

Custos em Redes de Computadores

Rosana C. de M. Grillo Gonçalves

Faculdade de Economia, Administração e Contabilidade - FEARP
Universidade de São Paulo - USP - Campus de Ribeirão Preto - Brasil
CEP: 14040-900 - RIBEIRÃO PRETO - SP
Fone: (016) 624-2846
FAX: (016) 633-6133 - e-mail: rosanagg@cat.cce.usp.br

RESUMO

Os custos das redes de computadores são abordados segundo duas óticas: a das empresas que prestam os serviços de rede, e a das empresas que utilizam-se das redes na consecução de suas diversas atividades operacionais e administrativas. Problemas como a mensuração dos custos das ligações públicas para as empresas por elas responsáveis (concessionárias), e como estes custos são repassados para diferentes empresas comerciais que se utilizam de seus serviços serão estudados. A alocação dos custos da rede aos diversos centros de custos da empresa usuária é discutida através de um exemplo. Além do custo relativo a tramitação pelas vias públicas, há diversos outros custos internos a serem considerados, e diferentes métodos de agrupá-los.

A discussão é ampliada com a consideração dos problemas intrínsecos à mensuração de consumo de recursos num ambiente inter-redes TCP-IP. É demonstrado que qualquer que seja o esquema utilizado para a mensuração do custo, este deve ser passado para usuário final, como um feedback que o oriente a fazer um melhor uso dos recursos globais da rede

1. Introdução

Os resultados que superaram as expectativas positivas da implantação e dissimulação de uso das redes de computadores, e seu grande avanço tecnológico, fazem com que o estudo de seus custos torne-se emergente, e ao mesmo tempo uma atividade complexa. A complexidade deste estudo advém, em grande parte, das mudanças freqüentes de paradigmas tecnológicos.

Pode-se dizer que os custos da primeira geração de redes corporativas de domínio de negócio único, fundamentalmente introduzidas por bancos, baseadas em equipamentos IBM, utilizando-se de um modelo computacional centralizar (um mainframe com terminais remotos) ainda não se encontravam suficientemente estudados, quando surgem os desafios propostos pelas inter-redes.

Inter-redes são macro redes que englobam diversas redes locais ou não, que geralmente são de tecnologia distinta. Redes de computadores IBM podem estar ligadas a redes de estações de trabalho Digital, e a uma rede local de microcomputadores. O surgimento das inter-redes que ocorreu em meios acadêmicos, deveu-se grandemente à flexibilidade proposta por seus protocolos (TCP-IP), e a sua descentralização administrativa. Desde sua origem, a Internet (mais famosa das inter-redes) não repassa a totalidade de seus custos a seus usuários finais, e tratando-se da comunidade científica, praticamente nenhum custo lhe é repassado. A discussão sobre os aspectos políticos e sociais do repasse destes custos vai além do escopo deste trabalho, seu ênfoque é mostrar que a mensuração de custos transcende esta discussão.

Várias outras inter-redes comerciais têm surgido trazendo à tona o problema da mensuração dos custos das ligações públicas para as empresas por elas responsáveis (concessionárias), e como estes custos são repassados para diferentes empresas comerciais que se utilizam de seus serviços [We 93]. Toda a interconexão feita através de vias públicas (fiação subterrânea em ruas, difusão de sinais via satélites de comunicação, etc.) faz parte do que é chamado de ligações públicas.

Outra perspectiva dos custos em redes de computadores a ser discutida neste artigo é alocação dos custos da rede aos diversos centros de custos da empresa. Além do custo relativo a tramitação pelas vias públicas, há diversos outros custos internos a serem considerados, e diferentes métodos de agrupá-los. Este item é desenvolvido-se para uma classe particular de rede.

Finalmente é feita uma ampla discussão sobre os problemas intrínsecos à mensuração do consumo de recursos num ambiente inter-redes TCP-IP. A ênfase desta discussão mostra que qualquer que seja o esquema utilizado para a mensuração do consumo, este deve ser passado para usuário de forma a orientá-lo a fazer um melhor uso dos recursos da rede, ou seja, deve haver um *feedback* para os usuários finais.

2 Mensuração dos Custos das Redes Públicas para as Concessionárias e seu Repasse às Empresas Clientes

Em vários países do mundo, o monopólio e o caráter social atribuído às redes de ligações públicas, ou seja, às redes que transportam mensagens por vias públicas consagrou um sistema único de tarifas.

