Análise Estatística do Trafego de E-Mail para o Projeto de um Cluster de Alto Desempenho

Antonio Carlos Oliveira Amorim
Instituto de Pesquisas Tecnológicas - IPT
amorim@ipt.br

Daniel Kobayashi Imori
Bastion Systems
daniel@bastion.com.br

Sérgio Takeo Kofuji
Laboratório de Sistemas Integráveis – PAD/LSI/POLI/USP
kofuji@poli.usp.br

Vidal Zapparoli Melo
Laboratório de Sistemas Integráveis – PAD/LSI/POLI/USP
vidal@bastion.com.br

Abstract

This paper shows an e-mail statistical analysis and the behavior of its users based on real servers utilization. The work is related to an specification project, under development in Institute for Technological Research (IPT-Brasil), for a high scalability and availability e-mail service (above 100.000 users).

The statistical results are a important step to the system performance evaluation, allowing an implementation of realistic workload simulations.

The data collected are related in three categories: the amount of messages per user, the time interval between messages and the size of each message. The time intervals and the size of messages are calculate in accordance to a model thats use probability density functions fitting the real data, scaled to desired quantity of users.

To carry through the simulations, Java applications will be implemented to generate messages with the results that best fits the realistic workload.

The work intends to simulate the requisitions generated by the users, to measure the system capacity according the servers quantity and performance, to characterize servers utilization patterns and users behavior in order to determine the restrictive performance factors.

1. Introdução

Este artigo apresenta uma análise estatística da utilização de servidores de e-mail e do comportamento de seus usuários, obtida a partir de logs de servidores reais em produção.

O trabalho foi desenvolvido no contexto de um projeto de cluster de servidores de e-mail com balanceamento de carga e tolerância a falhas, que está sendo desenvolvido no IPT – Instituto de Pesquisas Tecnológicas. O cluster está sendo planejado para suportar uma grande quantidade de usuários (100.000 ou mais) com características de alta escalabilidade e disponibilidade, baseado em servidores baratos de arquitetura PC utilizando software livre.

A análise estatística é uma etapa preliminar aos testes de desempenho do sistema e permite a implementação de simuladores de carga de trabalho realista. A programação das simulações é baseada em Java, permitindo a geração de mensagens em intervalos aleatórios, de tamanho aleatório e de acordo com as funções de densidade de probabilidade ajustadas aos dados reais. O trabalho continuará com o objetivo de simular as requisições geradas pelos usuários, medir a capacidade do sistema de acordo com a quantidade e performance dos servidores, caracterizar padrões de utilização de servidores e o comportamento de usuários, determinando os fatores limitantes de desempenho.

2. O Projeto

O projeto E-Gov consiste em um cluster de servidores que fornecem os serviços de e-mail com protocolos SMTP, POP3 e IMAP4, com objetivos similares a clusters como NinjaMail e Porcupine[8,9]. Os serviços podem ser acessados por um software cliente de e-mail comum (MUA) ou através de um Webmail. A solução engloba proteção contra spam e vírus, firewall, software para balanceamento de carga e tolerância a falhas.

O projeto está em fase de desenvolvimento em um laboratório do IPT – Instituto de Pesquisas Tecnológicas, com a colaboração de pesquisadores do grupo de computação pervasiva e alto desempenho do Laboratório de Sistemas Integráveis – PAD/LSI/ POLI/USP. Os desafios iniciais já foram suplantados: instalação de hardware e software no ambiente do

laboratório e testes exaustivos das funcionalidades básicas.

3. Software

A solução é totalmente implementada com software livre, utilizando o sistema operacional Debian Linux 3.0r2 (woody).

O software LVS é usado para o balanceamento de carga entre os servidores. A instalação do LVS exige a substituição do kernel original do Debian por outro compilado de modo a habilitar o IPVS (Internet Protocol Virtual Server). O IPVS funciona pela modificação da pilha de protocolos TCP/IP em um computador que é responsável pelo balanceamento de carga (LVS-Director). As requisições são encaminhadas a um endereço virtual e distribuídas a vários servidores reais (RealServers), de acordo com o algoritmo de escalonamento selecionado [1].

O software Heartbeat [6] é utilizado para implementar tolerância a falhas nos pontos mais críticos, como o LVS-Director , que são implantados com dois servidores. Se um dos servidores deixar de funcionar, o Heartbeat encarrega-se de ativar o outro.

O servidor de e-mail utilizado é o Qmail, associado com o Open-Ldap (que gerencia os dados dos usuários) e o Courier-mail (responsável pelo controle do protocolo IMAP4).

O sistema disponibiliza, também, as funcionalidades de Webmail, proteção anti-spam e proteção antivírus, implementadas respectivamente com o SquirrelMail, Spamassassin e o Clamav.

