PROJETO TOPOLÓGICO PARA INTERLIGAÇÃO DE REDES LOCAIS

Virgílio José Martins Ferreira Filho [1]

Paulo Henrique de Aguiar Rodrigues [±]

Roberto Diéguez Galvão [*]

[1] Aluno do mestrado do Programa de Engenharia de Produção COPPE/UFRJ (bolsista do CNPQ)

± Analista Consultor do NCE/UFRJ, Professor Adjunto do IM/UFRJ e Professor Colaborador COPPE/SISTEMAS

* Programa de Engenharia de Produção COPPE/UFRJ

Resumo

O projeto topológico de interligação de redes de computadores é um processo complexo, que consiste em selecionar um conjunto de redes e pontes determinando em que rede cada estação usuária será conectada, e como estas redes serão interligadas. A dificuldade do problema impõe que suposições simplificadoras sejam assumidas para que se possa resolvê-lo. Neste artigo o problema é enfocado de forma simplificada, na qual a alocação das estações é conhecida, tratando-se portanto de determinar a topologia da interligação. Para este problema é proposta uma formulação matemática para a qual foi desenvolvida uma relaxação Lagrangeana. A solução Lagrangeana fornece limites inferiores para a solução ótima, os quais são melhorados pela utilização do método de otimização por subgradiente

1 - Introdução

Redes locais são redes de comunicação de dados, tipicamente redes de comunicação chaveadas por pacotes, limitadas geograficamente, cobrindo distâncias variando de dezenas de metros até dezenas de quilômetros. As redes locais derivam diretamente da evolução da tecnologia de redes e da redução do custo dos equipamentos de computação, e tem como característica fundamental o uso de tecnologias de comunicação digital de alta velocidade. Decorrem deste fato redes com altas taxas de transmissão de dados e com custos de transmissão e controle muito baixos, especialmente quando comparados aos custos das demais redes de computadores.

No entanto redes locais apresentam limitações, principalmente no tocante as distâncias físicas que podem cobrir, ao número de estações que podem ser conectadas a rede e à capacidade de manusear o tráfego entre estações. Estas limitações têm provocado cada vez mais a necessidade de interligar redes locais, conseguindo desta forma um ou mais dos seguintes benefícios: extensão geográfica das redes locais, melhor contenção e isolamento de falhas, segurança e controle do acesso, interligação de redes locais heterogêneas, aumento da eficiência e performance, aumento da conectividade e disponibilidade.

A interligação de redes é estabelecida através de um elemento intermediário, denominado genericamente, na nomenclatura da ISO (International Organization for Standardization), por "estação de troca", recebendo o nome particular de repetidor se a interligação se dá no nível físico, de ponte se é no nível de enlace, de roteador no nível de rede ou de comporta ("gateway") em níveis mais altos que nível de rede.

Nas interligações de redes no nível físico basicamente o que se faz é estender a rede original, assim todo o tráfego gerado em um segmento aparece em todos os segmentos. Este tipo de interligação não é, portanto, muito útil.

Nas interligações nos níveis de enlace e de rede, os segmentos são interligados de forma mais inteligente, restringindo o tráfego às seções da rede nas quais ele é relevante. A seleção entre pontes e roteadores tem sido amplamente pesquisada (IEEE - Network, edição especial de janeiro de 1988 sobre interligação de redes locais), sendo o uso das pontes mais divulgado.

Uma interligação de redes locais deve preservar as características destas redes no tocante aos atrasos e capacidade de transmissão, devendo entretanto ter uma topologia mais dinâmica do que a de uma rede simples devido às possibilidades de falhas das redes e das pontes. Com a flexibilização da topologia, torna-se necessário um algoritmo que promova um roteamento das mensagens através das pontes. O algoritmo mais frequentemente utilizado é aquele em que o roteamento se baseia na obtenção de uma árvore geradora mínima [BACK 88], [HART 88], [SINC 88].

Sob estas condições o uso das pontes é transparente, isto é, nenhuma participação das estações finais é requerida para que se faça uso dos serviços providos pela ponte e não são necessárias modificações nas estações finais. Uma ponte transparente se inicializa, autoconfigura e roda sem intervenções do gerente da rede. Contudo, se possível, é altamente desejável que exista uma interface de gerenciamento remoto para configurações e performance especiais, segurança e monitoramento gerencial. Muito embora a topologia física possa ser arbitrária, a topologia lógica deve ser uma árvore, sendo portanto requerida a operação de um algoritmo de árvore geradora mínima. Este algoritmo deve também permitir a alocação automática de pontes sobressalentes sempre que houver falhas de pontes ou redes locais. Como as mensagens de dados são encaminhadas através de uma árvore, pode ocorrer de um caminho entre duas estações não ser ótimo. Contudo estas mensagens não entrarão em ciclos, e somente uma cópia de cada mensagem será recebida por cada rede local.