No Brasil, este modelo tem sido, presentemente questionado, devido ao fato de os recursos por ele gerado, não terem sido suficientes (ou não terem sido geridos) para que fossem feitos os investimentos necessários a fim de tais redes proporcionarem os recursos já disponibilizados pela tecnologia em países desenvolvidos. Maior polêmica encontra-se em discussões sobre os altíssimos preços das tarifas, atribuída ao modelo de monopólio por uns, e por outros, aos deveres sociais do provimento de recursos num país tão heterogêneo. As possíveis alterações legais, previstas para este ano, poderão alterar este quadro.

O modelo americano para a prestação dos serviços de redes de dados em via pública prevê, a exemplo dos serviços de telefonia, inúmeras empresas comerciais (AT&T, GTE Telenet, MCI, PSI Inc., etc.). Tais empresas operam as chamadas redes comerciais, isto é, as redes que oferecem serviços para quaisquer usuários, quaisquer que sejam seus ramos de negócio. Vários questionamentos têm sido feito pelos preços praticados por estas empresas comerciais, conhecidas por "carriers" [We 93].

Será feita uma discussão breve sobre os custos gerais de uma concessionária/carrier, os custos da geração de dois tipos de serviço por elas oferecido, e sobre como deveriam ser formados os preços para as empresas usuárias destes serviços.

(a) Os Custos Gerais de uma Concessionária/Carrier

Os custos gerais de uma concessionária/carrier indubitavelmente incluem o investimento inicial para formação da rede, os custos com sua manutenção, e despesas administrativas e com vendas. Além das depreciações convencionais, existe outro tipo de depreciação a ser discutida, a depreciação por ação do tempo em função da obsolescência tecnológica. Alguns estudiosos de rede¹ defendem que avanços além da tecnologia baseada em fibras ópticas, já disponibilizada no momento, são improváveis por esbarrarem em limites físicos. No entanto, existem razões históricas para advogar-se a favor da consideração da obsolescência tecnológica.

Outro ponto polêmico é até que ponto tais empresas devem estar engajadas com pesquisa e desenvolvimento. Existem indicadores que todos os avanços na área de Redes de Computadores continuarão a ser feitos através de consórcio de diferentes empresas, com seus resultados sendo disponibilizados através de especificações e softwares de domínio público. Mesmo assim, inúmeras empresas continuam a pesquisar soluções próprias, o que sem dúvida nenhuma, implica em enormes investimentos.

(b) Custo e Preço dos Serviços Fornecidos pelas Concessionárias/Carriers

(b1) Ligações Ponto a Ponto

Estas ligações são feitas entre apenas dois pontos da rede, por uma linha de capacidade (vazão ou *bandwidth*) pré-definida. Tais ligações são bastante utilizadas na ligação de terminais remotos a uma estação central de processamento. O custo do provimento de tal linha é fixo, em função de sua velocidade e da distância, e, portanto, fixo deve ser seu preço.

(b2) Ligações Múltiplos Pontos

Estas ligações estão geralmente ligadas às redes de comutação de pacotes. A idéia destas redes é o particionamento das mensagens em pacotes, que por serem de menor tamanho, consomem menos espaço de armazenamento nos nós da rede. Todos os pacotes possuem endereço de sua origem e de seu destino. Os nós responsáveis pela comutação dos pacotes, ou seja, por definirem o caminho que ele perseguirá até seu destino, fazem seu trabalho mais rapidamente por tratarem de unidades menores. A preocupação com a sequencialização dos pacotes é feita a posteriori, o que ajuda a transmissão bastante rápida. Não é feita uma reserva dos meios de transmissão da rede, os canais vão sendo alocados dinamicamente, conforme os endereços destino são interpretados. Há o risco do congestionamento de determinado canal, ou seja, podem ocorrer atrasos de transmissão variáveis e dependentes da carga da rede.

Com exceção dos casos em que a capacidade instalada da rede supre com folga todo tráfego da demanda agregada, tais redes podem oferecer serviços comprometidos por um tráfego intenso. Neste caso, seria razoável que os preços praticados variassem em função do número de pacotes ou de bytes transmitidos.

No mercado americano, para a transmissão de dados, os preços são fixos, variando em função do tráfego gerado apenas no caso da transmissão de voz. No Brasil, os preços são ainda determinados em função do número de mensagens transmitidas

¹ Esta e outras idéias apresentadas neste artigo foram discutidas com o Prof. Dr. Álvaro Garcia Neto, professor e pesquisador do IFSC - USP.