Em cada RealServer são instalados todos os softwares necessários aos serviços de e-mail e webmail: Qmail, Open-Ldap, Courier-IMAP, SquirrelMail, Clamav e SpamAssassin.

4. Arquitetura

A arquitetura LVS-DR (Diret-Routing) é indicada, pelos autores do LVS como a de maior escalabilidade [1]. Nessa arquitetura, o LVS-director e os RealServers devem estar em um mesmo seguimento físico de rede, mas o LVS-Director deve ser conectado à Internet por uma rede diferente através de outra placa de rede.

As requisições chegam ao LVS-Director pela Internet e são distribuídas - para a rede interna - trocando o MAC-address de destino dos pacotes para o MAC-address do RealServer selecionado pelo algoritmo de escalonamento.

Os RealServers são implementados com três placas de rede, uma ligada à rede interna do LVS-Director, pela qual recebe as requisições encaminhadas pelo balanceador de carga, outra ligada ao gateway, que permite encaminhar as respostas diretamente para a Internet, e uma terceira à rede do storage (figura 1).

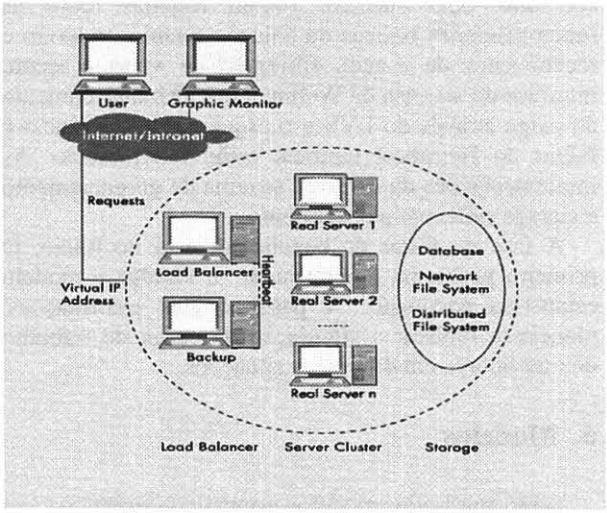

Figura 1: Arquitetura LVS-DR [1]

O servidor LDAP pode ser usado pelo QMail diretamente no mesmo equipamento ou através do direcionamento das requisições para um servidor remoto. A opção final será decidida após os testes de desempenho.

O serviço de armazenamento (storage), que contem as caixas postais, está implementado atualmente como um simples servidor NFS. Soluções mais apropriadas estão em estudo.

A solução adotada pode ser escalada de uma maneira muito flexível. Conforme aumenta a quantidade de usuários podem ser acrescentados mais Realservers. Segundo os desenvolvedores do LVS, na arquitetura LVS-DR o Linux-Director suporta até 2 milhões de requisições simultâneas com 256 MB de RAM livre [1], e dificilmente se tornará um gargalo.

As tecnologias de redes usadas para acesso à internet e internamente ao cluster, a tecnologia de storage e a capacidade de processamento dos RealServers e da Firewall devem ser os maiores limitadores de desempenho. Os fatores limitantes e as melhores configurações de software e hardware serão analisados com benchmarks, descritos adiante.

5. Testes

Os testes da solução são divididos em duas fases, tendo a primeira o objetivo de testar as funcionalidades e a segunda, caracterizar a capacidade e a escalabilidade do sistema com benchmarks. A capacidade será avaliada pela quantidade de usuários e mensagens que o sistema pode suportar. A escalabilidade será avaliada pelo aumento da

capacidade proporcionado pelo acréscimo de um ou mais servidores.

A fase de testes das funcionalidades já foi concluída com sucesso. Foram testadas todas as funcionalidades básicas da solução quanto ao envio e recebimento de e-mail, filtragem de vírus e spam, interface do sistema de Webmail etc. O balanceamento de carga através do LVS e o esquema de tolerância a falhas do Heartbeat também estão funcionando. As implementações da firewall, sistema de gerenciamento e storage estão em planejamento.

A fase de testes de benchmark está no início. O primeiro problema é determinar e validar o modelo estatístico adequado. A partir dessas informações, planeja-se realizar a simulação da carga de trabalho dos servidores em diferentes situações.

6. Modelos

Serão analisados modelos estatísticos analíticos que permitirão posteriormente a simulação da carga de trabalho. A literatura indica o uso de distribuições de probabilidade de cauda longa para modelar o tamanho e o tempo entre mensagens, pois existe uma probabilidade pequena, mas não desprezível, de ocorrerem valores que apresentam uma grande variabilidade - o que também reduz o significado estatístico dos resultados.

Berlotti e Calzarossa [3,4] sugerem que as distribuições de probabilidade para o tempo entre as mensagens consecutivas sejam: Weibull para o corpo do histograma, Pareto para cauda e a função Lognormal para o tamanho das mensagens.