A operação de uma ponte transparente pode ser resumida da seguinte forma: pontes recebem e transmitem mensagens das redes às quais estão conectadas. Uma tabela de endereços de estações é mantida pela observação dos endereços de origem das mensagens que chegam em cada uma das portas da ponte (aprendizado). Mensagens são transmitidas ou descartadas em função das informações contidas nestas tabelas. Assim, mensagens cujo destino é sabido estar do mesmo lado da ponte que recebeu a mensagem são descartadas; mensagens que estão de outro lado da ponte ou têm endereço desconhecido são transmitidas (transmissão); mensagens adicionais são trocadas entre as pontes com a finalidade de informar alterações na topologia da rede. Tais informações serão utilizadas para permitir a reconfiguração da rede, através de um algoritmo de árvore geradora mínima.

Como a topologia lógica da interligação deve ser uma árvore a topologia física da interligação deve ser criteriosamente determinada, de modo a se evitar que retardos excessivos sejam impostos às mensagens que trafegam pelo sistema de redes. O problema de se projetar a topologia da interligação é enfocado neste artigo, sendo o problema definido na seção 2, na seção 3 é proposta uma formulação matemática para a qual foi desenvolvida uma relaxação Lagrangeana, apresentada na seção 4. A solução Lagrangeana fornece limites inferiores para a solução ótima, os quais são melhorados pela utilização do método de otimização por subgradiente .

2 - Definição do problema

2.1 - Problema Geral

Um problema encontrado por projetistas e gerentes de redes de computadores é: Qual deve ser a configuração global de uma rede de computadores? Determinar esta configuração envolve selecionar a localização das pontes, decidir como as estações serão ligadas a estas pontes e ainda decidir qual será a rota a ser usada por cada par de usuários que deseja se comunicar de modo a assegurar um nível aceitável de performance a custo mimino. De maneira mais esquemática o problema pode ser colocado da seguinte forma:

DADOS:

i - número e posição das estações a serem interligadas

ii - requisitos de tráfego: taxa de chegada a cada estação e probabilidade de que uma mensagem que chegue em uma estação S se destine a outra estação T.

iii - restrições de confiabilidade.

iv - custos unitários (estrutura das tarifas, custos de hardware, manutenção,etc).

OBTER:

a - número de redes a serem interligadas.

b - a alocação das estações às redes.

c - a topologia da interligação, isto é, o número e a posição das pontes.

d - a rota a ser seguida por uma mensagem que se origina numa estação S e se destina a uma estação T.

OBJETIVO:

Minimizar o custo global do sistema. Este custo é composto de duas parcelas básicas: uma que retrata os custos de instalação e operação da rede e outra que retrata o custos dos retardos impostos pela rede às mensagens que por ela circulam.

2.2 - Problema simplificado

O problema acima colocado é computacionalmente intratável e para solucioná-lo devem ser estabelecidas, a priori, algumas suposições. Uma delas, comum à maioria dos trabalhos existentes na literatura sobre projetos de redes, é a que estabelece que as mensagens na rede seguem rotas estáticas e não bifurcadas. Esta suposicão tem aceitação geral, o que se deve ao fato de ser esta a estratégia usada pela maioria das redes em operação, bem como pelo fato de existirem resultados de simulação [GERL 75] que sugerem que, para redes de porte médio e grande em regime permanente, não existem diferenças significativas entre os retardos nas redes operando com roteamento dinâmicos ou com bons roteamentos estáticos.

Uma outra suposição estabelecida a priori por alguns pesquisadores, como G. Bernard em [BERN 83] e L. Lellis em [LELL 85] é a topologia da rede global, isto é, a posição das pontes e das redes a ela conectadas. O problema então consiste em se alocar as estações a redes pré-definidas. Em contraposição a esta estratégia não será considerado que a topologia da rede global seja conhecida a priori, mas pelo contrário, são supostos conhecidos o número de redes a interligar, bem como a alocação das estações a estas redes. Uma justificativa para que esta estratégia seja seguida é o fato de existir, dentro de uma organização, agrupamentos naturais de estações (estações de um mesmo departamento, divisão, etc) que seriam então interligadas numa rede. O projeto se resume então em interligar todas as redes da organização. Este problema simplificado fica então semelhante ao de projetar uma rede mestra ("backbone network") interligando redes regionais, que vem sendo intensivamente estudado ([ALTI 89], [GAVI 86a], [GAVI 88], [GAVI 89a], [GAVI 89b]).

O problema simplificado aqui tratado pode ser esquematicamente colocado como:

Dados:

i - o conjunto de redes (número, posição e capacidade) a serem interligadas.

ii - as requisições de tráfego a esta redes: taxa de chegada a cada rede e probabilidade de transição entre redes.

iii - que as pontes que farão as interligacões são dotadas de duas portas, isto é, apenas duas redes são conectadas a cada ponte.

iv - que a topologia de interligação é em árvore.

v - que as rotas seguidas pelas mensagens não se bifurcam.

Obter:

a topologia da interligação, isto é, a posição das pontes com o objetivo de minimizar o retardo médio fim a fim na rede global.

Observe-se que uma vez estabelecido que a topologia da interligação é uma árvore, e que cada ponte tem duas portas, então ficam automaticamente determinados: o número de pontes (número de redes - 1) e também as rotas, que serão únicas para cada par origem-destino, dado que elas não se bifurcam. Neste caso a função objetivo se reduz a minimizar o retardo médio fim a fim pois os custos associados à instalação da rede serão fixos, e os custos associados a operação de redes locais podem ser considerados desprezíveis.