Deve ser observado, que dependendo do alcance geográfico da rede, suas mensagens poderão trafegar por linhas públicas de diferentes concessionárias/*carriers* neste caso, os preços podem ser dependentes de acordos e tratados semelhantes àqueles praticados pela telefonia.

3. Alocação dos Custos da Rede a Diversos Centros de Custos - Um Modelo

Nesta seção os custos das redes de computadores é visto sob a ótica das empresas que utilizam das redes para levarem a cabo seus negócios. A gama de usuários, tradicionalmente encabeçada por bancos e agências de passagens aéreas, hoje se estende incluindo até mesmo pequenas e médias empresas.

O modelo a ser discutido diz respeito a uma rede baseada em tecnologia proprietária, tecnologia da IBM. Através deste exemplo pretende-se evoluir para um consenso de que a recuperação dos custos de redes requer *a identificação dos serviços produzidos, a mensuração dos custos de produção dos serviços, a identificação dos clientes dos serviços, e a caracterização do consumo por cliente (quantas vezes consome, e a duração de cada sessão de consumo)*.

Este modelo simplificadamente assume que os serviços correspondem às diversas sessões de acesso às aplicações disponíveis no sistema. A mensuração dos custos de produção dos serviços é feita basicamente através de pacotes com inventários dos ativos em hardware e software.

Os pacotes de inventário de ativos são basicamente usados para identificar todos os equipamentos envolvidos na operação da rede de forma que seus custos possam ser distribuídos no processo de recuperação.

A identificação dos clientes dos serviços é facilitada por algumas características deste ambiente. Nestas redes, para que as aplicações sejam disponibilizadas a outros terminais ou outros computadores, são utilizadas bibliotecas de softwares de comunicação específicos (como VTAM e NCP), que fornecem os dados para o seguinte mapeamento: terminais x aplicações acessadas, que mostra a quais aplicações cada terminal teve acesso.

O problema maior é a caracterização do consumo por cliente. Alguns autores propõem que os clientes de determinado centro de custo sejam conectados através de um mesmo "cluster", que pode ser um concentrador de terminais. A idéia é a maior adequação possível entre as hierarquias de centros de custo da organização e o layout da rede. Assim sendo, a mensuração do consumo seria feita numa granularidade menos fina, e portanto, simplificada [Keys 92]. Esta mensuração pode ser feita através de amostragens estatísticas utilizando-se os monitores/gerenciadores dos *hosts*.

Os monitores dos *hosts* executam determinadas funções de contabilidade durante uma sessão tais como registrar o início e o fim de uma sessão, monitorar os dados trocados entre o transmissor e o receptor, identificando o tráfego devido a troca de dados, versus o o tráfego total, registrando o nome do terminal (SLU) e o nome da aplicação (PLU) e registrando o nome da subárea e dos *links*. Os dados de contabilidade da sessão são descarregados num banco de dados (por exemplo, NetView insere 39 registros de *log* no SMF).

Exemplos de softwares gerenciadores de *hosts* incluem o Netview da IBM, NetSpy da Legend, NetMaster do Systems Center, e TMON/DB2 do Landmark Systems. Eles basicamente medem a utilização em termos de bytes, ocupação e conexões.

De um modo geral, a recuperação dos custos da rede é muito semelhante à recuperação de custos de computadores. O usuário final consome os serviços enquanto executa suas tarefas. Os recursos são consumidos na produção destes serviços. Os usuários finais são agrupados em centros de custo e os recursos são agrupados em conjunto de custos ("cost pool").

Os centros de custos frequentemente estão de acordo com a estrutura organizacional da empresa, e cada centro de custos tem seu próprio orçamento a ser gasto com os serviços que consome. O "cost pool", que abrange os recursos, no caso das funções de transmissão de dados, inclui terminais, controladores de terminais, processadores front-end, modems, linhas, software de comunicação (VTAM, NCPP), o centro de controle da rede, pessoal de suporte à comunicação, etc.

Os serviços providos aos usuários finais são geralmente definidos como sessões entre o terminal de um usuário final (por exemplo, um terminal no departamento de finanças) e uma aplicação (por exemplo, uma aplicação CICS que faz contas a pagar). O usuário final vê o serviço como uma sessão on-line que lhe permite executar certos tipos de transações.