Menascé e Almeida [2,12] sugerem, em seus estudos para o protocolo HTTP, que o tempo entre requisições seja modelado com a função de Weibull e que o tamanho dos arquivos seja modelado com a função Lognormal para o corpo do histograma e Pareto para a cauda. Esse modelo será aplicado aos dados experimentais mapeando o tamanho dos arquivos para o tamanho das mensagens e o tempo entre requisições para o tempo entre mensagens [11,12]. As fórmulas 1 a 3 apresentam as funções de densidade de probabilidade utilizadas.

$$f(x)=\left(\frac{b}{a^b}\right)*x^{b-1}*e^{\frac{-x^b}{a^b}}$$

Fórmula 1 – Função Weibull

$$f(x)=\frac{a*b^a}{x^{a+1}}$$

Fórmula 2 – Função Pareto

$$f(x)=\frac{e^{\frac{-1}{2}*\left(\frac{\ln x-\mu}{\sigma}\right)^2}}{\sigma*x*\sqrt{2*\pi}}$$

Fórmula 3 – Função Lognormal

7. Ajuste do Modelo

O programa que faz o ajuste das funções de densidade de probabilidade foi desenvolvido em Java por um dos autores [7] e utiliza um algoritmo genético para encontrar os valores dos parâmetros.

Os algoritmos genéticos [10] são uma alternativa ao Método dos Mínimos Quadrados (MMQ) para ajustar funções não-lineares, como as distribuições de probabilidade utilizadas, pois são métodos iterativos que dispensam a linearização, podendo apresentar bons resultados com poucas iterações.

O MMQ funciona bem quando as funções ajustadas são lineares ou polinômios, mas no caso de funções não-lineares é necessário linearizar os dados, levando a uma solução aproximada. O processo de linearização tende a apresentar imprecisões, especialmente quando existe uma grande quantidade de dados. O ajuste obtido dessa forma é a melhor solução para os dados linearizados, e corresponde apenas a uma estimativa inicial para o ajuste dos dados originais que precisam ser melhorados com outros métodos[11].

8. Análise dos Dados

Para verificação do modelo foram avaliados servidores e usuários reais através da análise dos logs das mensagens de e-mail de duas empresas, referenciadas como A, B e C por questões de sigilo.

Na época os servidores das empresas A e B não possuíam software anti-spam e anti-vírus e apenas exigiam autenticação pelo protocolo POP3 antes do uso do protocolo SMTP para controle de relays indevidos.

Os dados foram analisados estatisticamente, obtendo-se os histograma de quantidade, tamanho e tempo entre mensagens recebidas e enviadas por cada empresa. O quadro 1 mostra as características dos dados analisados.

Usuários válidos são aqueles cadastrados no servidor e que usam o serviço de e-mail durante o período de coleta das informações. Os inválidos são usuários que não existem no servidor e correspondem a tentativas de uso inadequado (spam, relay, etc).

As mensagens de controle ocorrem quando um servidor tenta avisar o remetente da mensagem ou o administrador do sistema que por algum motivo uma

mensagem não foi processada corretamente (usuário inválido, falha de comunicação, queda da rede, etc).

Servidor A (empresa)
7 domínios
Mensagens Recebidas:
676.459 total de mensagens
416.984 mensagens com endereço válido
2.483 endereços inválidos
87 endereços válidos
Mensagens enviadas:
958.785 total de mensagens
802.331 mensagens de controle
28.693 mensagens com endereço válido
58.670 endereços inválidos
85 endereços válidos
logs de 4 meses analisados
de 13/04/2004 8:53:50 até 07/08/2004 19:00:24
Total de 377 endereços válidos cadastrados.

Servidor B (empresa):
25 domínios
Mensagens Recebidas:
256.973 total de mensagens
171.587 mensagens com endereço válido
2.600 endereços inválidos
99 endereços válidos
Mensagens enviadas:
283.518 total de mensagens
103.725 mensagens de controle
10.513 mensagens com endereço válido
24.226 endereços inválidos
59 endereços válidos
logs de 4 meses analisados
de 12/02/2004 19:36:56 até 17/08/2004 16:32:22
Total de 220 endereços válidos cadastrados.

Servidor C
1 domínio
Mensagens Recebidas:
1.462.000 total de mensagens
86.109 mensagens com endereço válido
49.847 endereços inválidos
987 endereços válidos
Mensagens enviadas:
1.464.000 total de mensagens
200.474 mensagens com endereço válido
143.381 endereços inválidos
1763 endereços válidos
logs de 1 meses analisados
de 01/01/2005 00:00:00 até 31/01/2005 23:59:59
Total de 2500 endereços válidos cadastrados.