Este problema é de natureza combinatória, o que pode ser ilustrado considerando-se o seguinte: seja um sistema com n redes, então o número de possíveis ligações entre elas é :

$$C_n^2 = \frac{n\,(n-1)}{2}$$

Se a topologia desejada é uma árvore, então somente (n-1) ligações serão usadas. O número de topologias com (n-1) ligações é então dado por:

$$C^{n-1}_{n(n-1)/2}$$

Qualquer algoritmo que se baseie na enumeração de todas estas topologias é de complexidade muito elevada para qualquer caso prático.

3 - Formulação matemática do problema simplificado

3.1 - Representação do problema

O problema consiste em interligar as redes através de pontes, o qual será representado por um grafo onde os nós representarão as redes e os arcos representarão as pontes (Observa-se aqui que caso as pontes tivessem mais de duas portas esta representação não seria possivel). Se for considerado que tanto as redes quanto as pontes são indiferentes ao sentido que a mensagem percorre (canais "full duplex") podemos representar o problema por um grafo não orientado $\mathbb{G} = (\mathbb{N}, \mathbb{L})$, onde $\mathbb{N}$ representa o conjunto de vértices (redes) e $\mathbb{L}$ o conjunto de ligações (pontes). As requisições de tráfego serão representadas por uma matriz $\mathbb{D} = [\, d_{ij} \,]_{\mathbb{N} \times \mathbb{N}}$ onde d_{ij} representa a demanda média de tráfego do nó N_i para o nó N_j.

Para que a matriz de requisições de tráfego fique coerente com o grafo não orientado será adotada a matriz de intensidade de tráfego:

$$\bar{D} = [\, \bar{d}_{ij} \,]_{\mathbb{N} \times \mathbb{N}}$$

onde:

$$\bar{d}_{ij} = \begin{cases} d_{ij} & \text{se } i = j \\ d_{ij} + d_{ji} & \text{se } i < j \\ 0 & \text{se } i > j \end{cases} \qquad (1)$$

3.2 - Notações:

Com a finalidade de simplificar a notação, o termo comodidade será usado para representar um par origem-destino de redes que se comunicam e que tem que ser suportado pela interligação de redes.

Com o objetivo de formular o problema, são definidos:

Conjuntos

Π - conjunto de todos os pares origem-destino que se comunicam na rede, isto é, o conjunto de todas as comodidades.

S_p - conjunto de índices de todas as rotas que podem suportar o tráfego gerado pela comodidade **p**, isto é, o conjunto de todas as rotas que se iniciam no nó de origem da comodidade **p** e que terminam no nó de destino da mesma comodidade **p**.

$\mathbb{R} = \bigcup_{p \in \Pi} S_p$ - conjunto de todas as rotas da interligação de redes.

$\mathbb{N}$ - conjunto de índices de todas as pontes a serem usadas na interligação.

$\mathbb{A}$ - qualquer subconjunto de $\mathbb{N}$.

$\overline{\mathbb{A}} = \mathbb{N} - \mathbb{A}$ - conjunto complementar de $\mathbb{A}$ em relação a $\mathbb{N}$.

$\mathbb{L}_r$ - conjunto de índices de todas as ligações usadas na rota **r**.

$\mathbb{L} = \bigcup_{r \in \mathbb{R}} \mathbb{L}_r$ - conjunto de todas as ligações candidatas a serem usadas na interligação.

Índices

i, j, n - índices que indicam nós (redes).

l, b - índices que indicam ligações (pontes).

r - índice que indica rota.

p - índice que indica um determinado par origem-destino (comodidade).

Dados

λ_p - intensidade de tráfego associada a comodidade p, em mensagens por unidade de tempo.

Se a comodidade p tem como nó origem o nó N_i e como nó destino o nó N_j, então:

$$\lambda_p = \overline{d}_{ij} \tag{2}$$

$\lambda = \sum_{p \in \Pi} \lambda_p$ - taxa global de chegadas externas.

λ_r - intensidade de tráfego associada à comodidade que é suportada pela rota r.

τ_i - retardo de propagação e deteção da rede i em unidades de tempo.

$\overline{X}_{N_i}$ - tempo médio de transmissão de uma mensagem na rede **i** em unidades de tempo.

$\overline{X^2_{N_i}}$ - segundo momento do tempo de transmissão de uma mensagem na rede **i**.

$\overline{X}_{B_{ij}}$ - tempo médio de processamento de uma mensagem na ponte **ij** em unidades de tempo.

$\overline{X^2_{B_{ij}}}$ - segundo momento do tempo de processamento de uma mensagem na ponte **ij**.

