamente representadas pelos modelos markovianos. Entretanto, este novo cenário de pesquisa tem como grande problema a falta de ferramentas de modelagem e análise. Na realidade, o que temos hoje é uma coleção de métodos de análise para definir a dependência temporal do tráfego e fórmulas matemáticas fechadas para o cálculo de certas características dos modelos. Infelizmente, o uso destes métodos e fórmulas não é uma tarefa simples. É necessário usar vários métodos matemáticos para se obter a caracterização da dependência temporal do tráfego. Além disso, muitas das formulações existentes só são aplicáveis a uma determinada classe de tráfego, ou seja, não são de uso geral.

Este trabalho teve basicamente dois objetivos: a coleta de tráfego de dados da UNIRIO e o desenvolvimento da ferramenta SET com as seguintes facilidades:

1. Análise dos dados coletados usando diferentes métodos (R/S [12], Higuchi [13] e estimador AV baseado em *wavelets* [14]);
2. Regressão;
3. Cálculo do parâmetro Hurst - este parâmetro fornece o grau de dependência temporal dos dados estudados;
4. Gráficos das curvas dos *traces* e dos dados gerados pelos métodos de análise.

É importante fazer duas observações em relação ao desenvolvimento da ferramenta SET. Em primeiro lugar, existe uma ferramenta desenvolvida pelo laboratório do IME chamada HEP (*Hurst Estimator Package* – veja [12, 13]) e que serviu de base para o desenvolvimento da nova ferramenta. A ferramenta HEP implementa os métodos R/S, Higuchi e estimador AV baseado em *Wavelets*. Basicamente, o usuário fornece, na linha de comandos, o nome do arquivo de entrada e o nome do método a ser utilizado. A saída é gravada em um outro arquivo.

Em segundo lugar, existem ferramentas que executam alguns dos passos acima definidos para a ferramenta SET, mas não existe uma ferramenta que execute todos os passos. Por exemplo, pode-se usar alguma ferramenta que implementa o método R/S (por exemplo, a ferramenta HEP) para se obter o resultado da análise no formato (x,y). Em seguida, deve-se usar alguma outra ferramenta para fazer regressão (por exemplo, a ferramenta Grace (<http://plasma-gate.weizmann.ac.il/Grace>) e assim, encontrar o parâmetro Hurst. Outra observação importante, é que não se usa apenas um método para encontrar o parâmetro Hurst. O estudo é feito usando os três métodos e analisando os resultados obtidos de forma a se chegar a um único valor. Pode ser necessário usar o mesmo método várias vezes até se obter um valor aproximado entre eles.

1.2 Objetivos

O objetivo principal dessa monografia foi o desenvolvimento de uma ferramenta gráfica para auxiliar a medição do tráfego em uma rede e analisar suas características. Como exemplo do uso da ferramenta, também foi feito o estudo do tráfego que entra e sai na UNIRIO. Para facilitar a compreensão deste trabalho, o projeto será aqui apresentado em duas partes: primeiro, a medição do tráfego e, segundo, o desenvolvimento da ferramenta.

1.2.1 Medição do tráfego multimídia

O objetivo desta parte do projeto foi fazer a medição do tráfego na UNIRIO em vários horários. Esses dados foram usados então para estudar o comportamento de

rede e caracterizar o tráfego que entra e sai do Departamento de Informática Aplicada (DIA).

Primeiramente, a estrutura da rede da UNIRIO precisou de ajustes para que se adaptasse às necessidades do projeto. Após o término da análise foi decidido que o melhor seria que a infra-estrutura de medição ficasse na DMZ (DeMilitarized Zone). A DMZ é criada especificamente para portar servidores (que possuem proteções próprias por não terem resguardo do firewall) cujos serviços também são oferecidos para máquinas fora da rede interna, ou seja, na Internet (serviços como webmail, website, ftp, etc...). Todas as máquinas da DMZ possuem IP válido na Internet. Os IP's mais importantes neste projeto foram: 200.143.194.3 (Firewall da UNIRIO) e 200.143.194.6 (Computador de coleta e desenvolvimento da ferramenta SET).

O próximo passo foi selecionar uma ferramenta para fazer a coleta. Foram feitas comparações entre as ferramentas disponíveis para auxiliar neste trabalho. Descobriu-se, entre outras, uma ferramenta chamada **Ethereal** (<http://www.ethereal.com>), um analisador de tráfego de rede com licença **GNU** (Licença de Software Livre). Esta ferramenta foi escolhida, por ser a mais completa, rápida e a mais fácil de configurar entre as ferramentas estudadas.

1.2.2 Desenvolvimento da ferramenta

O objetivo desta parte do projeto foi desenvolver a ferramenta SET (Scaling Estimation Tool) para análise de características temporais do tráfego de dados.

A ferramenta é constituída por duas partes:

- Solver: possui os algoritmos necessários para o cálculo das características estatísticas e analíticas dos tráfegos de dados. Ele foi desenvolvido em linguagem C++. O HEP (veja na seção 22) já havia sido construído em C++ e foi integrado à ferramenta. As outras funcionalidades foram também integradas à ferramenta, e alguns desses algoritmos foram escritos em C; para atender às necessidades do Solver, foram convertidas para C++.

- Interface gráfica: desenvolvida em linguagem Java por esta possuir uma ampla biblioteca de objetos gráficos e por ter um suporte a diferentes formatos de arquivos para impressão. Além disso, esta linguagem é independente de plataforma, possibilitando o uso da ferramenta em Windows e Linux.

1.3 Organização do trabalho

Esta monografia tem a seguinte organização: no Capítulo 2 são discutidos os problemas e a implementação da ferramenta. No Capítulo 3 é feito um estudo do tráfego da UNIRIO usando a ferramenta SET. E no Capítulo 4 são apresentadas as conclusões e referências de pesquisa.

Capítulo 2 – Implementação da Ferramenta

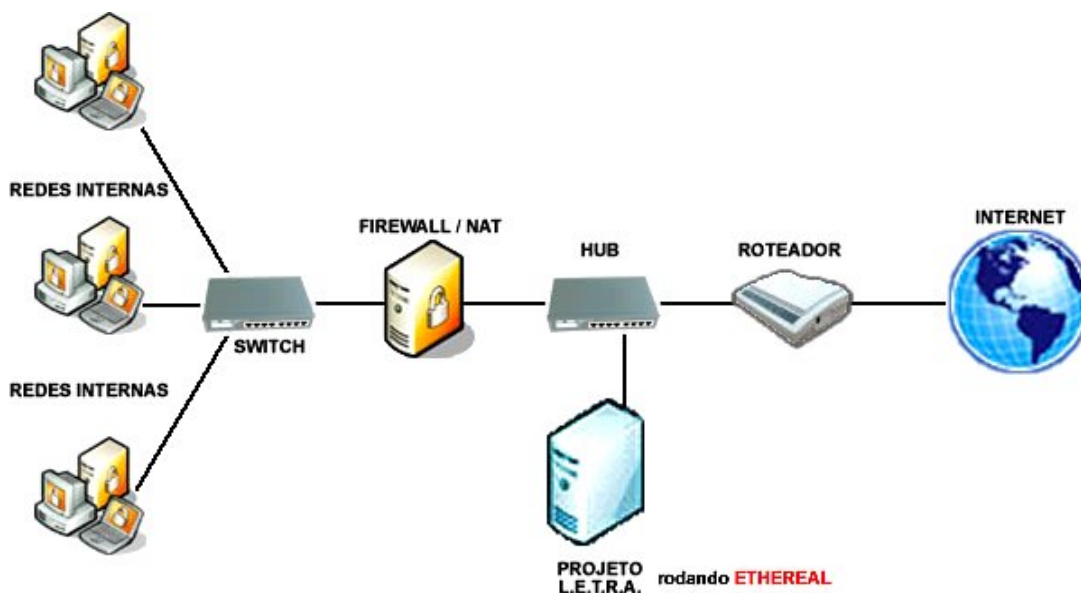
2.1 Coleta de dados

2.1.1 Estrutura da Rede da UNIRIO

Para podermos realizar nossos estudos, primeiramente foi necessário capturar o tráfego da rede interna da UNIRIO. Para isso tivemos que instalar um computador ligado a um Hub agindo como um M.I.T.M. (Man-in-the-Middle) entre o roteador da Internet e o Firewall que protege a rede interna. Assim todo o tráfego que entra ou sai da UNIRIO passa pelo Hub, que por sua vez distribui este tráfego para todos os computadores nele conectados, inclusive o computador de coleta.

Na ilustração 2.1 destacamos a parte da estrutura da rede que importa para nossa pesquisa. Escolhemos o lado externo (DMZ → DeMilitarized Zone) da rede da universidade para colocar o equipamento do projeto, pois assim temos um controle maior devido ao número diminuído de IP's próximos.

AMBIENTE DE REDE DA PESQUISA



2.1. O ambiente usado na coleta de dados da rede UNIRIOTEC. [19]

Para o estudo da rede, o computador de coleta ficou desprovido de acesso à Internet nos horários de coleta (12:00 → 23:00) para que não se resultasse na adulteração da performance da banda compartilhada.

2.1.2 **Ethereal**

Utilizamos a versão 0.10.9 do Ethereal (versão compilada e atualizada), que proporciona o melhor ambiente de captura necessário para os requisitos do projeto. A versão empacotada (RPM), que vêm pré-compilada, não estava atendendo aos requisitos de desempenho e estabilidade devido ao fato de ter sido previamente compilada em outro computador com configurações distintas. [18]

O Ethereal, corretamente configurado, captura em **modo promíscuo de rede** informações específicas e selecionadas de cada pacote que trafega pela interface de rede (ETH) selecionada. Quais informações serão mostradas podem ser escolhidas nas configurações da ferramenta. Dentre as várias possíveis, para nossos estudos precisávamos de apenas algumas:

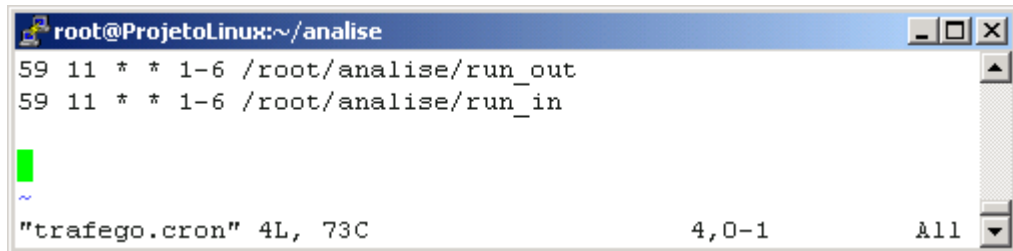
1. Hora exata (*com precisão de 10^{-6} (0.000001) segundos*)
2. Tamanho do pacote (em Bytes)
3. Protocolo do pacote

Foram programados e agendados para execução (em CRONTAB) dois scripts com a linha de comando (Vide ilustração 2.2) do Tethereal (**Sub-programa do Ethereal que gera coletas em texto ASCII**) para captura em arquivos distintos do tráfego de entrada e tráfego de saída da rede.

Para cada dia de coleta é criado pelo script um arquivo que possui os formatos **OUT_[mês]_[dia]** (Ex: out_11_19) ou **IN_[mês]_[dia]** (Ex: in_11_19), sendo estes respectivamente o tráfego de saída e de entrada do dia 19 de novembro. Foram feitos logs diários entre 12h e 23h, de segunda à sexta, independente de feriados. Considerados os tamanhos extensos de cada arquivos de log (entre 15mb e 40mb), decidiu-se compactar grupos de logs a cada mês.