Apesar deste modelo enfatizar a necessidade de medir-se o custo por uso, ou seja pela demanda efetiva, deve ser ressaltada a inexistência de ferramentas que o façam no nível de precisão desejado. Portanto, uma alternativa é medir-se a capacidade instalada de cada centro de custo, ou seja, qual é o máximo de recursos que poderia por ele ser consumido, e somar-se à medida estimada de seu consumo. Os custos serão então atribuídos proporcionalmente ao valor deste total.

4. Mensuração do Consumo de Recursos num Ambiente Inter-redes e Sua Apresentação para os Usuários Finais

Esta seção faz uma discussão geral, da perspectiva do usuário, sobre a questão de seu consumo de recursos de inter-redes de comutação de pacotes, assumindo que aplicações envolvendo dados, voz e imagem estejam disponíveis. Muitas das considerações sobre as classes de aplicações disponibilizadas fazem com que o estudo esteja bastante voltado ao modelo proposto pela maior inter-rede do momento, a Internet.

Atualmente diversas redes comerciais têm explorado este modelo. A forma mais comum de recuperação de custos hoje nas redes comerciais de comutação de pacotes americanas é uma taxa fixa por conexão física, onde a taxa sempre é uma função da *bandwidth* das linhas utilizadas pela conexão. Exceções existem nos casos de canais de transmissão de voz, onde os preços podem ser função do tráfego gerado.

A falta de um mecanismo de feedback que penalize o usuário por um uso indevido (excessivo) pode causar a degradação da rede como um todo. Um caso típico ocorre quando usuários da Internet resolvem enviar propaganda de determinados produtos, a uma gama extensa de destinatários via difusão (*broadcast*).

Um modelo mais adequado de tarifação deve ser baseado em características relativas ao tráfego gerado pelo usuário. É necessário um *feedback* que motive os usuários a fazerem um uso global eficiente dos recursos existentes. Da perspectiva do sistema,

quando este estiver levemente carregado, o *feedback* deve encorajar (ou pelo menos não desencorajar) os usuários a maximizar a vazão (*throughput*) do sistema. Quando o sistema estiver pesadamente carregado (isto é, quando a demanda aproxima-se ou excede a capacidade finita) o *feedback* deve motivar os usuários de aplicações não tão urgentes a adiar a geração de seu tráfego ou forçar que os usuários que estão demandando recursos em demasia do sistema, cessem de fazê-lo. A ênfase é que o *feedback* ideal a ser provido para o usuário deve encorajar o usuário a um uso eficiente, e ao mesmo tempo garantir que o sistema não seja sobrecarregado.

Um exemplo de *feedback* que motiva um comportamento eficiente do usuário é aquele que o incentiva a excluir as aplicações não tão sensíveis ao tempo de espera dos horários de pico. A Internet, em seu uso atual, pode ser considerada como fornecedora de um *feedback* muito distante do desejado. Por exemplo, no Brasil, durante as horas de pico a performance da rede degrada-se de tal forma que os usuários sensíveis a performance são forçados a protelarem seus trabalhos para horas de utilização não tão intensa. Os usuários menos sensíveis à degradação de performance podem até mesmo terem um incentivo para transmitirem mais para compensar as perdas causadas pela congestão. Os usuários ineficientes não são adequadamente penalizados pela aumento do tempo de espera devido ao aumento da fila de pacotes a serem servidos, ou pela perda de pacotes, que é causada devido às suas ações.

Atualmente a Internet provê um serviço de entrega de pacotes do tipo FCFS (First Come, First Served), que implica numa distribuição uniforme de todos os tipos de pacotes segundo a ordem de chegada. Portanto, todos atrasos e perdas são compartilhados por todos os usuários. Quando a demanda excede a capacidade, o resultado é uma rede que é congestionada durante as horas de pico e conseqüentemente atende precariamente a todos usuários. Em outras palavras, o que é localmente um comportamento eficiente para alguns usuários, resulta num uso de recursos ineficiente globalmente da perspectiva dos recursos da rede.

4.1 Modelos do Feedback para o Usuário dos Efeitos de seu Consumo

Existem quatro alternativas possíveis para a implementação de um modelo de mensuração de uso que gere uma tarifa com o efeito do feedback desejado nos usuários: tarifa simples por pacote ("flat per-packet fee"), por diferentes tipos de serviços (types of service - TOS), pela carga de pico e baseada em prioridades.