Quadro 1: Dados Analisados

9. Resultados Obtidos

Os dados coletados foram analisados estatisticamente e os histogramas são apresentados nessa seção. Os histogramas do intervalo entre mensagens e do tamanho de mensagens são apresentados com o ajuste das funções de densidade de probabilidade.

No máximo as primeiras 25 classes de cada histograma são apresentadas. Devido à ocorrência de valores extremos, alguns histogramas chegam a ter quase mil classes, mas a quantidade de ocorrências a partir da 26ª classe é pouco significativa, abrangendo menos de 3% das amostras.

As análises de quantidade de mensagens serão feitas para usuários válidos. Foram considerados para essa análise os usuários que receberam ou enviaram ao menos uma mensagem por dia (em média). A análise da quantidade de mensagens consiste apenas de um estudo que permite avaliar o volume de mensagens durante o período de análise, pois é dependente do tempo entre mensagens, que é analisado em detalhes mais adiante.

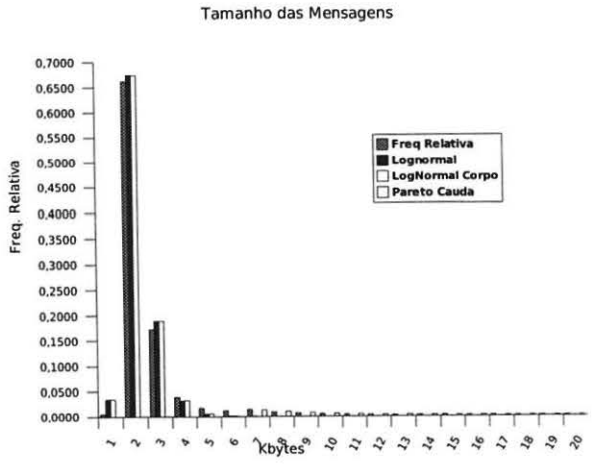

Fig. 2a: Tamanho das mensagens da empresa A

No caso das mensagens recebidas, foram analisados 118 usuários da empresa A, que apresenta uma mediana de 749,5 mensagens por usuário, média de aproximadamente 1664 mensagens por usuário e desvio padrão de aproximadamente 2806 mensagens por usuário. No caso da empresa B foram analisados 50 usuários, apresentando uma mediana de 839 mensagens por usuário, média de aproximadamente 2064,5 mensagens por usuário e desvio padrão de aproximadamente 3779 mensagens por usuário.

No caso das mensagens enviadas, foram analisados 128 usuários da empresa A, que apresenta uma mediana de 242 mensagens por usuário, média de

aproximadamente 371,5 mensagens por usuário e desvio padrão de aproximadamente 347,5 mensagens por usuário. Foram também analisados 56 usuários da empresa B, que apresenta uma mediana de 424 mensagens por usuário, média de aproximadamente 911 mensagens por usuário e desvio padrão de aproximadamente 1571 mensagens por usuário.

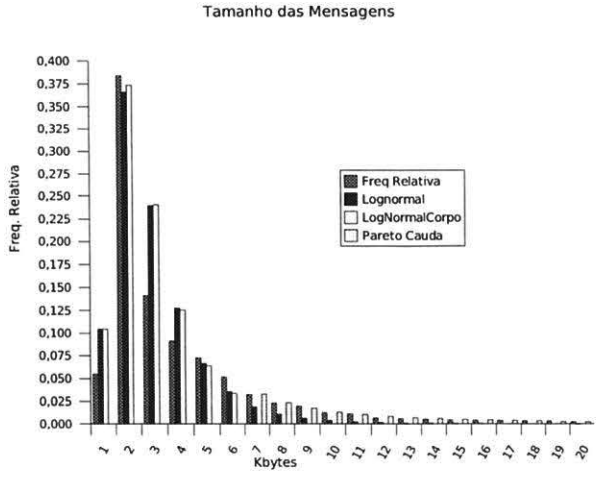

Fig. 2b: Tamanho das mensagens da empresa B

O histograma do tamanho das mensagens das empresas A e B com as funções de probabilidade ajustadas são mostrados nas figuras 2a e 2b respectivamente. Foram analisadas 953112 amostras da empresa A, que apresenta uma mediana de 1,883 kbytes, média de aproximadamente 5,595 kbytes e desvio padrão de aproximadamente 27,341 kbytes. O melhor ajuste foi obtido com uma função Lognormal ajustada ao corpo do histograma (de 0 a 5,5 kbyte) com coeficientes mi=0,43 e sigma=0,39 e erro quadrático total de 0,015 e uma função Pareto a cauda (a partir de 5,5 Kbyte) com coeficientes a=1,20 e b=0,70 e erro quadrático total de 0,000005.