2.1.2.1 AGENDAMENTO POR CRONTAB

Os scripts abaixo foram agendados na sintaxe de CRONTAB para serem ambos rodados paralelamente e iniciando-se ao mesmo tempo a partir de 11:59:00, pois se fossem definidas a partir do horário que queríamos (12:00), os pacotes entre o exato meio-dia e o tempo necessário para ele reconhecer e iniciar os scripts seriam perdidos, iniciando-se na verdade por volta de 12:00:03.700000.



```

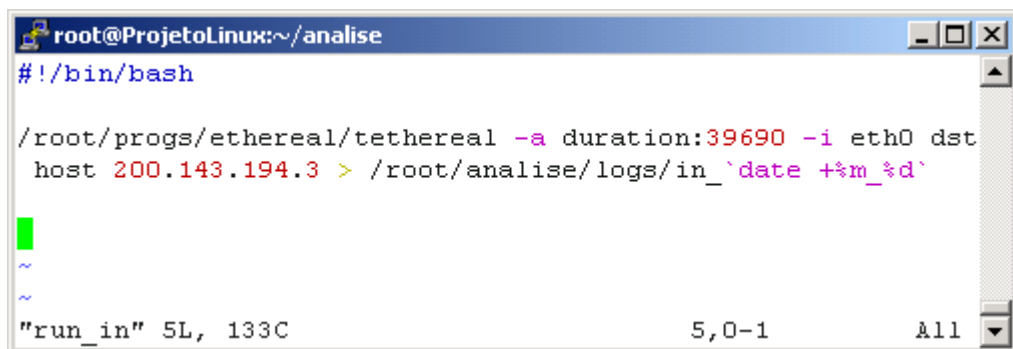
root@ProjetoLinux:~/analise
59 11 * * 1-6 /root/analise/run_out
59 11 * * 1-6 /root/analise/run_in

"trafego.cron" 4L, 73C          4,0-1      All

```

2.2 Script de agendamento do Crontab

2.1.2.2 SCRIPTS DE COLETA



```

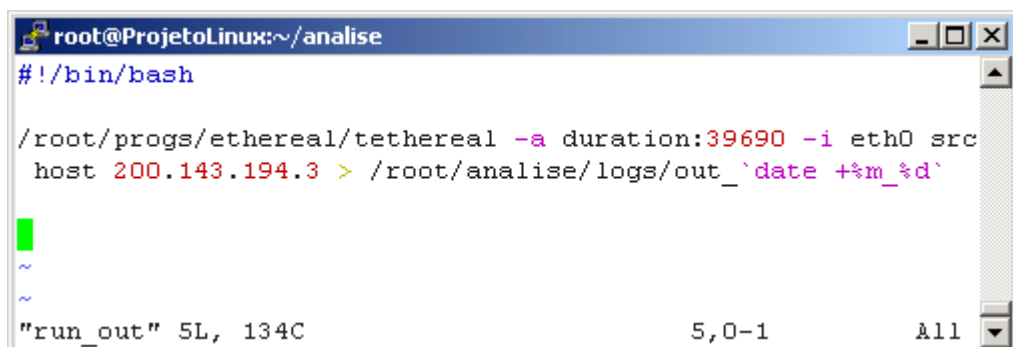
root@ProjetoLinux:~/analise
#!/bin/bash

/root/progs/ethereal/tethereal -a duration:39690 -i eth0 dst
host 200.143.194.3 > /root/analise/logs/in_`date +%m_%d`

"run_in" 5L, 133C          5,0-1      All

```

2.3 RUN_IN: coleta com duração de 39690 segundos (11 horas e alguns minutos de folga) na interface de rede ETH0 onde o pacote se **destina** ao IP 200.143.194.3 (Firewall da Unirio) e gravando, no diretório de logs, a saída em um arquivo de log padronizado dependendo dos dia e mês atuais. Quando o pacote se destina ao IP do Firewall significa que é um **tráfego de entrada**.



```

root@ProjetoLinux:~/analise
#!/bin/bash

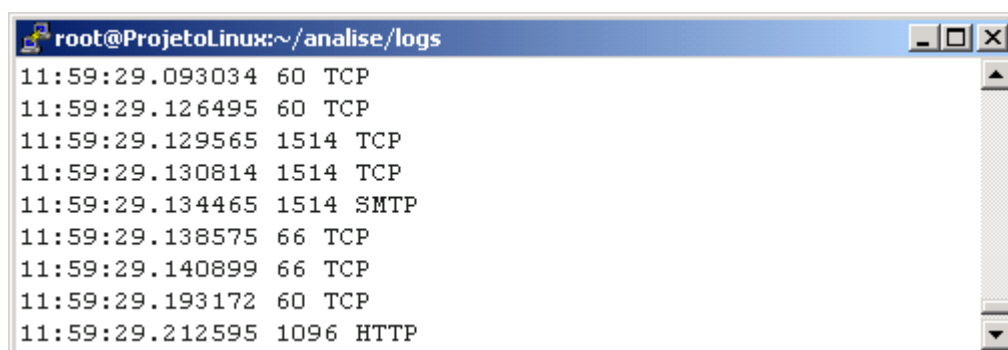
/root/progs/ethereal/tethereal -a duration:39690 -i eth0 src
host 200.143.194.3 > /root/analise/logs/out_`date +%m_%d`

"run_out" 5L, 134C          5,0-1      All

```

2.4 RUN_OUT: coleta com duração de 39690 segundos (11 horas e alguns minutos de folga) na interface de rede ETH0 onde o pacote se origina do IP 200.143.194.3 (Firewall da Unirio) e gravando, no diretório de logs, a saída em um arquivo de log padronizado dependendo dos dia e mês atuais. Quando o pacote se origina do IP do Firewall significa que é um **tráfego de saída**.

2.1.2.3 AMOSTRA DE COLETA

A terminal window titled 'root@ProjetoLinux:~/analise/logs' displays a list of network traffic logs. Each line represents a data point with a timestamp, bytes, and a protocol. The data is as follows:

Timestamp	Bytes	Protocol
11:59:29.093034	60	TCP
11:59:29.126495	60	TCP
11:59:29.129565	1514	TCP
11:59:29.130814	1514	TCP
11:59:29.134465	1514	SMTP
11:59:29.138575	66	TCP
11:59:29.140899	66	TCP
11:59:29.193172	60	TCP
11:59:29.212595	1096	HTTP

2.5 Pequena amostra de uma coleta onde cada coluna define um dado: hora exata, bytes e protocolo. Separados por espaço simples para que pudessem ser reconhecidos por qualquer programa leitor no futuro.

2.2 Solver

O Solver é uma ferramenta multifunção que foi desenvolvida neste projeto. Ela possibilita a execução de vários cálculos independentes que resultam em dados apropriados para o estudo dos tráfegos de rede.

O Solver é composto por vários processos (algoritmos de medição) onde cada processo recebe um arquivo de entrada, parâmetros específicos de cada algoritmo e gera um arquivo de saída, sendo parametrizado em linha de comando. As sintaxes do Solver e os objetivos de suas funcionalidades podem ser conferidos no Anexo I.

2.3 Interface gráfica

2.3.1 Descrição conceitual

A ferramenta SET fornece as seguintes facilidades, tendo em vista seus principais objetivos:

- Análise dos dados coletados usando diferentes métodos (R/S, Higuchi, Var e estimador AV baseado em wavelets);
- Regressão linear e auto-correlação;
- Cálculo do parâmetro Hurst;
- Gráficos das curvas dos traces e dos dados gerados pelos métodos de análise;
- Medidas estatísticas (média e variância);

- DCP (Distribuição de Cauda Pesada) e FDP (Frequência de Densidade Probabilística).

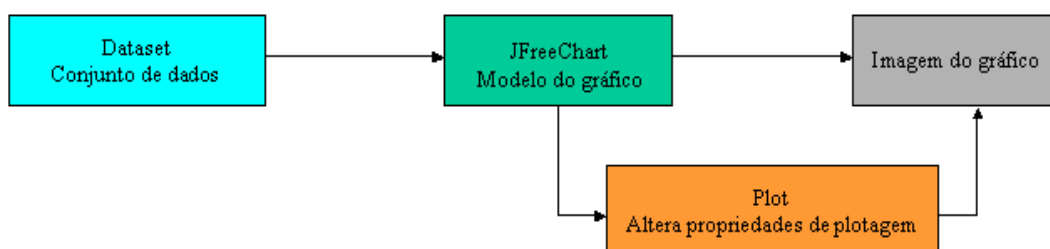
Finalmente, o Solver é utilizado na ferramenta SET. Através de interfaces gráficas, o usuário pode utilizar o Solver de maneira muito mais convidativa e simples, pois não precisa lidar com linhas de comando. Todos os parâmetros que devem ser especificados, de acordo com cada funcionalidade do Solver, são apresentados ao usuário em forma de campos de formulário.

Nesta questão, além da facilidade de uso permitida ao usuário, a ferramenta faz o cálculo de número de amostras dado um determinado arquivo (tal parâmetro é importante para o uso dos estimadores e para o uso de Autocorrelação, discutida adiante) e faz o controle de validação de campos preenchidos pelo usuário. O cálculo do parâmetro Hurst também é feito através da ferramenta e só pode ser realizado através da regressão linear do arquivo gerado para os estimadores Higuchi, Var e R/S. A regressão linear é disponibilizada pelo Solver e utilizada pelo SET. Para o caso do estimador AV baseado em Wavelets, a regressão linear deve ser ponderada e já é realizada pelo Solver quando o arquivo de saída é gerado. Esta regressão linear ponderada não é disponibilizada pelo SET.

Para que os dados gerados pelos métodos de análise possam ser analisados graficamente, a ferramenta utiliza uma biblioteca de classes Java conhecida como JFreeChart. Esta biblioteca gera gráficos, tais como: gráficos de pizza, linhas, barras, áreas, combinações, entre outros. Todo o código-fonte é disponibilizado sob a licença GPL (General Public License), então não há problemas em alterá-la para fins pessoais ou comerciais, fato importante dentro do contexto da ferramenta SET, pois não será usada apenas para fins pessoais. Como outras facilidades que são disponibilizadas, podemos contar também com: zoom, impressão, exportar para figuras JPEG e PNG, exportar para PDF (via API iText) e SVG (como o Apache Batik), suporte para servlets e JSP (através da taglib Cewolf é possível gerar gráficos dinamicamente em aplicações web) e documentação em Javadoc (documentação padrão de classes Java).

Existem duas famílias principais de gráficos no JFreeChart: de categorias e XY. Para o SET, foi utilizada apenas a categoria XY.

Para a construção de gráficos com o JFreeChart deve-se seguir alguns passos: definição de um conjunto de dados a serem plotados (Dataset); criação do modelo do gráfico; alteração de propriedades de plotagem (opcional); e geração da imagem do gráfico a partir do seu modelo.



2.6: Fluxo do pacote JFreeChart

Abaixo se pode observar um exemplo de uso da biblioteca, que pode ser encontrado no site oficial dos desenvolvedores.