Variáveis de decisão (binárias)

$$Y_{ij} = \begin{cases} 1 & \text{se a ligação entre os nós } \mathbf{i} \text{ e } \mathbf{j} \text{ é estabelecida diretamente.} \\ 0 & \text{caso contrário.} \end{cases}$$

$$W_i = \begin{cases} 1 & \text{se a rota } \mathbf{r},\ \mathbf{r} \in \mathbb{R} \text{ é selecionada para suportar a comodidade apropriada.} \\ 0 & \text{caso contrário.} \end{cases}$$

Funções indicadoras (binárias)

Diferem das variáveis de decisão pois uma vez que valores tenham sido definidos para estas, as funções indicadores são parâmetros dados.

$$\delta^r_{ij} = \begin{cases} 1 & \text{se a ligação } \mathbf{(i,j)} \text{ aparece na rota } \mathbf{r}. \\ 0 & \text{caso contrário.} \end{cases}$$

$$\rho^r_i = \begin{cases} 1 & \text{se o nó } \mathbf{i} \text{ aparece na rota } \mathbf{r}. \\ 0 & \text{caso contrário.} \end{cases}$$

Variáveis (Taxas efetivas)

$$\lambda_{B_{ij}} = \sum_{r\in\mathbb{R}} W_r \lambda_r \delta^r_{ij} \quad \text{- taxa efetiva de mensagens na ligação } \mathbf{(i,j)}. \tag{3}$$

$$\lambda_{N_i} = \sum_{r\in\mathbb{R}} W_r \lambda_r \rho^r_i \quad \text{- taxa efetiva de mensagens no nó j.} \tag{4}$$

3.3 - Formulação matemática

Trata-se aqui de dar uma representação mais formal ao problema apresentado na seção 2.2 sob forma esquemática, usando as notações definidas na seção 3.2.

A função objetivo que é o retardo médio fim a fim na rede global é obtida pela soma dos retardos sofridos por uma mensagem em cada uma das comodidades, ponderada pela probabilidade de ocorrência desta comodidade. O retardo sofrido em uma comodidade será a soma dos retardos sofridos em cada uma das redes e pontes presentes na rota que suporta esta comodidade. O retardo médio sofrido nas redes é calculado à partir da expressão simplificada apresentada por Bertsekas e Gallager em [BERT 87] para redes tipo ETHERNET, sob as seguintes suposições: não persistente CSMA/CD; distribuição de chegada das mensagens Poisson com taxa λ_N e número finito de estações. O retardo médio sofrido por uma mensagem para ser processada em uma ponte pode ser obtido através do modelo M/G/1 [KLEI 75].

Seguindo as considerações acima obtém-se [FERR 89] para expressão do retardo a seguinte expressão:

$$T = \frac{1}{\lambda}\left\{ \sum_{i \in N}\left(d_i + \frac{a_i + b_i\lambda_{N_i}}{2 - c_i\lambda_{N_i}} \right)\lambda_{N_i} + \sum_{(i,j) \in L}\left(e_{ij} + \frac{f_{ij}\lambda_{B_{ij}}}{2 - 2e_{ij}\lambda_{B_{ij}}} \right)\lambda_{B_{ij}} \right\} \tag{5}$$

onde os coeficientes da expressão de retardo são dados por:

$$a_i = 4.62\,\tau_i$$

$$b_i = 2\tau_i\,\bar{X}_{N_i} + \overline{X^2_{N_i}}$$

$$c_i = 2\left(3.31\,\tau_i + \bar{X}_{N_i} \right)$$

$$d_i = \bar{X}_{N_i}$$

$$e_{ij} = \bar{X}_{B_{ij}}$$

$$f_{ij} = \overline{X^2_{B_{ij}}}$$

O problema pode então ser formulado como:

Problema P

$$Z = \text{Min} \left\{ \frac{1}{\lambda} \left[\sum_{i \in \mathbb{N}} \left(d_i \lambda_{N_i} + \frac{a_i \lambda_{N_i} + b_i \lambda_{N_i}^2}{2 - c_i \lambda_{N_i}} \right) + \right. \right.$$

$$\left. \left. + \sum_{(i,j) \in \mathbb{L}} \left(e_{ij} \lambda_{B_{ij}} + \frac{f_{ij} \lambda_{B_{ij}}^2}{2 - 2e_{ij} \lambda_{B_{ij}}} \right) \right] \right\} \qquad (6)$$

sujeito a:

$$\sum_{i \in \mathbb{A}} \sum_{j \in \overline{\mathbb{A}}} Y_{ij} \geq 1 \qquad \forall\ \mathbb{A} \subseteq \mathbb{N}, \quad \overline{\mathbb{A}} = \mathbb{N} - \mathbb{A} \qquad (7)$$

$$\sum_{(i,j) \in \mathbb{L}} Y_{ij} = \| \mathbb{N} \| - 1 \qquad (8)$$

$$W_r \leq Y_{ij} \qquad \forall\ (i,j) \in \mathbb{L}_r \qquad (9)$$

$$\sum_{r \in S_p} W_r = 1 \qquad \forall\ p \in \Pi \qquad (10)$$

$$\lambda_{B_{ij}} = \sum_{r \in \mathbb{R}} W_r \lambda_r \delta_{ij}^r \qquad \forall\ (i,j) \in \mathbb{L} \qquad (11)$$

$$\lambda_{N_i} = \sum_{r \in \mathbb{R}} W_r \lambda_r \rho_i^r \qquad \forall\ i \in \mathbb{N} \qquad (12)$$

$$\lambda_{B_{ij}} < \frac{1}{e_{ij}} \qquad \forall\ (i,j) \in \mathbb{L} \qquad (13)$$