(a) tarifa simples por pacote ("flat per-packet fee")

Esta abordagem independe da carga total do sistema, portanto não encoraja os usuários a adiarem o uso protelável.

(b) por diferentes tipos de serviços (types of service - TOS)

Se as redes oferecerem diferentes tipos de serviços, a mensuração seria baseada na qualidade de determinado serviço. Quando a rede estiver sobrecarregada, contudo, o tráfego relativo a novos usuários deverá ser vetado para garantir a qualidade de serviço dos usuários atuais. Neste caso, os usuários podem prever o custo de determinado nível de serviço. Os usuários tem duas opções: ou recebem o serviço no nível exigido, ou não recebem nada.

Se o preço por utilização numa rede com garantia de serviços for independente do tráfego no momento, ou do tráfego esperado, a rede simplesmente aplica uma política FCFS para a demanda de recursos, os que requisitarem os serviços por último, durante horários de pico serão forçados a utilizarem em outro período. Contudo, pode ser for

preferível encorajar a demanda nos momentos em que a rede está sub-utilizada através de uma diferenciação de preços.

(c) carga de pico

Um esquema de preços baseado nos horários de pico provê diferentes *feedbacks* (diferentes preços) dependendo da demanda agregada existente no sistema. Se existirem períodos do dia em que regularmente a rede encontra-se sobrecarregada, este esquema significativamente afastará os usuários proteláveis dos horários de pico. Medidas de tráfego gerado por usuários distintos mostraram padrões de tráfego a grosso modo consistentes nos horários de maior ocupação das redes [Amer 86]. Se os padrões de tráfego não forem previsíveis, as taxas de pico podem variar dinamicamente com a carga da rede. Contudo, as fontes de tráfego podem não ser capazes de prever suas demandas de forma precisa, o que impossibilita o uso deste *feedback* para fins de planejamento orçamentário.

Nesta abordagem, os recursos podem ser expandidos para atenderem a demanda de uma forma dinâmica, isto é, o fornecedor da rede pode conectar recursos adicionais para atender a demanda em horários de pico. Estabelecendo-se um preço mais elevado para as horas de pico, esta receita poderá ser usada para cobrir os custos extras da anexação de recursos.

(d) baseada em prioridades

Neste caso, a rede servirá usuários segundo as suas prioridades, e a taxa relativa ao tráfego gerado será calculada acordemente. Os preços por classe de prioridade dever ser divulgados publicamente a todos usuários. Este esquema é mais adaptativo do que o anterior (por pico de tráfego), uma vez que o estabelecimento de prioridades fornece uma base para a rede adiar seu tráfego de menor prioridade, em favor do tráfego de maior prioridade, nos momentos em que a rede estiver congestionada.

5. Conclusão

O estudo de custos em redes de computadores encontra-se num estágio embrionário. Modelos globais devem ser amadurecidos de forma a influenciarem os fabricantes de produtos, tanto hardware, como software, a incorporarem em seus produtos ferramentas mais eficientes para a mensuração efetiva de uso.

Pesquisas também têm sido feitas quanto a diversificação dos serviços a serem oferecidos. Em [Cocchi 93] é apresentada uma simulação, que baseada no Equilíbrio de Nash revela que é possível estabelecer-se preços de forma a que os usuários de cada tipo de aplicação fiquem mais satisfeitos com a combinação de diferentes combinações de preço e performance, do que serem arbitrariamente submetidos a performances insatisfatórias.

Entende-se que a realização de diferentes simulações poderão contribuir para o estabelecimento de preços sensíveis ao nível de utilização efetiva dos recursos de rede pelo usuário, o que indubitavelmente auxiliará a distribuição dos custos por diferentes centros de custos.

BIBLIOGRAFIA

[Amer 86] Amer, P. et alii - "Local Area Broadcast Network Measurement: Traffic Characterization". In *University of Delaware Tech report 86-12*, January 1986.

[Cocchi 93] Cocchi, R. et alii. - "Pricing in Computer Networks: Motivation, Formulation, and Example". In *IEEE/ACM Transactions on Networking*, pp. 614-627, vol. 1, no. 6, December 1993.

[Keys 92] Keys, S - "Recovering Network Costs". In *Capacity Management Review*, pp. 1-4, vol.21, no. 1, January 1993.

[We 93] Weichselbaum, P. - "Public Pricing Puts Users in the Driver's Seat". In *Network Management*, pp. 17-19, June 1993.