2.7 Demonstração da utilização do JFreeChart

Para que o usuário tenha algumas informações adicionais a respeito dos estimadores, dos métodos auxiliares e da geração dos gráficos, a ferramenta ainda disponibiliza um tutorial, em formato PDF. Para que o usuário não precise se

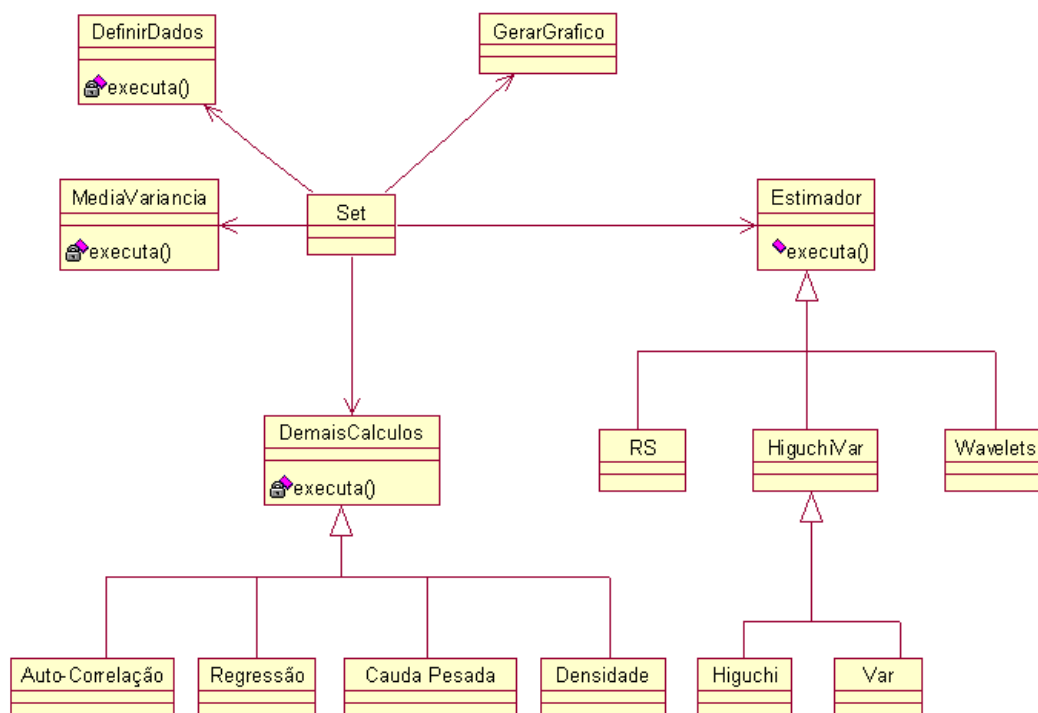
preocupar se possui algum visualizador, foi utilizada uma outra biblioteca de classes Java, conhecida como JPedal, desenvolvida e mantida por uma empresa conhecida como IDR Solutions. Além disso, foi levado em conta que o visualizador fosse independente de plataforma, pois também foi desenvolvido em Java.

Dentre os objetivos alcançados, a ferramenta ainda não permite que o usuário faça a delimitação gráfica de um intervalo de tempo para que efetue o cálculo do parâmetro Hurst. Este procedimento é importante para que seja possível fazer uma melhor estimação do mesmo, fazendo uma seleção dos pontos que irão sofrer uma regressão linear. Tal procedimento somente não é utilizado na análise de dependência temporal para o estimador Wavelets, pois “o estimador AV implementa internamente o estágio de regressão linear ponderada (não necessitando de qualquer software gráfico/matemático para tal) fornecendo diretamente o valor do parâmetro H”. [22]

É importante observar que a ferramenta SET não usa nenhum acesso a qualquer banco de dados. Conforme especificado, a ferramenta Solver, a princípio, trabalha baseado em serialização, ou seja, em manipulação de arquivos (I/O). Para cada estimador, são definidas extensões de arquivo. Por exemplo, para o estimador R/S, a extensão será um arquivo com extensão .rs, para o Higuchi a extensão .hig, para o Var a extensão .var e finalmente para o AV baseado em wavelet com extensão .wav. Para que algum estimador possa ser utilizado, previamente deve ser gerado algum arquivo através do Solver, dependendo dos requisitos especificados pelo usuário na interface gráfica. Tal funcionalidade será permitida através do SET.

Analogamente, O cálculo da média recebe a extensão .media, da variância .variance, auto correlação .ac, regressão .regr, cauda pesada .ht (heavy tail), densidade .den. Para complementar, a geração de estatísticas, recebe a extensão .pac para arquivos gerados para o número de pacotes, e .byt para o número de bytes.

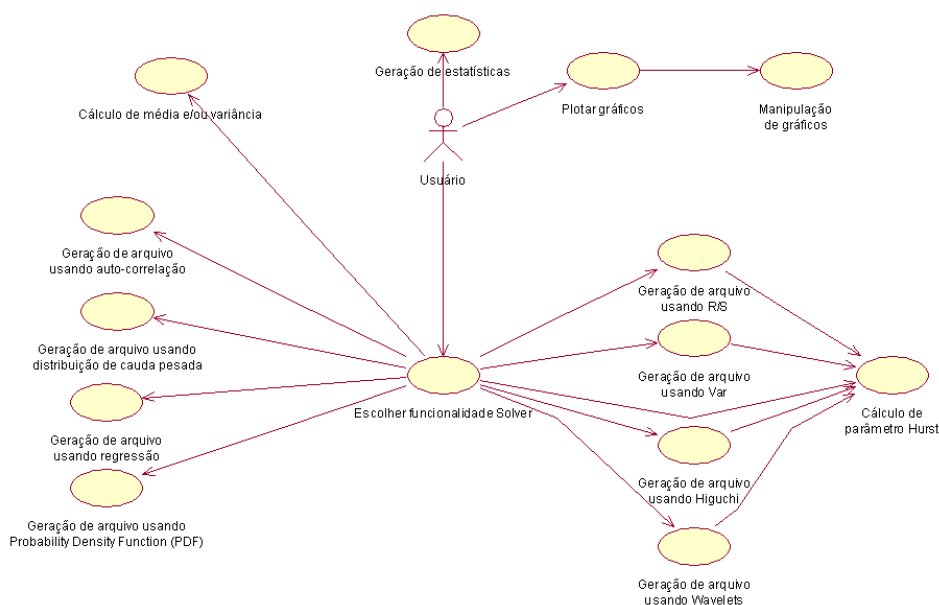
2.3.2 Diagrama de classes (modelo conceitual)



2.8 Diagrama de classes*

*Não reflete o modelo atual da ferramenta, consta como sugestão para alteração futura.

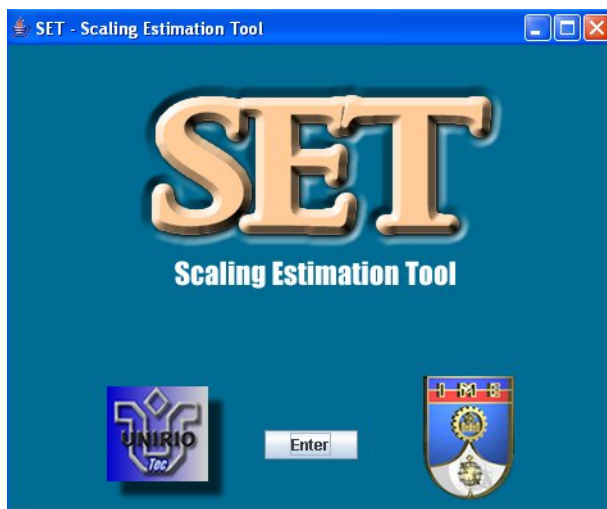
2.3.3 Diagrama de casos de uso



2.9 Diagrama de casos de uso

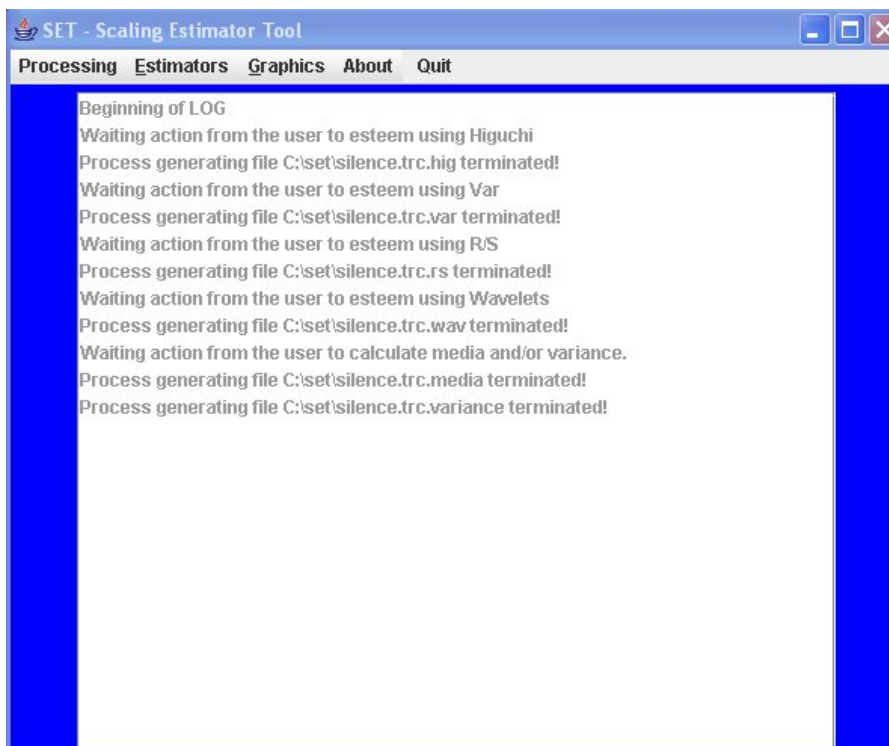
2.3.4 Descrição da interface gráfica – visão geral

Assim que o usuário entra na ferramenta, aparece uma tela inicial informando o nome da ferramenta e a parceria que foi realizada para tornar possível a criação da mesma.



2.10 Tela inicial do SET

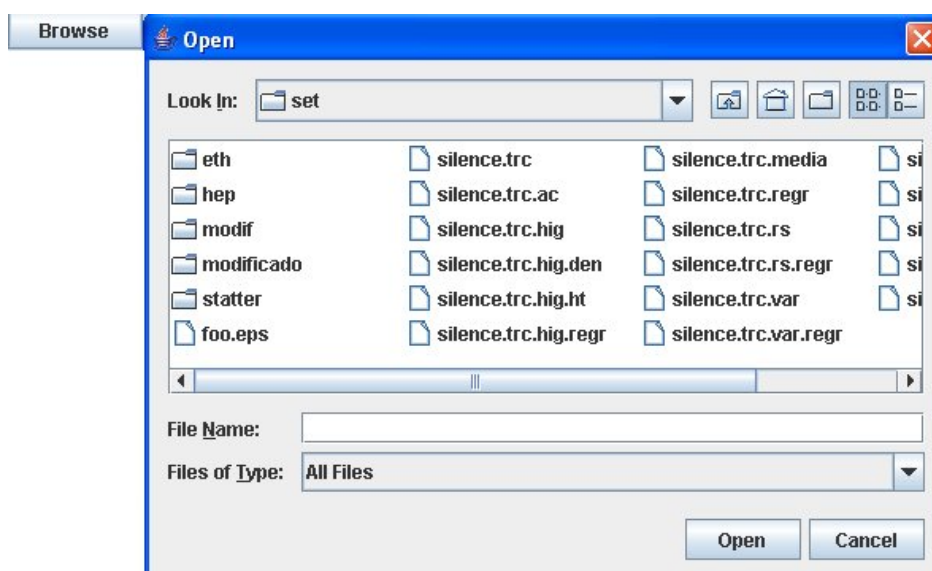
Para prosseguir, deve-se clicar com o mouse no botão apresentado ou pressionar a tecla “Enter” do teclado. Com isso, a tela principal do programa aparece, oferecendo ao usuário diversas opções de uso.



2.11 Console de mensagens

Como se observou na imagem anterior, para oferecer maior controle das ações executadas na ferramenta, foi proposto o uso de uma lista, que possui a mesma funcionalidade de um logger (uma lista que apresenta informações de processamentos realizados pelo usuário).