Para a empresa B foram analisadas 282657 amostras apresenta uma mediana de 2,210 kbytes, média de aproximadamente 13,072 kbytes e desvio padrão de aproximadamente 72,131 kbytes. O melhor ajuste foi obtido com uma função Lognormal ajustada ao corpo do histograma (de 0 a 5,5 kbyte) com coeficientes mi=0,71 e sigma=0,63 e erro quadrático total de 0,014 e uma função Pareto a cauda (a partir de 5,5 Kbyte) com coeficientes a=1,33 e b=1,65 e erro quadrático total de 0,000002.

As figuras 3a, 3b mostram os histogramas do tempo entre mensagens recebidas pelas empresas A e B.

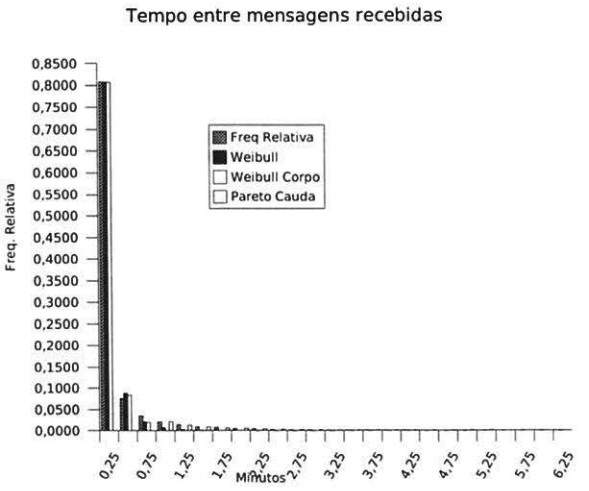

Fig. 3a: Tempo entre as mensagens recebidas pela empresa A

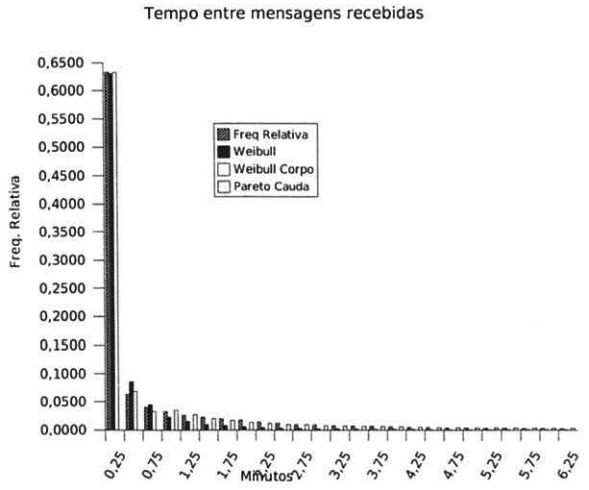

Fig. 3b: Tempo entre as mensagens recebidas pela empresa B

Foram analisadas 676454 amostras da empresa A, que apresenta uma mediana de 4,04 segundos, média de aproximadamente 435 segundos e desvio padrão de aproximadamente 3269 segundos. O melhor ajuste foi obtido com uma função Weibull ajustada ao corpo do histograma (de 0 a 0,75 minutos) com coeficientes a=0,21 e b=0,50 e erro quadrático total de 0,0003 e uma função Pareto a cauda (a partir de 0,75 minutos) com coeficientes a=0,92 e b=0,01 e erro quadrático total de 0,000007.

Para a empresa B foram analisadas 256965 amostras apresenta uma mediana de 2,745 segundos, média de aproximadamente 1852 segundos e desvio padrão de aproximadamente 674572 segundos. O melhor ajuste foi obtido com uma função Weibull

ajustada ao corpo do histograma (de 0 a 0,75 minutos) com coeficientes a=0,005 e b=0,29 e erro quadrático total de 0,0001 e uma função Pareto a cauda (a partir de 0,75 minutos) com coeficientes a=0,27 e b=0,0003, com quadrático total de 0,00007.

As figuras 4a, 4b apresentam os histogramas do tempo entre mensagens enviadas pelas empresas A e B.