$$\lambda_{N_i} < \frac{2}{c_i} \qquad \forall\ i \in \mathbb{N} \qquad (14)$$

$$W_r \in \{0,1\} \qquad \forall\ r \in \mathbb{R} \qquad (15)$$

$$Y_{ij} \in \{0,1\} \qquad \forall\ (i,j) \in \mathbb{L} \qquad (16)$$

As restrições 7 e 8 garantem que a topologia obtida na solução é uma árvore. A restrição 7 garante que o grafo é conexo, pois garante que para todo subconjunto $\mathbb{A}$ do conjunto de vértices existe ao menos um sucessor que não pertença ao subconjunto $\mathbb{A}$ (ou seja pertence ao seu conjunto complementar $\overline{\mathbb{A}}$). A restrição 8 garante que existem exatamente $\| \mathbb{N} \| -1$ arcos no grafo.

A restrição 9 impõe que uma rota só pode ser estabelecida se todas as ligações pelas quais ela passa tiverem sido selecionadas pelo modelo.

A restrição 10 diz que exatamente uma rota deve ser selecionada para cada comodidade que é suportada pela interligação de redes. É assumido para tanto que as rotas não se bifurcam.

As equações 11 e 12 expressam as taxas de fluxo efetivo de mensagens respectivamente nas ligações (pontes) e nos nós (redes).

As restrições 13 e 14 estabelecem limites superiores para o tráfego nas ligações (pontes) e nós (redes), respectivamente. Este limite superior representa o instante em que os dispositivos ficam saturados quando então o retardo nestes dispositivos tende para o infinito.

As retrições 15 e 16 simplesmente expressam a natureza binária das variáveis de decisão W_r e Y_{ij}.

4 - Obtenção de um limite inferior para a solução do problema simplificado usando Relaxação Lagrangeana

O problema de projeto de interligação de redes dado pelas expressões 6 a 16 é um problema de otimização combinatória não linear, NP completo. Nesta seção será desenvolvida uma relaxação Lagrangeana do problema que fornece um limite inferior para o valor da solução ótima. A diferença entre a solução viável de menor valor e o limite inferior mais alto é um limite superior para o erro introduzido pela seleção desta solução viável.

4.1 - Reformulação do problema

Com o objetivo de desenvolver o problema Lagrangeano, tornando-o separável nas variáveis Y_{ij} e W_r, o problema P é reformulado como:

Problema P

$$Z = \text{Min} \left\{ \frac{1}{\lambda} \left[\sum_{i \in \mathbb{N}} \left(d_i \lambda_{N_i} + \frac{a_i \lambda_{N_i} + b_i \lambda_{N_i}^2}{2 - c_i \lambda_{N_i}} \right) + \right. \right.$$

$$\left. \left. + \sum_{(i,j) \in \mathbb{L}} \left(e_{ij} \lambda_{B_{ij}} + \frac{f_{ij} \lambda_{B_{ij}}^2}{2 - 2e_{ij} \lambda_{B_{ij}}} \right) \right] \right\} \qquad (6)$$

sujeito a:

$$\sum_{i \in \mathbb{A}} \sum_{j \in \bar{\mathbb{A}}} Y_{ij} \geq 1 \qquad \forall\ \mathbb{A} \subseteq \mathbb{N},\ \bar{\mathbb{A}} = \mathbb{N} - \mathbb{A} \qquad (7)$$

$$\sum_{(i,j) \in \mathbb{L}} Y_{ij} = \| \mathbb{N} \| - 1 \qquad (8)$$

$$\sum_{r \in S_p} W_r = 1 \qquad \forall\, p \in \Pi \qquad (10)$$

$$\lambda_{N_i} < \frac{2}{c_i} \qquad \forall\, i \in \mathbb{N} \qquad (14)$$

$$W_r \in \{0,1\} \qquad \forall\, r \in \mathbb{R} \qquad (15)$$

$$Y_{ij} \in \{0,1\} \qquad \forall\, (i,j) \in \mathbb{L} \qquad (16)$$

$$\lambda_{N_i} \geq \sum_{r \in \mathbb{R}} W_r \lambda_r \rho_i^r \qquad \forall\, i \in \mathbb{N} \qquad (17)$$

$$\lambda_{B_{ij}} \geq \sum_{r \in \mathbb{R}} W_r \lambda_r \delta_{ij}^r \qquad \forall\, (i,j) \in \mathbb{L} \qquad (18)$$

$$\lambda_{B_{ij}} < \frac{1}{e_{ij}} Y_{ij} \qquad \forall\, (i,j) \in \mathbb{L} \qquad (19)$$

$$\lambda B_{ij} \geq 0 \qquad \forall\, (i,j) \in \mathbb{L} \qquad (20)$$

As igualdades em 11 e 12 podem ser substituidas pelas desigualdades 18 e 17, respectivamente, pois a função objetivo é estritamente crescente em relação as variáveis $\lambda_{B_{ij}}$ e λ_{N_i} e portanto em qualquer solução ótima $\lambda_{B_{ij}}$ e λ_{N_i} terão sempre o mínimo valor possivel. Esta alteração é significativa na relaxação Lagrangeana pois restringe o valor dos multiplicadores a valores negativos, permitindo uma convergência mais rápida.