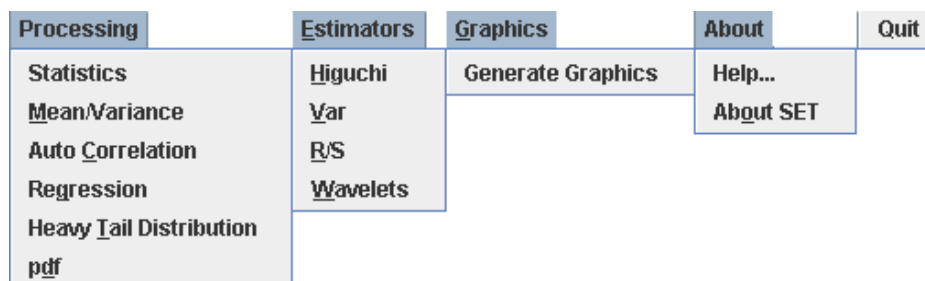
A interface gráfica, por usar o Solver de maneira transparente ao usuário, precisa também de uma outra interface que auxilie o carregamento de arquivos, pois todas as suas funcionalidades (Solver) para gerarem um arquivo de saída precisam de um arquivo de entrada. Portanto, todas estas funcionalidades possuem esta interface auxiliar. Para o mesmo tipo de dados a ser usado como arquivo de entrada, não é necessário que o usuário carregue novamente o mesmo arquivo, basta que ele faça o processamento necessário para que o arquivo correspondente seja gerado.



2.12 Seleção de arquivos

A interface gráfica gera os arquivos de saída no mesmo diretório do arquivo de entrada especificado pelo usuário. Após um debate com a professora orientadora, definiu-se que a ferramenta teria esta característica. Além disso, para cada arquivo gerado foram propostas extensões de arquivo dedutivas, para que o usuário saiba quais arquivos deseja manipular. Para isso, o nome do arquivo de saída leva o mesmo nome do arquivo de entrada, terminado pela extensão proposta.

A ferramenta, conforme suas funcionalidades, apresenta um menu de opções, organizado segundo a estrutura a seguir:



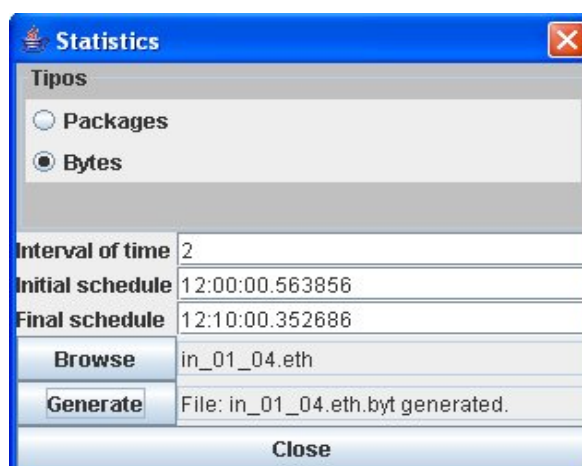
2.13 Menus da ferramenta

As opções que se referem às funcionalidades do Solver são as dos menus “Processing” e “Estimators”. Os controles de gráficos, tais como geração e manipulação, estão presentes no menu “Graphics”. Em menu “About”, o usuário pode contar com um item de ajuda (“Help...”) e outro com informações sobre o SET. Adiante, serão explicadas detalhadamente.

2.3.4.1 Processing

2.3.4.1.1 Statistics

Esta funcionalidade é responsável pela geração de estatísticas, através do Solver. Através da interface, o usuário tem o conhecimento de todos os dados que devem ser especificados para que seja possível gerar o arquivo de saída. O arquivo de entrada, a ser especificado pelo botão “Browse”, deve ser um arquivo gerado pela ferramenta Ethereal, através da coleta de dados.



2.14 Janela de geração de estatísticas

2.3.4.1.2 Mean/Variance

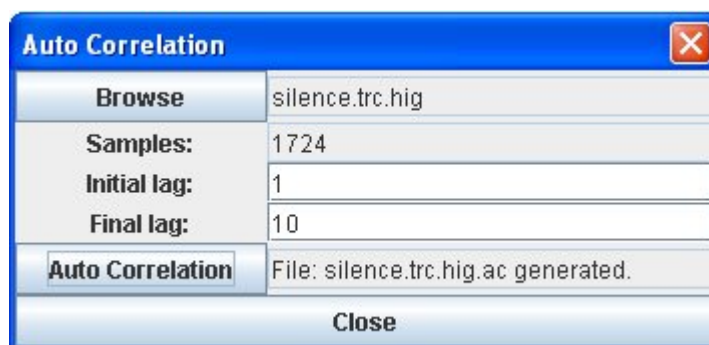
Esta funcionalidade é responsável pelo cálculo da média e/ou da variância, dado um arquivo de entrada. Os arquivos de saída são gerados, mesmo que não se precise manipulá-los, apenas os respectivos valores desejados.



2.15 Janela de cálculo da média/variância

2.3.4.1.3 Auto-Correlation

Esta funcionalidade é responsável pela geração do arquivo de saída representando a autocorrelação do arquivo de entrada.



2.16 Janela de autocorrelação

2.3.4.1.4 Regression

Esta funcionalidade é responsável pela geração do arquivo de saída representando a regressão do arquivo de entrada. O arquivo de entrada (especificado à direita do botão “Browse”) mantém-se o mesmo que o especificado pela autocorrelação, ou vice-versa, pois se tratam de arquivos de entrada que obedecem ao mesmo padrão de especificação de dados.



2.17 Janela de regressão

2.3.4.1.5 Heavy tail distribution

Esta funcionalidade é responsável pela geração do arquivo de saída representando a geração de distribuição de cauda pesada do arquivo de entrada.



2.18 Janela de cauda pesada

2.3.4.1.6 pdf (probability of density function)

Esta funcionalidade é responsável pela geração do arquivo de saída representando a geração de pdf do arquivo de entrada. O arquivo de entrada (especificado à direita do botão “Browse”) mantém-se o mesmo que o especificado pelo heavy tail distribution, ou vice-versa, pois se tratam de arquivos de entrada que obedecem ao mesmo padrão de especificação de dados.



2.19 Janela de pdf

2.3.4.2 Estimators

Todos os estimadores possuem o mesmo tipo de arquivo de entrada (seguem o mesmo padrão). Portanto, todos eles mantêm o mesmo arquivo de entrada para facilitar o uso pelo usuário.

Além disso, todos os estimadores possuem a funcionalidade de cálculo do parâmetro Hurst. Porém, para que o cálculo seja realizado, inicialmente deve ser gerado o arquivo de saída através do botão “Estimation”.

2.3.4.2.1 Higuchi

Esta funcionalidade é responsável pela geração do arquivo de saída representando o estimador Higuchi do arquivo de entrada.



2.20 Janela do estimador Higuchi

2.3.4.2.2 VAR

Esta funcionalidade é responsável pela geração do arquivo de saída representando o estimador Var do arquivo de entrada.



2.21 Janela do estimador VAR

2.3.4.2.3 R/S

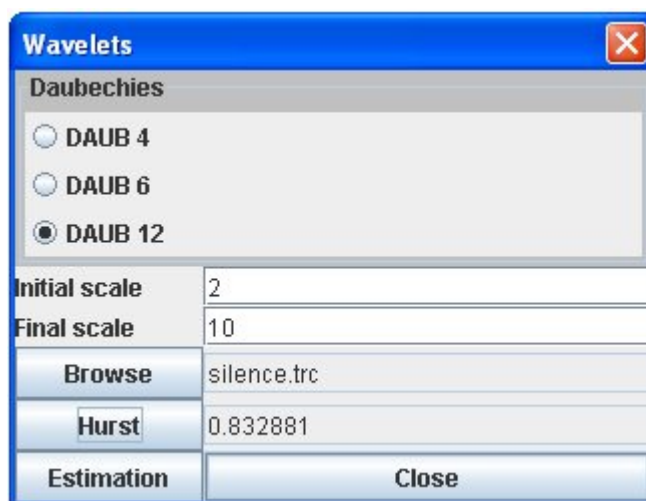
Esta funcionalidade é responsável pela geração do arquivo de saída representando o estimador R/S do arquivo de entrada.



2.22 Janela do estimador R/S

2.3.4.2.4 Wavelets

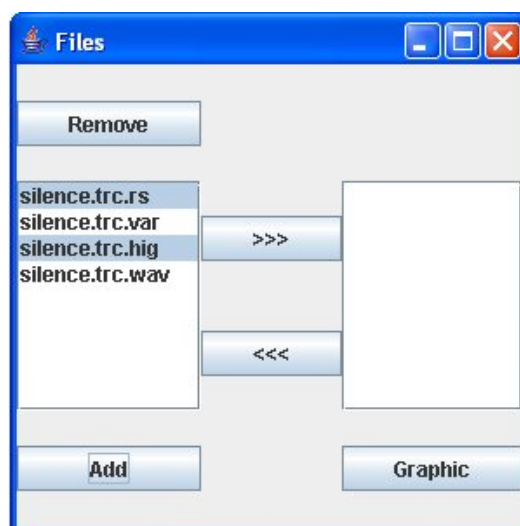
Esta funcionalidade é responsável pela geração do arquivo de saída representando o estimador R/S do arquivo de entrada.



2.23 Janela do estimador Wavelets

2.3.4.3 Graphics

Este recurso independe do solver, pois se encarrega do tratamento de arquivos gerados graficamente. Para que o usuário pudesse ter maior flexibilidade, foi criada uma interface que auxilia na manipulação destes arquivos.



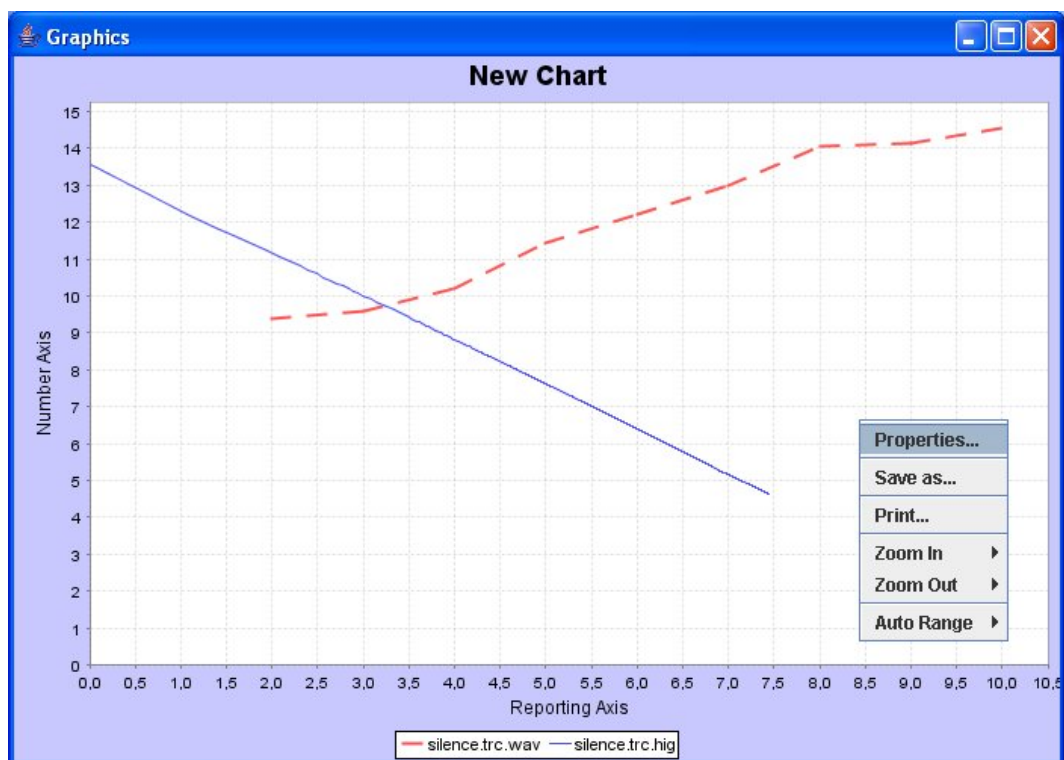
2.24 Janela de adição dos gráficos

Através desta interface, o usuário pode escolher quais arquivos deseja manipular. Para isso, deve clicar no botão “Add”, para carregar os arquivos na lista à esquerda. O gráfico a ser gerado deve conter arquivos na lista à direita, podendo-se selecionar múltiplos arquivos da lista esquerda. Caso o usuário deseje remover algum item desta lista (ou mais de um), pode fazê-lo selecionando os arquivos desejados da lista à esquerda e pressionando-se o botão “Remove”. Para a geração do gráfico, deve-se pressionar o botão “>>>”, que representa o fluxo dos arquivos de uma lista à outra, neste caso da lista esquerda para a direita e, posteriormente, o botão “Graphic”. Caso se queira remover os itens da lista à direita, o usuário pode selecionar os arquivos desejados e clicar em “<<<”, que representa o fluxo dos arquivos da lista direita para a lista esquerda. Conclui-se que o gráfico a ser gerado baseia-se nos itens que estão na lista à direita. Além disso, pode-se selecionar múltiplos arquivos também na interface de carregamento de arquivo, como se pode observar na figura adiante.