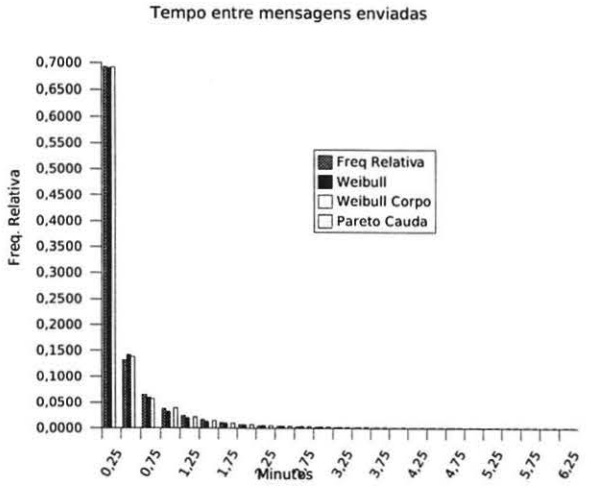

Fig. 4a: Tempo entre as mensagens enviadas pela empresa A

Foram analisadas 546220 amostras da empresa A, que apresenta uma mediana de 7,47 segundos, média de aproximadamente 539 segundos e desvio padrão de aproximadamente 36379 segundos. O melhor ajuste foi obtido com uma função Weibull ajustada ao corpo do histograma (de 0 a 0,75 minutos) com coeficientes a=0,02 e b=0,34 e erro quadrático total de 0,0001 e uma função Pareto a cauda (a partir de 0,75 minutos) com coeficientes a=1,30 e b=0,05 e erro quadrático total de 0,00002.

Para a empresa B foram analisadas 256965 amostras apresenta uma mediana de 2,745 segundos, média de aproximadamente 1852 segundos e desvio padrão de aproximadamente 674572 segundos. O melhor ajuste foi obtido com uma função Weibull ajustada ao corpo do histograma (de 0 a 0,75 minutos) com coeficientes a=20.4 e b=0,59 e erro quadrático total de 0,0007 e uma função Pareto a cauda (a partir de 0,75 minutos) com coeficientes a=0,42 e b=0,02, com quadrático total de 0,0007.

A análise dos dados do servidor C ainda não foi concluída, mas a avaliação de médias mensais permitiu calcular algumas estatísticas do tempo entre mensagens: média de 0,037 minutos, mediana de 0,039 minutos e desvio padrão de 0,06 minutos.

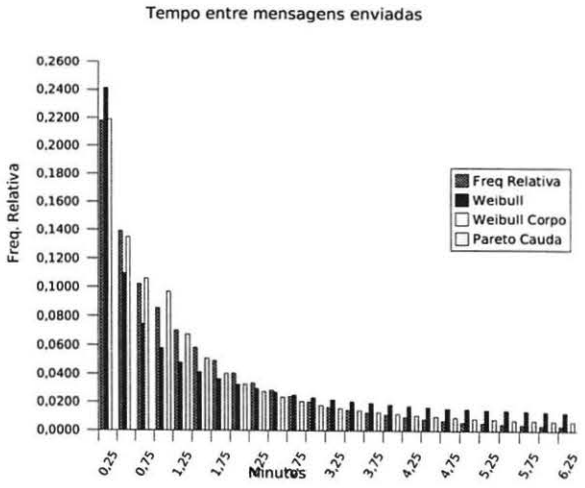

Fig. 4b: Tempo entre as mensagens enviadas pela empresa B

10. Análise dos Resultados

De modo geral, os histogramas mostram uma grande concentração de amostras nas primeiras classes, e uma extensa cauda, indicando que valores extremos distorcem os resultados, reduzindo o significado estatístico das análises. A grande diferença entre a média e a mediana e o valor exagerado do desvio padrão confirmam essa consideração.

As funções de densidade de probabilidade se ajustaram bem ao tempo de mensagem das duas empresas, apresentando melhor resultado o modelo proposto por Berlotti e Calzarossa [3,4]: Função de Weibull para o corpo do histograma e Pareto para a cauda. No caso do tamanho das mensagens o modelo mais adequado para a empresa B foi o proposto por Menascé e Almeida [2,12].

Apesar do bom resultado dos ajustes, uma análise do erro da função ajustada em relação aos dados do histograma permite identificar os limites dos modelos.

A figura 5a mostra a variação do erro absoluto do melhor ajuste obtido para o intervalo entre mensagens enviadas pela empresa B. Observa-se que o erro tende a diminuir a partir de 6,75 minutos.

O mesmo não ocorre quando se analisa o erro relativo (Fig. 5b). O erro relativo fica abaixo de 30% nas primeiras classes, porém acima de 6,75 minutos, explode para mais de 50%, pois até a função Pareto apresenta dificuldades de ajuste para a longa cauda do histograma

Felizmente até 6,75 minutos concentram-se 97,5% das amostras. A situação para os outros ajustes é similar, o ponto de corte muda um pouco, mas a porcentagem de amostras se mantém.