As restrições 9 e 13 podem ser substituidas pelas restrições 19 e 20, como se demonstra pela seguinte análise:

Quando $Y_{ij} = 1$ - A restrição 9 será sempre satisfeita, e torna-se então desnecessária,
- as restrições 13 e 19 são iguais,
- a restrição 20 é supérflua.

Quando $Y_{ij} = 0$ - As restrições 9 e 11, na formulação original estabelecem que $\lambda_{B_{ij}} = 0$,
- de maneira equivalente as restrições 19 e 20, na formulação modificada, também impõe que $\lambda_{B_{ij}} = 0$,
- a restrição 13 é supérflua.

4.2 - Relaxação Lagrangeana

O problema relaxado é obtido multiplicando-se as restrições 17 e 18 por um vetor de multiplicadores $\mathbb{M} = [\ \mu_i,\ \eta_{ij}]$ e então adiciondo-as à função objetivo.

Problema Lagrangeano $P(\mathbb{M})$

$$Z(\mathbb{M}) = \text{Min} \left\{ \frac{1}{\lambda} \left[\sum_{i \in \mathbb{N}} \left(d_i \lambda_{N_i} + \frac{a_i \lambda_{N_i} + b_i \lambda_{N_i}^2}{2 - c_i \lambda_{N_i}} \right) + \right. \right.$$

$$\left. + \sum_{(i,j) \in \mathbb{L}} \left(e_{ij} \lambda_{B_{ij}} + \frac{f_{ij} \lambda_{B_{ij}}^2}{2 - 2e_{ij} \lambda_{B_{ij}}} \right) \right] +$$

$$\left. + \sum_{i \in \mathbb{N}} \mu_i \left(\lambda_{N_i} - \sum_{r \in \mathbb{R}} W_r \lambda_r \rho_i^r \right) + \sum_{(i,j) \in \mathbb{L}} \eta_{ij} \left(\lambda_{B_{ij}} - \sum_{r \in \mathbb{R}} W_r \lambda_r \delta_{ij}^r \right) \right\}$$

que pode ser reescrito como:

$$Z(\mathbb{M}) = \text{Min} \left\{ \frac{1}{\lambda} \left[\sum_{i \in \mathbb{N}} \left((d_i + \lambda\mu_i)\lambda_{N_i} + \frac{a_i \lambda_{N_i} + b_i \lambda_{N_i}^2}{2 - c_i \lambda_{N_i}} \right) + \right. \right.$$

$$+ \sum_{(i,j) \in \mathbb{L}} \left((e_{ij} + \lambda\eta_{ij})\lambda_{B_{ij}} + \frac{f_{ij} \lambda_{B_{ij}}^2}{2 - 2e_{ij} \lambda_{B_{ij}}} \right) +$$

$$\left. \left. + \sum_{r \in \mathbb{R}} \left(\sum_{i \in \mathbb{N}} \left(-\mu_i \rho_i^r \right) + \sum_{(i,j) \in \mathbb{L}} \left(-\eta_{ij} \delta_{ij}^r \right) \right) W_r \lambda_r \lambda \right] \right\} \qquad (21)$$

sujeito às restrições: 7, 8, 10, 14, 15, 16, 19 e 20

É conhecido da teoria da otimização (Geofrion [GEOF 74]) que para um dado vetor de multiplicadores $\mathbb{M} = [\ \mu_i,\ \eta_{ij}]$ com $\mu_i < 0$ e $\eta_{ij} < 0$ $(\forall\ i,\ j)$ o valor da função objetivo do problema Lagrangeano $Z(\mathbb{M})$ é um limite inferior para o valor da função objetivo do problema original Z, isto é, $Z(\mathbb{M}) \le Z$ $(\forall\ \mathbb{M} \le 0)$. Como o interesse neste ponto é obter o limite superior mais alto possível, é necessário encontrar o vetor $\mathbb{M}^*$ tal que:

$$Z(\mathbb{M}^*) = \underset{\mathbb{M}}{\text{Max}} \left\{ Z(\mathbb{M}) \right\}.$$

Qualquer procedimento computacional para se determinar $\mathbb{M}^*$ depende da habilidade em se resolver o problema Lagrangeano $P(\mathbb{M})$ rapidamente.

As relaxações introduzidas no problema P permitem separar o problema Lagrangeano em subproblemas que são mais fáceis de se resolver. Para qualquer vetor de multiplicadores $\mathbb{M}$ o problema Lagrangeano $P(\mathbb{M})$ pode ser decomposto em três subproblemas $P_1(\mathbb{M})$, $P_2(\mathbb{M})$, $P_3(\mathbb{M})$ tais que:

$$Z(\mathbb{M}) = Z_1(\mathbb{M}) + Z_2(\mathbb{M}) + Z_3(\mathbb{M})$$

onde:

Subproblema $P_1(\mathbb{M})$

$$Z_1(\mathbb{M}) = \text{Min} \left\{ \frac{1}{\lambda} \left[\sum_{i \in \mathbb{N}} \left((d_i + \lambda\mu_i)\lambda_{N_i} + \frac{a_i\lambda_{N_i} + b_i\lambda_{N_i}^2}{2 - c_i\lambda_{N_i}} \right) \right] \right\} \tag{22}$$

sujeito a restrição: 14

Subproblema $P_2(\mathbb{M})$

$$Z_2(\mathbb{M}) = \text{Min} \left\{ \frac{1}{\lambda} \left[\sum_{(i,j) \in \mathbb{L}} \left((e_{ij} + \lambda\eta_{ij})\lambda_{B_{ij}} + \frac{f_{ij}\lambda_{B_{ij}}^2}{2 - 2e_{ij}\lambda_{B_{ij}}} \right) \right] \right\} \tag{23}$$

sujeito às restrições: 7, 8, 16, 19 e 20

Subproblema $P_3(\mathbb{M})$

$$Z_3(\mathbb{M}) = \text{Min} \left\{ \sum_{r \in \mathbb{R}} \left[\left(\sum_{i \in \mathbb{N}} \left(-\mu_i \, \rho_i^r \right) + \sum_{(i,j) \in \mathbb{L}} \left(-\eta_{ij} \, \delta_{ij}^r \right) \right) W_r \lambda_r \right] \right\} \tag{24}$$

sujeito às restrições: 10 e 15

Fisicamente podemos interpretar a decomposição do problema de minimizar o retardo global na interligação de redes como a resolução de três subproblemas. O primeiro trata de minimizar o retardo nos nós (redes locais); o segundo se refere a minimização do retardo nas ligações (pontes); o terceiro trata de minimizar a distância total percorrida nas rotas selecionadas.

4.3 - Solução dos subproblemas Lagrangeanos

4.3.1 - Subproblema $P_1(\mathbb{M})$

Este subproblema pode ser decomposto em $||\mathbb{N}||$ subproblemas, um para cada nó (rede local). Onde o i-ézimo subproblema será:

$$Z_i^1(M) = \text{Min}\left\{ \frac{1}{\lambda}\left[(d_i + \lambda\mu_i)\,\lambda_{N_i} + \frac{a_i\lambda_{N_i} + b_i\lambda_{N_i}^2}{2 - c_i\lambda_{N_i}}\right]\right\} \qquad (25)$$

sujeito a: $0 \le \lambda_{N_i} < \dfrac{2}{c_i}$

Este é um problema de otimização não linear, cuja solução é dada por:

$$\frac{\partial Z_i^1(M)}{\partial \lambda N_i} = 0 \Rightarrow d_i + \lambda\mu_i + \frac{2a_i + 4b_i\lambda_{N_i}^{\bullet} - b_i c_i \lambda_{N_i}^{\bullet 2}}{(2 - c_i\,\lambda_{N_i}^{\bullet})} = 0$$

$$\lambda_{N_i}^{\bullet} = \begin{cases} \dfrac{2}{c_i}\left[1 - \sqrt{1 - \dfrac{2\,c_i(d_i + \lambda\mu_i) + a_i c_i}{2\,c_i(d_i + \lambda\mu_i) - 2b_i}}\,\right] & \text{p/ } \mu_i < -\dfrac{d_i}{\lambda} - \dfrac{a_i}{2\lambda} \\ 0 & \text{caso contrário} \end{cases}$$

Substituindo-se $\lambda_{N_i}^{\bullet}$ na expressão 25 obtém-se o valor ótimo de

$Z_i^1(M) = \tilde{Z}_i^1(M)$

A solução do subproblema $P_1(M)$ é dada então por:

$$\tilde{Z}_1(M) = \sum_{i \in N} \tilde{Z}_i^1(M)$$

4.3.2 - Subproblema $P_2(M)$

O subproblema $P_2(M)$ é um problema de otimização combinatória não linear e para ser resolvido será usado o seguinte fato:

Seja $\tilde{Y}_{ij}$ a solução de $P_2(M)$, então a otimização em relação a variável $\lambda_{B_{ij}}$ é separável sobre ligações (pontes).

Pela expressões 19 e 20 quando $\tilde{Y}_{ij} = 0 \Rightarrow \lambda_{B_{ij}} = 0$,

e quando $\tilde{Y}_{ij} = 1 \Rightarrow \lambda_{B_{ij}} = \lambda_{B_{ij}}^{\bullet}$

Onde $\lambda_{B_{ij}}^{\bullet}$ é a solução ótima do seguinte problema:

$$V_{ij} = \text{Min}\left\{ \frac{1}{\lambda}\left[(e_{ij} + \lambda\eta_{ij})\lambda_{B_{ij}}^{\bullet} + \frac{f_{ij}.\lambda_{B_{ij}}^{\bullet 2}}{2 - 2e_{ij}\lambda_{B_{ij}}^{\bullet}}\right]\right\} \qquad (26)$$

sujeito a: $0 \le \lambda_{B_{ij}}^{\bullet} \le \dfrac{1}{e_{ij}} Y_{ij}$

que é um problema de otimização não linear que pode ser resolvido tomando-se $\dfrac{\partial V_{ij}}{\partial \lambda B_{ij}} = 0$ e cuja solução é:

$$\lambda^{*}_{B_{ij}} = \begin{cases} \dfrac{1}{e_{ij}}\left(1 - \sqrt{1 - \dfrac{2\,e_{ij}(e_{ij} + \lambda\eta_{ij})}{2\,e_{ij}(e_{ij} + \lambda\eta_{ij}) - f_{ij}}}\right) & p/\ \eta_{ij} < -\dfrac{e_{ij}}{\lambda} \\ 0 \quad \text{caso contrário} \end{cases}$$

A substituição de $\lambda^{*}_{B_{ij}}$ na equação 26 leva então ao valor ótimo de V_{ij}.