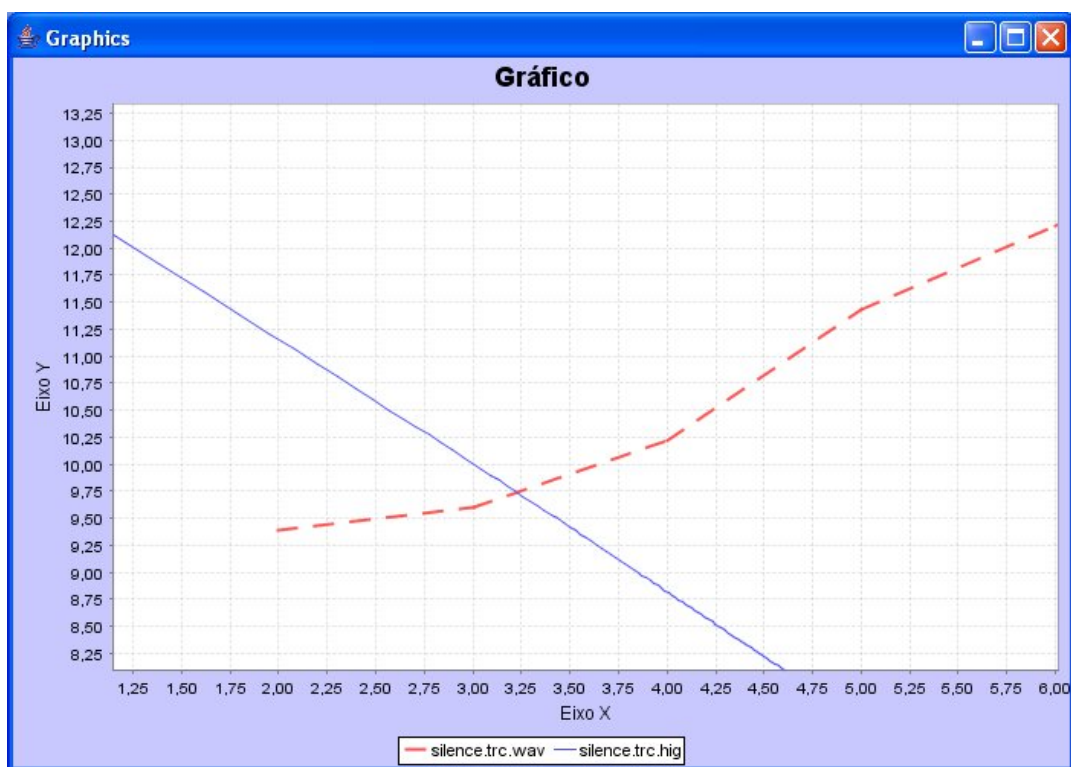
2.25 Janela de seleção dos arquivos/gráficos

Adiante, um exemplo de um gráfico gerado. Clicando-se com o botão direito do mouse, pode-se alterar diversas propriedades do gráfico, assim como salvar como uma figura PNG, imprimir e opções de zoom (Auto Range também está incluído nas características de zoom).



2.26 Janela dos gráficos

Através do item “Properties...” pode-se alterar todas as propriedades visuais do gráfico gerado, tais como: título, fontes, cores, intervalos de área visualizada (eixos X e Y, além de seus respectivos nomes). Adiante pode-se observar um exemplo:

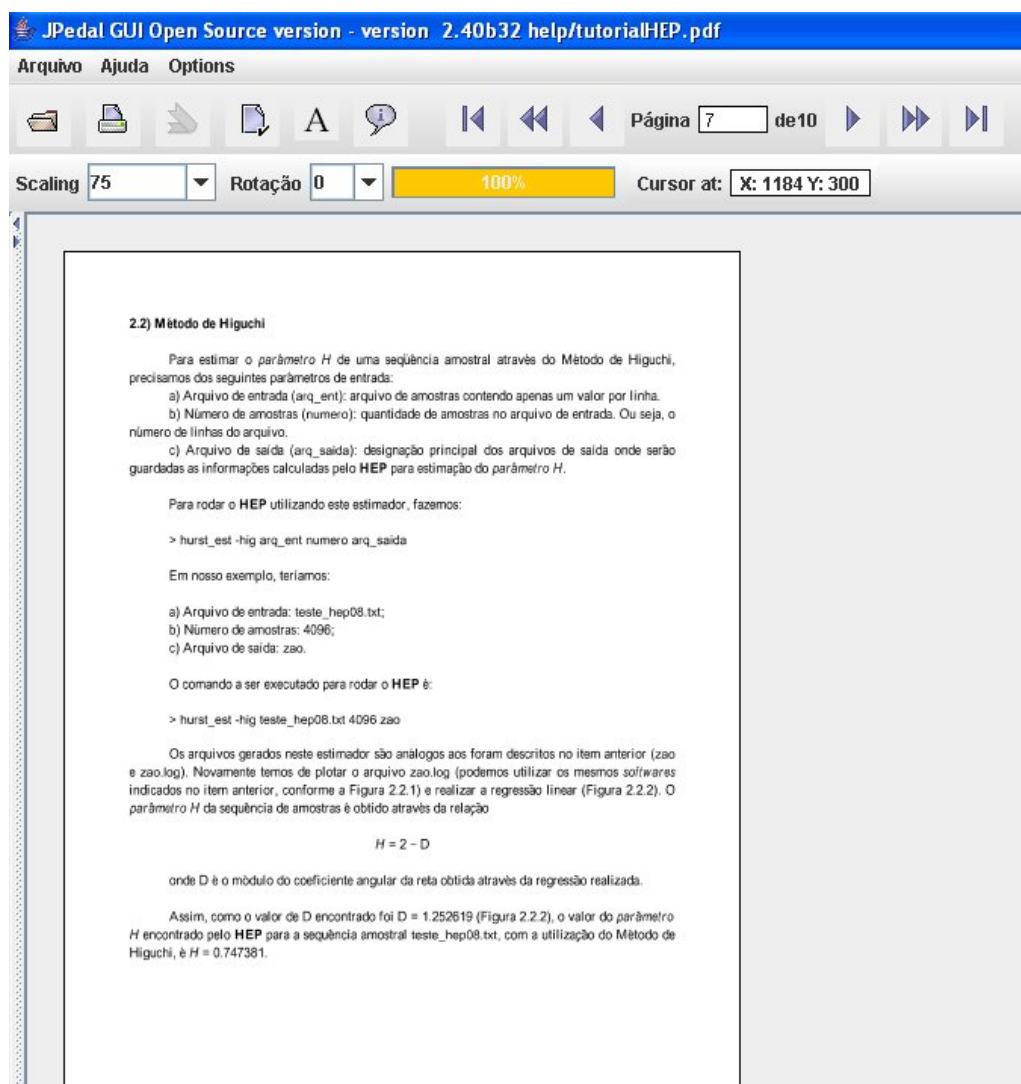


2.27 Zoom de um gráfico

2.3.4.4 About

2.3.4.4.1 Help...

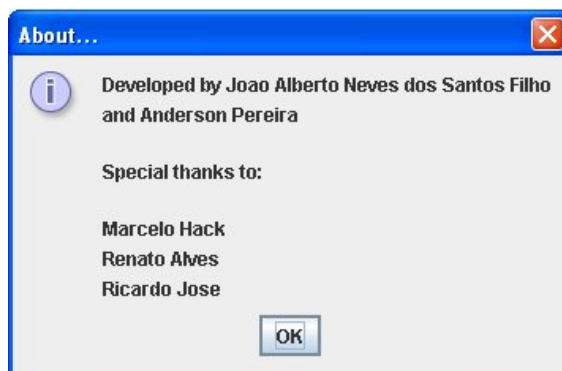
Aqui, como exemplo, foi disponibilizado um tutorial ao usuário, em formato pdf. Futuramente, o tutorial será substituído por outro mais adequado à ferramenta, porém a funcionalidade já está implementada.



2.28 Help da ferramenta

2.3.4.4.2 About SET

Este item não implementa nenhuma funcionalidade, apenas se trata dos desenvolvedores do SET e agradecimentos.



2.29 Sobre os autores

2.3.4.5 Quit

O usuário pode fechar a ferramenta por este item. Aparece uma caixa de diálogo confirmando a operação. Caso o usuário não deseje sair, basta responder à opção “No” ou “Cancel”.



2.30 Janela de saída

Capítulo 3 – Estudo de Casos

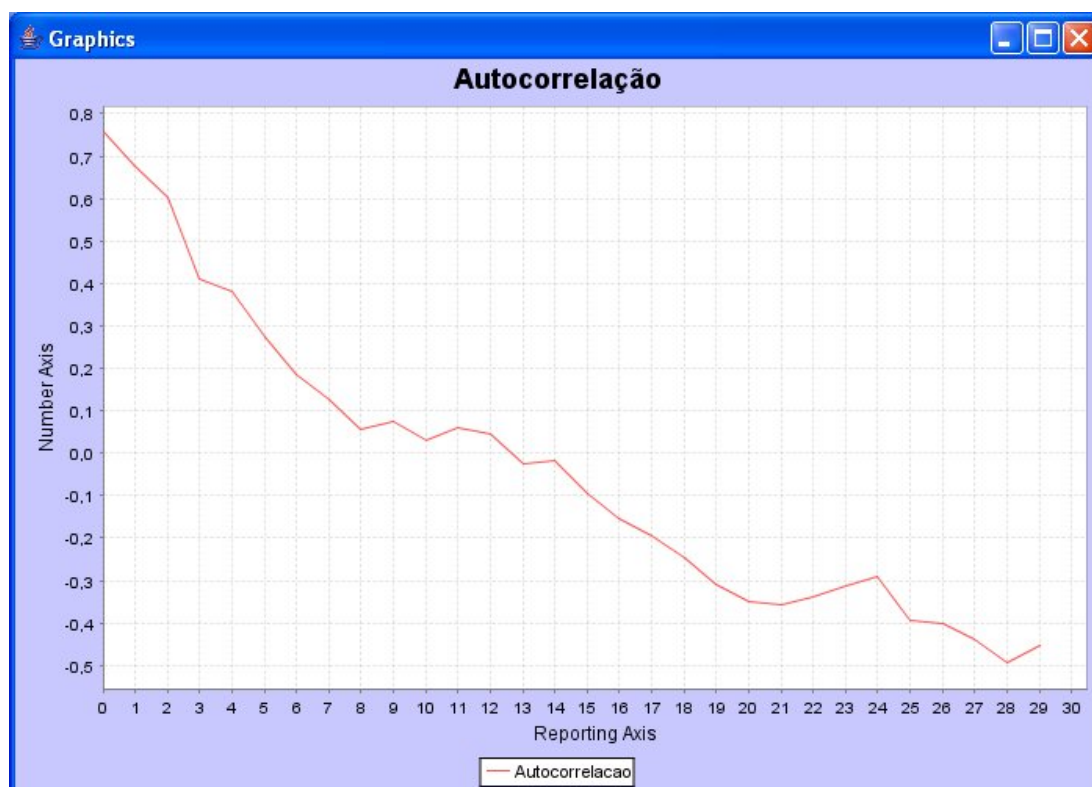
3.1 Banda de rede da UNIRIO

Para verificar a potencialidade da ferramenta desenvolvida neste trabalho, é feito nesta seção o estudo do tráfego de dados da UNIRIO. Um dia foi selecionado para servir de base para o estudo de caso e aleatoriamente destacamos o dia 06/05/2005 (Seis de Maio). Utilizamos um horário de pico de tráfego (19:00:00 → 20:00:00) para afirmar a utilidade da ferramenta respondendo mais à frente uma questão que pode ser comprovada: “A banda de rede atual é suficiente?”