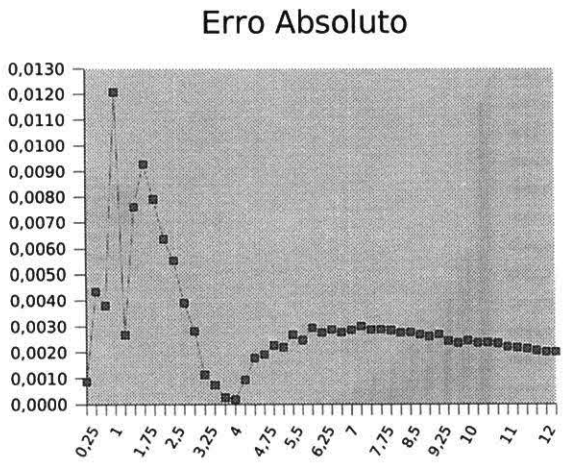

Fig. 5a: Erro absoluto do ajuste

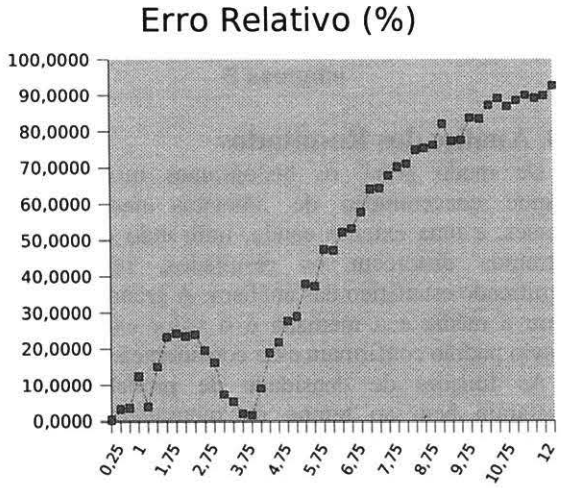

Fig. 5b: Erro relativo do ajuste

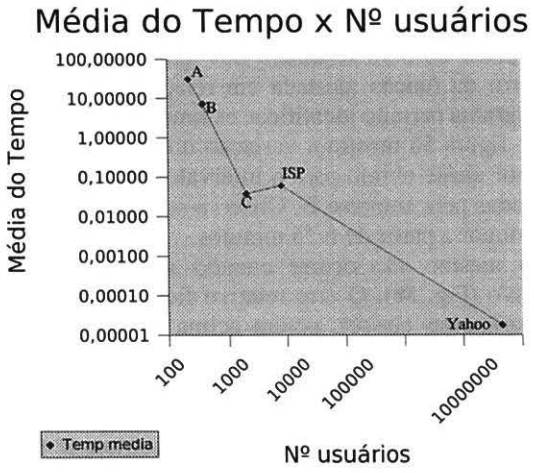

Fig. 6: tempo entre mensagens x nº de usuários

Uma vez que o tempo entre mensagens é dependente da quantidades de usuários, os resultados obtidos foram combinados com alguns dados da literatura, para permitir a avaliação da escalabilidade do modelo. A figura 6 ilustra a variação do tempo entre mensagens recebidas com o número de usuários cadastrados no servidor. A sigla ISP no gráfico simboliza o provedor analisado nos trabalhos de Berlotti e Calzarossa [3,4], com 8057 usuários e média=0,06 minutos. O Yahoo é representado com dados de março de 1999 [8], com 45 milhões de usuários e a média = 1,06 ms. Ajustando a função adequada a esse gráfico, é possível estimar o valor médio do tempo entre mensagens para a quantidade de usuários a ser simulada.

11. Simulação

Uma vez selecionados os modelos adequados e obtidos os histogramas e o ajuste correspondente das funções de densidade de probabilidade, serão desenvolvidos programas em Java para simular a carga de trabalho, usando a API Java Mail para enviar e receber automaticamente mensagens de e-mails artificiais em grande quantidade para usuários previamente cadastrados no sistema.

O tamanho de cada mensagem artificial com caracteres aleatórios será obtido a partir de um gerador de números ramdomicos que obedecem a distribuição Lognormal e Pareto ajustadas aos histogramas do tamanho das mensagens dos servidores analisados. O tempo entre mensagens consecutivas será calculado de acordo com as distribuições de Weibull e Pareto ajustadas aos histogramas do tempo entre as mensagens. Um exemplo de mensagem gerada é apresentado no quadro 2.

Assunto: GKoQQEdXlFsp-SSKNec~6u~xGR<
Texto: P$/(W}ZyLxG0n>k-,_DN9CV&
[n0UbnmIV$w1i-&:kL%(\6Us]zk[_+5i}K[#Fu
LRGgvV+2dr@eAGJiwFRmM$['
87 caracteres, gerados às Tue Jun 07 10:45:23 BRT 2005

Quadro 2: Exemplo de uma pequena mensagem criada pelos programas geradores de carga.

Quando os softwares de simulação de carga estiverem concluídos, o sistema poderá ser estimulado de maneira quase realística, permitindo estudar, através de dados coletados com SNMP, como a carga de trabalho afeta a utilização dos recursos, como CPU, memória, disco e rede, e quais os componentes críticos em diferentes situações [2,12].