3.2 Gráficos de amostras

Esta seção apresenta os gráficos gerados pela ferramenta SET. Os gráficos foram gerados a partir dos dados coletados no tráfego da UNIRIO (nas ilustrações: dia 6 de Maio de 2005, no intervalo entre 19h e 20h)

3.2.1 Autocorrelação



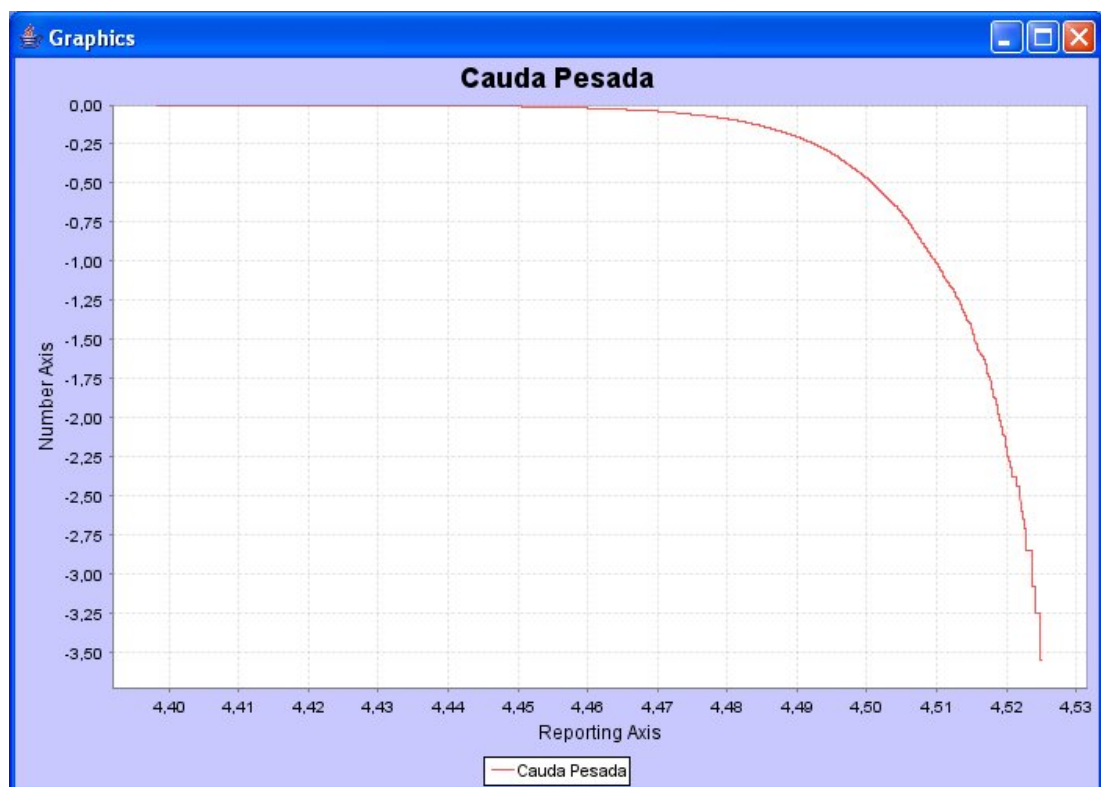
3.1 Gráfico de autocorrelação

3.2.2 Regressão



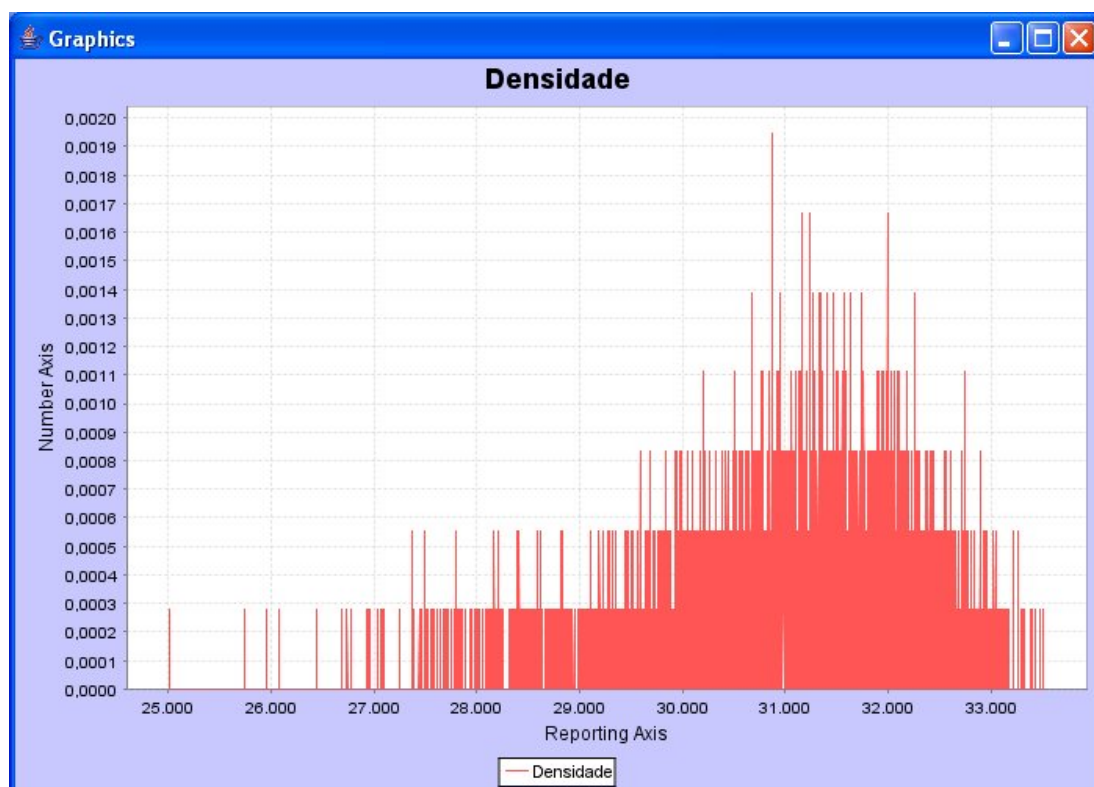
3.2 Gráfico de regressão

3.2.3 Cauda Pesada



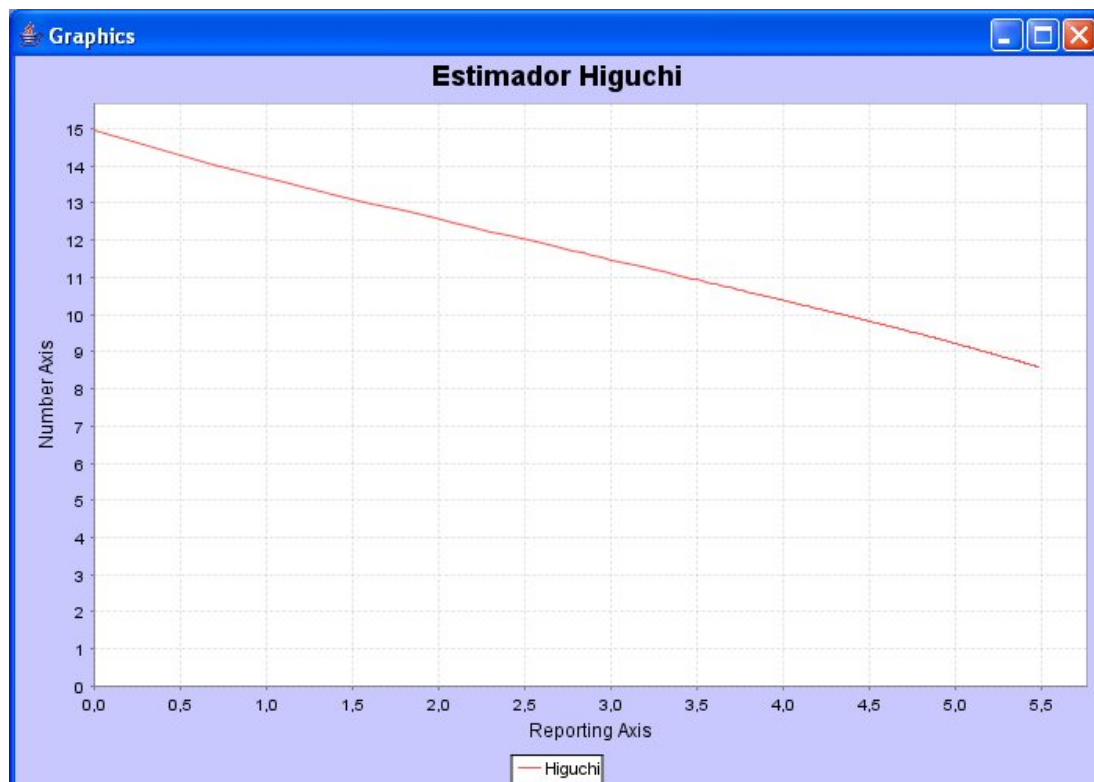
3.3 Gráfico de cauda pesada

3.2.4 Densidade



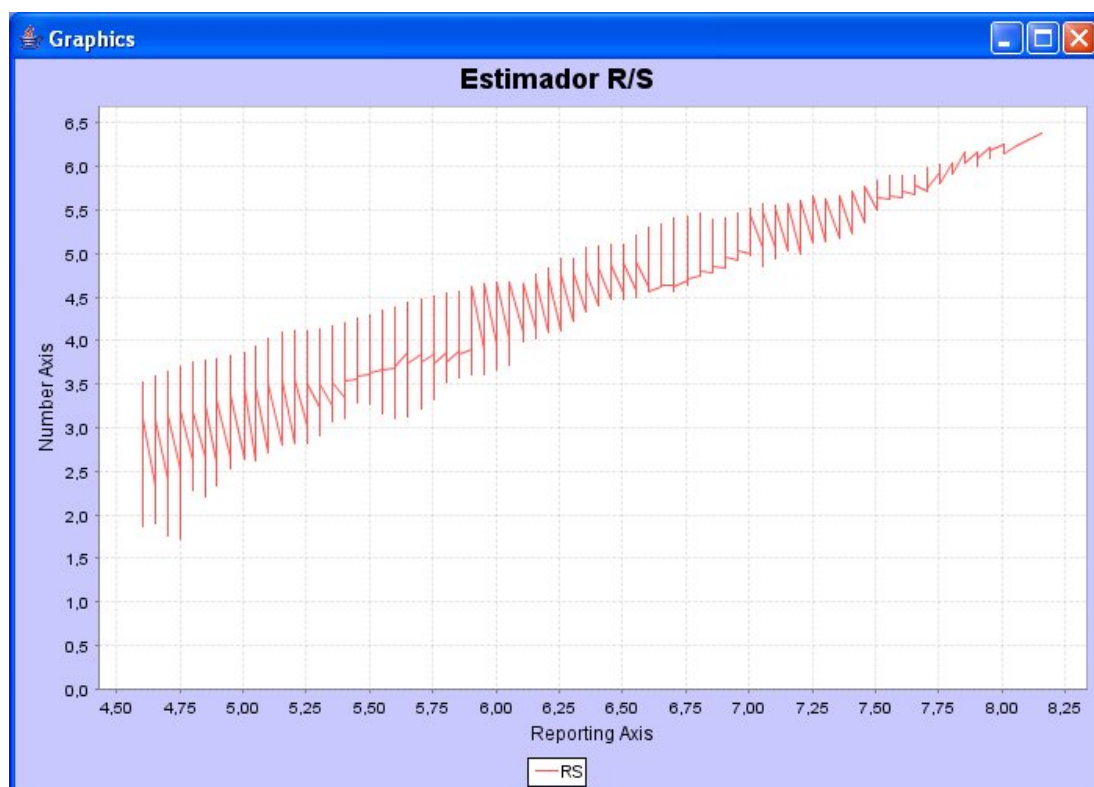
3.4 Gráfico de densidade

3.2.5 Estimador Higuchi



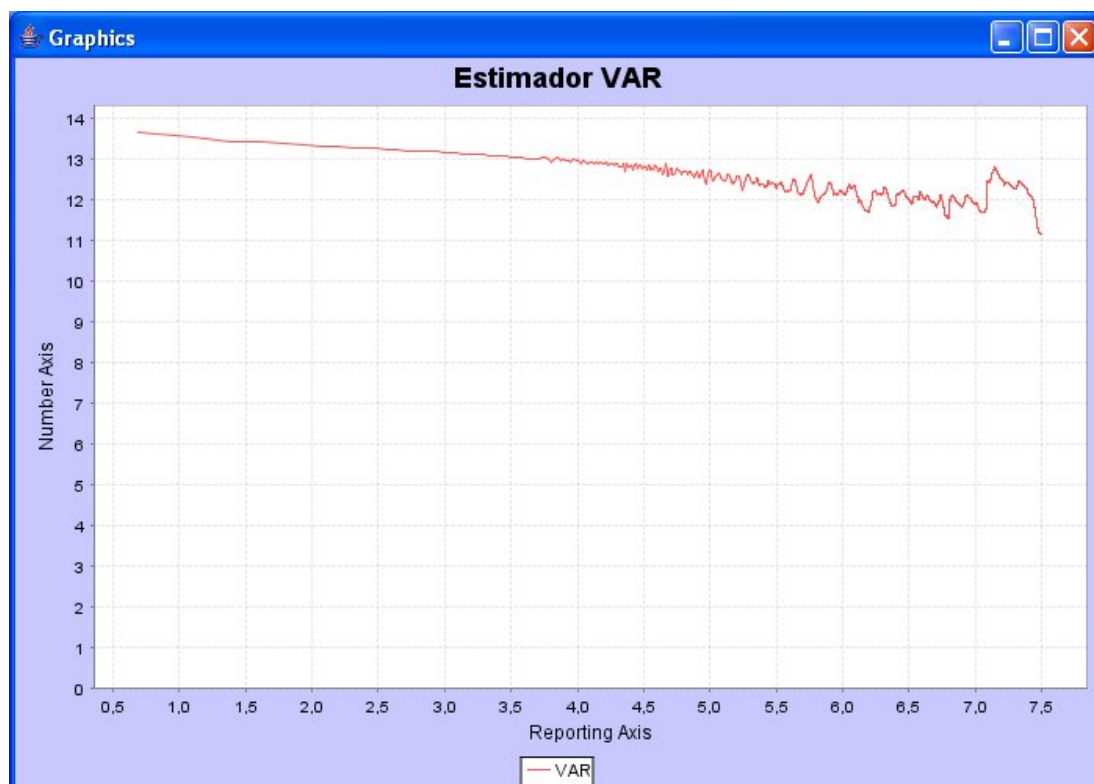
3.5 Gráfico de Higuchi

3.2.6 Estimador R/S



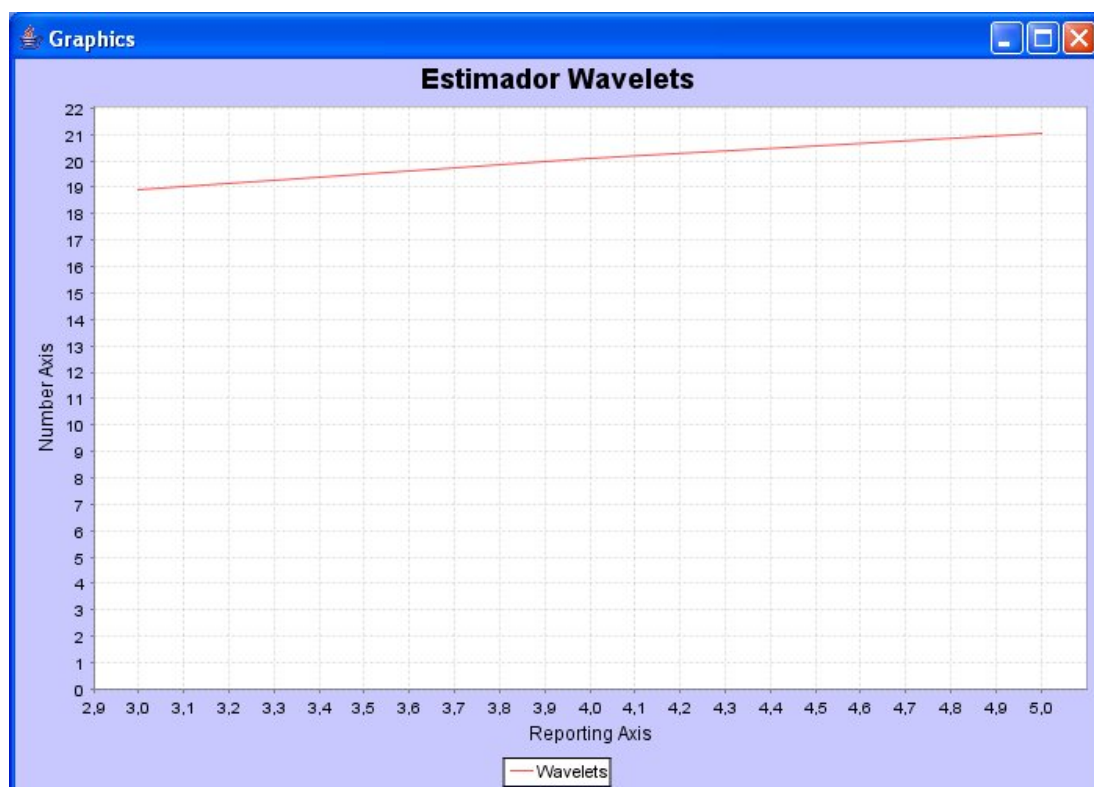
3.6 Gráfico de R/S

3.2.7 Estimador VAR



3.7 Gráfico de VAR

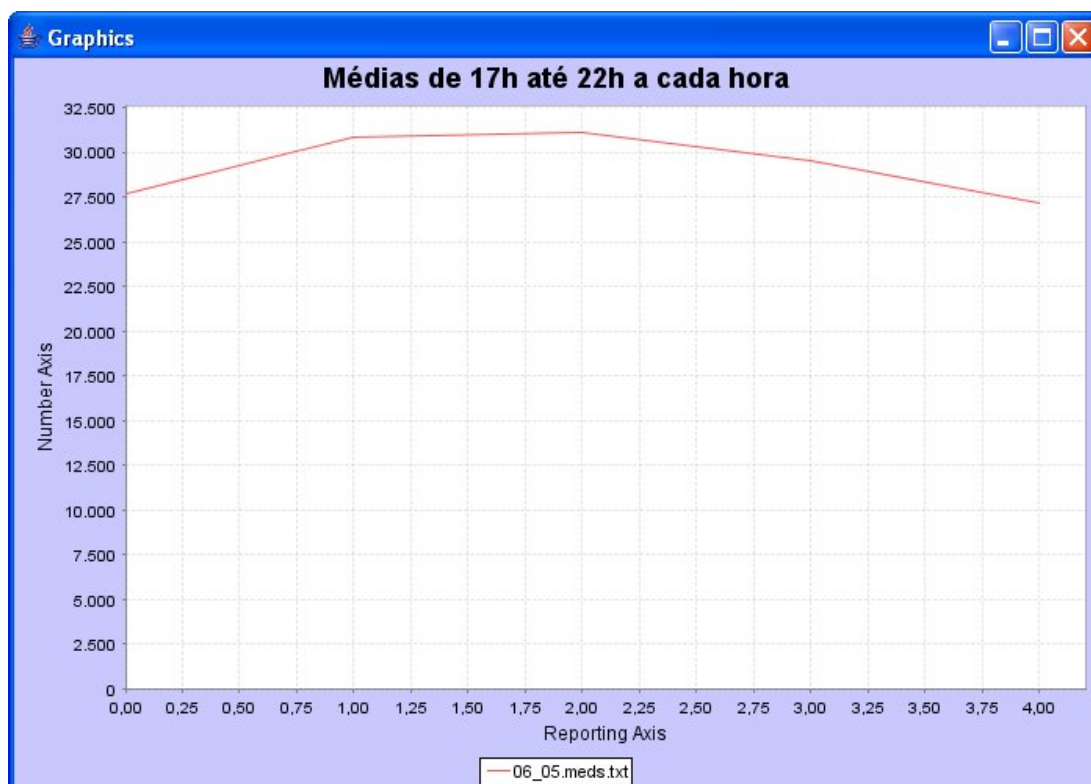
3.2.8 Estimador Wavelets



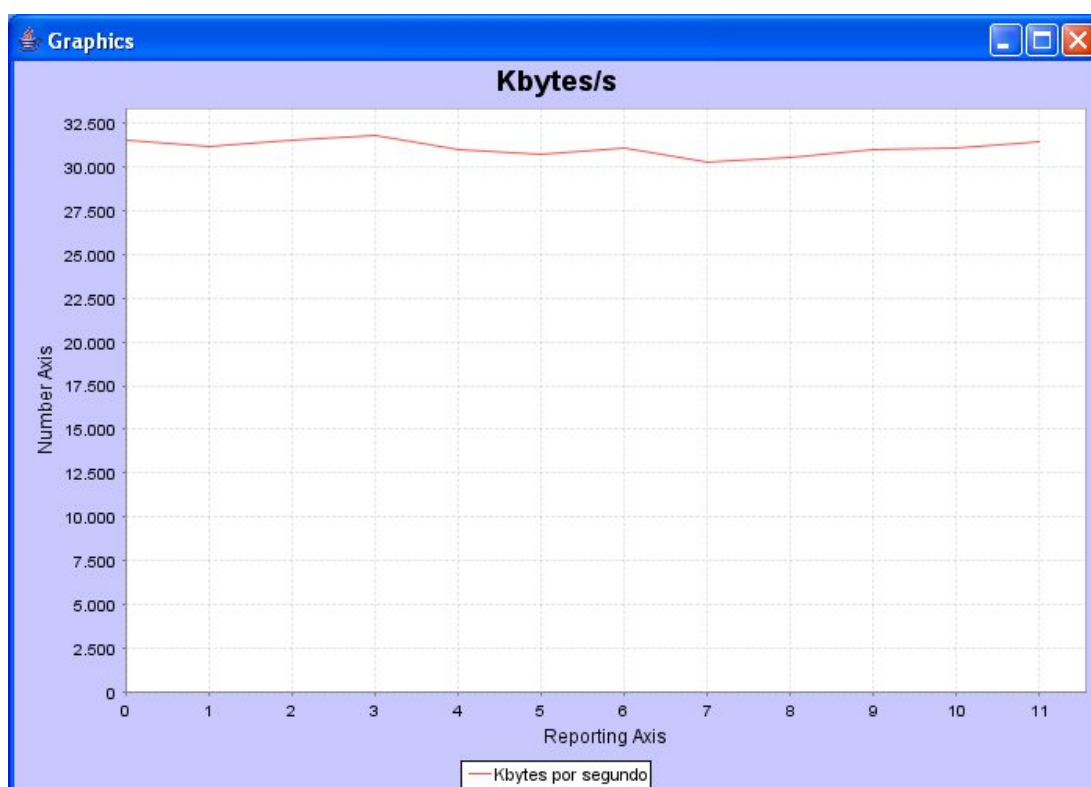
3.8 Gráfico de Wavelets

3.3 Resultado do estudo de caso

Considerando que a banda atual da UNIRIO possui 256 kbits/s ($256/8 = 32$ kbytes/s) e os gráficos abaixo da utilização no dia citado anteriormente, podemos afirmar que, pela média calculada (na ferramenta SET) nas ilustrações abaixo de 31,067 kbytes/s ser muito próxima ao total permitido pela banda, que esse tráfego é intenso demais e assim pode-se considerar esta banda de 256 kbits/s insuficiente para seu uso correto na UNIRIO.