12. Conclusão

O cluster de servidores de e-mail proposto oferece propriedades que possibilitarão suportar um grande número de usuários. O dimensionamento do cluster é uma questão desafiadora, que exige um estudo cuidadoso da carga de trabalho a partir de situações reais, caracterização de usuários e simulações quase realísticas, que permitirão determinar as melhores configurações e possíveis gargalos do sistema. O sistema de teste resultante deverá ser extremamente flexível, sendo extensível para o uso de diferentes protocolos.

O estudo do tráfego de e-mails com modelos estatísticos se mostrou adequado se os modelos forem corretamente ajustados, desde a escolha do tamanho das classes, até a definição das funções de densidade de probabilidade que melhor se ajustam aos histogramas. Os resultados obtidos são aderentes ao modelo sugerido por Berlotti e Calzarossa [3,4], mas a combinação com os modelos sugeridos por e Menascé e Almeida [2,12] geram resultados melhores para os dados analisados. Os ajustes possuem um intervalo de validade, pois acima de um certo limite o erro relativo se torna muito grande. Os resultados sugerem que o modelo precisa ser complementado pelo ajuste de uma distribuição uniforme para os valores que estão acima do limite dos modelos utilizados, de forma a permitir que a simulação gere os valores extremos que ocorrem na realidade. Os autores estão analisando dados de mais organizações para garantir a validade do modelo e sua escalabilidade; a obtenção de dados reais é difícil porque muitas empresas consideram os logs dos servidores de e-mail confidenciais. O modelo também pode ser complementado com outras técnicas de análise, como a Teoria das Filas, Caos Determinístico e Fractais, Análise Espectral e Wavelets.

Os parâmetros obtidos com os ajustes das funções de densidade de probabilidade permitem o desenvolvimento de geradores de carga com a finalidade de simular a carga de trabalho de servidores reais. O estudo das estatísticas por usuário deve permitir aperfeiçoamentos nos modelos para a simulação de usuários reais de acordo com perfis de comportamento, criando assim um novo personagem: o usuário virtual.

Simulando o sistema com a carga de trabalho realística, e obtendo medições detalhadas de seu desempenho em relação a cada recurso computacional, será possível especificar a infraestrutura necessária para implementação de uma solução adequada à demanda de milhares de usuários.

13. Referências

[1] ZHANG, Wensong, ZHANG, Wenzhuo "Linux Virtual Server Clusters - Build highly-scalable and highly-available network services at low cost" , Linux Magazine, Nov. 2003.

[2] MENASCÉ, D.A. "Loading Testing of Web sites", IEEE Internet computing, IEEE, Jul./Ago. 2002.

[3] BERLOTTI, L; CALZAROSSA, M.C. "Models of Mail Server Workloads", Performance Evaluation 46, Elsevier, pp 65-76, 2001

[4]BERLOTTI, L; CALZAROSSA M.C. "Workload Characterization of Mail Servers", Proceedings of SPECTS2000, Vancouver, Canadá, Jul. 2000.

[5]ZHANG, Wensong, , "IPVS Home Page", Internet, Wensong, "HTTP://www.linuxvirtualserver.org", Acessado em 9 jan. 2005.

[6] ROBERTSON, A, "Highly-affordable High Availability", Linux Magazine, Nov . 2003.

[7] AMORIM, A.C.O., "Ajuste de Funções com Algoritmos Genéticos"[1], disponível em <http://cognitio.incubadora.fapesp.br/portal/atividades/cursos/posgrad/vida_artificial/2005/trabalhosfinais/Antonio_TF_Monografia.pdf>, Acessado em 1 set. 2005.

[8] SAITO, Yasushi; BERSHAD, Brian; LEVI, Henry. Manageability, availability and performance in Porcupine: a highly scalable, cluster based mail server. University of Washington. Dez. 1999.

[9] BEHREN, J.R.V., CZERWINSKI, S., Joseph, A.D., BREWER, E.A., KUBIATOWICZ, J., "NinjaMail: The design of a High-Performance clustered, Distributed E-mail System", IEEE, 2000.

[10] MITCHELL, M., "An Introduction to Genetic Algorithms", Cambridge : MIT, 1996

[11] MILONE, G, "Estatística Geral e aplicada", Thomson, São Paulo, 2004.

[12] MENASCÉ, D. A., ALMEIDA, "Planejamento de Capacidade para Serviços na Web", Campus, Rio de Janeiro, 2002.

Agradecimentos

Agradecemos ao IPT/Cenatec, ao PAD/LSI e aos professores Claúdio Marte e Antonio Rigo, pelo apoio na realização do projeto, e aos colegas da Prodesp e do Metrô de São Paulo, pela contribuição na fase inicial dos trabalhos.

[1] Monografia Final do Curso PSI-5000 – POLI/USP