3.9 Médias geradas dos kbytes a cada 1 hora entre 17:00 e 22:00



3.10 Médias geradas dos kbytes a cada 5 minutos entre 19:00 e 20:00

Capítulo 4 – Conclusão

A divisão da ferramenta em linguagens de programação distintas teve como objetivo aproveitar o maior número de recursos que cada linguagem disponibiliza e maior facilidade em conversão. Estão presentes em C++ todos os algoritmos de processamento interno da ferramenta. Esta linguagem de programação foi escolhida pois alguns algoritmos de estimadores já se encontravam em C++, pela maior facilidade de converter os demais algoritmos de C para C++ e pela robustez de desempenho que a linguagem Java não pôde oferecer. Para que a agilidade de processamento fosse maior ainda, tudo foi feito baseado em linha de comando. Em Java foi dado destaque as interfaces, pois a linguagem possui um maior número de pacotes que dão suporte a interfaces. Então, seus objetivos foram prover uma melhor usabilidade pelo usuário e facilitar a análise de desempenho de rede, através de gráficos. O uso de projetos extras, integrados à ferramenta foi fundamental, pois tornam a ferramenta mais profissional.

A ferramenta SET já pode ser utilizada, mas ainda não está completa. Por motivos de tempo e necessidade de maior aprofundamento, ela continuará sendo desenvolvida por outros alunos da própria UNIRIO, sendo possível tornar a ferramenta mais completa. Além disso, serão analisadas novas funcionalidades que futuramente poderão ser realizadas através do SET.

Referências

- [1] J.F. Kurose e K.W. Ross. *Redes de Computadores e a Internet*. Makron Books. 2003.
- [2] L. Kelenrock. *Queueing Systems. Volume 1: Theory*. A Wiley-Interscience Publication. John Wiley & Sons. 1975.
- [3] W.E. Leland, M.S. Taqqu, W. Willinger and D. Wilson. On the Self-Similar Nature of Ethernet Traffic. *IEEE/ACM Trans. On Networking*, v.2, n.1, pag. 1-15. February 1994.
- [5] K. Park e W. Willinger. *Self-Similar Network Traffic and Performance Evaluation*. John Wiley & Sons, Inc. 2000.
- [6] A. Iyengar, M. S. Squillante e L. Zhang. Analysis and Characterization of Large-Scale Web Server Access Patterns and Performance. *World Wide Web*, V. 2, N. 1-2, pags. 85-100, 1999.
- [7] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E. Kolaczyk e N. Taft. Structural analysis of network traffic flows. In: *Proc. ACM SIGMETRICS*. June 2004.
- [8] J. L. Jerkins et al. Operations Measurements for Engineering Support of High-Speed Networks with Self-Similar Traffic. In: *Proc. Of 16th ITC*, V. 3A, pag. 895-906, UK, June 1999.
- [9] A. Guimarães e R. Coelho. Video Traffic Models Performance with Restricted and Non-Restricted Scaling Characterization. In: *Proc. of the International Telecommunications Symposium*, pag. 646-650. September 2002.
- [10] R. Pontes e R. Coelho. The scalling characteristics of the vídeo traffic and its impact on acceptance regions. *Proceedings of the 17th ITC*, v. 4. September 2001.
- [11] R. Pontes e R. Coelho. Admission control for vídeo traffic streams with scaling characteristics. *Revista da Sociedade Brasileira de Telecomunicações*. V. 17, N. 2. Dezembro de 2002.
- [12] H.E. Hurst. Long-term storage capacity of reservoirs. *Transactions of American Society of Civil Engineers*, 116:770-799, 1951.
- [13] M. Barnsley. *The Science of Fractal Images*. Springer-Verlag Inc., New York, USA, 1988.
- [14] P. Abry e D. Veitch. Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory*, 44(1):2-15, 1998.
- [15] H.E. Hurst. Long-term storage capacity of reservoirs. *Transactions of American Society of Civil Engineers*, 116:770-799, 1951.

- [16] M. Barnsley. The Science of Fractal Images. Springer-Verlag Inc., New York, USA, 1988.
- [17] P. Abry e D. Veitch. Wavelet analysis of long-range dependent traffic. IEEE Transactions on Information Theory, 44(1):2-15,1998.
- [18] Doc. Ethereal: <http://www.ethereal.com/distribution/docs/user-guide.pdf>
- [19] Doc. Infra-estrutura de rede da UNIRIOTEC
- [20] Site do projeto JFreeChart: <http://www.jfree.org/jfreechart/>
- [21] Site do projeto JPedal: <http://www.jpedal.org/>
- [22] Zão, Leonardo. Tutorial de uso do HEP. Larso - Laboratório de redes de comunicação e sistemas ópticos

ANEXO I

SINTAXES DO SOLVER

Notações:	<conteúdo>	→	preenchimento obrigatório
	[conteúdo]	→	preenchimento opcional
	(X,Y,W,Z)	→	um ou mais do conjunto

Sintaxe solver: solver <Nº processo> <arquivo de entrada> [parâmetros]

Onde os processos podem ser:

1) Geração de Estatísticas

Sintaxe adicional: <arquivo de saída> [-t segundos] [-i horário inicial] [-f
horário final] [-C (ini,fim,byt,pac)]

Objetivo: Processar os dados de entrada coletados pelo Ethereum e, dependendo do pedido passado à linha de comando, gerar uma lista de resultados que posteriormente será utilizada como entrada para algoritmos complexos de medição.

[-t segundos] → Especifica com quantos segundos (ou fração de segundos) se
forma um “grupo” de pacotes ou bytes
(Padrão: 1 segundo)
(Ex: -t 1.55)

[-i horário inicial] → Especifica o horário exato de início dos cálculos
(Padrão: início do arquivo de log)
(Ex: -i 12:23:21.000000)

[-f horário final] → Especifica o horário exato de fim dos cálculos
(Padrão: fim do arquivo de log)
(Ex: -f 12:57:34.023021)

[-C (ini,fim,byt,pac)] → Escolhe quais colunas devem ser mostradas na saída.
Podem ser escolhidas uma ou mais, sempre separadas por
vírgulas e nenhum espaço.

(Padrão: todas)

(Ex: -C byt,pac)

Exemplo Gerador: `solver 1 log_entrada arq_saida -i 12:00:00.000000 -f
14:30:00.000000 -C byt -t 10`

Grava no arquivo “arq_saída” uma lista de bytes somados de cada 10 segundos do horário entre 12:00 e 14:30 do trace “log_entrada” anteriormente coletado.

2) Média (output padrão)

Sintaxe adicional: [inicio] [final]

Objetivo: Retorna para a saída padrão o resultado da média dos valores contidos no arquivo de entrada.

Exemplo: `solver 2 log_entrada`

3) Variância (output padrão)

Sintaxe adicional: [inicio] [final]

Objetivo: Retorna para a saída padrão o resultado da variância dos valores contidos no arquivo de entrada.

Exemplo: `solver 3 log_entrada`

4) Média (output arquivo)

Sintaxe adicional: <arquivo de saída> [inicio] [final]

Objetivo: Salva no arquivo de saída o resultado da média dos valores contidos no arquivo de entrada.

[inicio] → Linha onde deve se começar a leitura do arquivo de entrada

(Padrão: primeira linha)

[final] → Linha onde deve se terminar a leitura do arquivo de entrada

(Padrão: última linha)

Exemplo: `solver 4 log_entrada media_log`

5) Variância (output arquivo)

Sintaxe adicional: <arquivo de saída> [inicio] [final]

Objetivo: Salva no arquivo de saída o resultado da variância dos valores contidos no arquivo de entrada.

[inicio] → Linha onde deve se começar a leitura do arquivo de entrada

(Padrão: primeira linha)

[final] → Linha onde deve se terminar a leitura do arquivo de entrada

(Padrão: última linha)

Exemplo: `solver 5 log_entrada variancia_log`

6) Autocorrelação

Sintaxe adicional: `<arquivo de saída> [lag inicial] [lag final]`

Objetivo: é uma medida de dependência entre observações da mesma série separadas por um determinado intervalo chamado retardo.

[lag inicial] → Valor inicial do lag.

[lag final] → Valor final do lag.

Exemplo: `solver 6 log_entrada result_saída 1 10`

Obs: A `lag_final` deve sempre ser menor que o número de amostras.

7) Regressão

Sintaxe adicional: `<arquivo de saída>`

Objetivo: é uma metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativas (ou qualitativas) de tal forma que uma variável pode ser predita a partir da outra ou outras. Exemplo: `solver 7 log_entrada result_saída`

8) Cauda Pesada

Sintaxe adicional: `<arquivo de saída>`

Objetivo: é caracterizado pela ocorrência de eventos pouco frequentes, mas de grande impacto, como uma rajada de sessões se iniciando em um curto período de tempo.

Exemplo: `solver 8 log_entrada result_saída`

9) Densidade

Sintaxe adicional: `<arquivo de saída>`

Objetivo: representa uma distribuição da probabilidade nos termos das funções integrais.

Exemplo: solver 9 log_entrada result_saída

10) Estimador VAR

Sintaxe adicional: <arquivo de saída> <amostras>

Objetivo: permite testar as limitações lineares na presença de mudanças estruturais. É pouco utilizada, pois constatou-se que é muito imprecisa.

<amostras> → Número de linhas do arquivo de entrada a serem lidas.

Exemplo: solver 10 log_entrada result_saída 10

11) Estimador HIG

Sintaxe adicional: <arquivo de saída> <amostras>

Objetivo: medir a dimensão fractal de uma série de tempo não-periódica e irregular.

<amostras> → Número de linhas do arquivo de entrada a serem lidas.

Exemplo: solver 11 log_entrada result_saída 20

12) Estimador RS

Sintaxe adicional: <arquivo de saída> <tamanho do bloco> <amostras>

Objetivo: é a escala de somas parciais dos desvios da média, re-escaladas por seu desvio padrão.

<tamanho do bloco> → Tamanho do bloco usado pelo estimador.

<amostras> → Número de linhas do arquivo de entrada a serem lidas.

Exemplo: solver 12 log_entrada result_saída 1000 15

13) Estimador WAV

Sintaxe adicional: <arquivo de saída> <(1,2,3)> <amostras> <escala inicial>
<escala final>

Objetivo: fatia dados ou funções ou operadores em componentes frequenciais diferentes, e então estuda cada componente com uma resolução casada com sua escala.

<(1,2,3)> → Filtro usado pelo estimador (respectivamente DAUB4, DAUB6 e DAUB12)

<amostras> → Número de linhas do arquivo de entrada a serem lidas.

<escala inicial> → Escala inicial utilizada pelo estimador.

<escala final> → Escala final utilizada pelo estimador.

Exemplo: solver 13 log_entrada result_saída 1 80 10 20