O problema $P_2(M)$ pode agora ser reescrito sem perda de generalidade, como:

$$Z_2(M) = \text{Min}\left\{ \sum_{(i,j)\in \mathbb{L}} V_{ij}\, Y_{ij} \right\} \qquad (27)$$

sujeito às restrições: 7, 8 e 16

Se associarmos à ligação (i,j) o custo V_{ij}, então o problema $P_2(M)$ consiste simplesmente em encontrar uma árvore geradora mínima, para o que já existem algoritmos eficientes desenvolvidos, como os algoritmos de Kruscal e de Dijkstra [BOAV 79].

4.3.3 - Subproblema $P_3(M)$

Este subproblema pode ser reescrito como:

$$Z_3(M) = \text{Min}\left\{ \sum_{p\in\Pi} \sum_{r\in S_p} h_r W_r \right\} \qquad (28)$$

sujeito às restrições: 10 e 15

$$\text{onde:} \quad h_r = \left[\sum_{i\in \mathbb{N}} \left[-\mu_i\, \rho^r_i\right] + \sum_{(i,j)\in \mathbb{L}} \left[-\eta_{ij}\, \delta^r_{ij}\right] \right] \lambda_r$$

Este problema pode ser decomposto em $\|\pi\|$ subproblemas $P^p_3(M)$, um para cada par origem-destino, onde o p-ézimo problema será:

Subproblema $P^p_3(M)$

$$Z^p_3(M) = \text{Min}\left\{ \sum_{r\in S_p} h_r W_r \right\}$$

sujeito a:

$$W_r \in \{0,1\} \qquad \forall\ r \in S_p$$

$$\sum_{r \in S_p} W_r = 1$$

O subproblema $P_3^p(M\ )$ consiste em escolher uma rota de custo mínimo para o par origem-destino p, sobre o conjunto de todas as possíveis rotas que suportam este par. Consequentemente este subproblema é equivalente ao de se encontrar o caminho mínimo da origem ao destino deste par em um grafo. Para que se possa estabelecer esta equivalência é necessário expressar o coeficiente h_r em termos de custos sobre as ligações.

O custo da rota no subproblema $P_3^p(M\ )$ é proporcional ao tráfego na rota e é constituído de duas parcelas: uma referente a soma de $(-\eta_{ij})\lambda_r$ sobre todas as ligações usadas por esta rota, e a outra referente a soma de $(-\mu_i)\lambda_r$ sobre todos os nós usados por esta rota. Como não existe bifurcação de rota, pode-se, equivalentemente, considerar que a cada ligação (i,j) corresponderá um custo α_{ij} dado por:

$$\alpha_{ij} = -(\ \eta_{ij} + \mu_i\ )\ \lambda_r$$

Ou seja, o custo da ligação (i,j) engloba o custo da ligação (i,j) propriamente dita, mais o custo do nó (rede) onde se origina a ligação.

Evidentemente o custo da rota, referente a determinado par origem-destino, obtido somando-se o custo de todas as ligações existentes na rota, difere do custo real da rota por uma constante $(-\mu_{j\cdot})\lambda_r$ que é o custo referente nó destino $j^{\cdot}$. Como este custo é constante para um determinado par origem-destino, ele não interfere na escolha da rota ótima que suporta este par.

Observa-se que como os multiplicadores de Lagrange η_{ij} e μ_i são não positivos, então os custos de quaisquer ligações serão sempre não negativos, e portanto não existirá no grafo nenhum ciclo negativo.

Seja CM^p - o custo do caminho mais curto associado ao par origem -destino p, obtido usando-se α_{ij} como custos das ligações. Então pode-se relacionar $Z_3^p(M\ )$ e CM^p por:

$$Z_3^p(M\ ) = CM^p - \mu_{j\cdot}\lambda_r$$

onde: $j\cdot$ - índice correspondente ao nó destino do par origem-destino p

λ_r - taxa de tráfego correspondente ao par origem-destino p

Deste modo o problema $P_3(\mathbb{M})$ se relaciona ao problema de encontrar o caminho mais curto entre todos os pares origem-destino em um grafo, usando α_{ij} como custo das ligações. Para este problema já existem algoritmos eficientes desenvolvidos, como os algoritmo de Djkstra e Floyd [BOAV 79].

A solução de $P_3(\mathbb{M})$ será dada por:

$$Z_3(\mathbb{M}) = \sum_{p\in\pi} Z_3^p(\mathbb{M})$$

4.4 - Maximização do limite inferior

A função $Z(\mathbb{M})$ é uma função não diferenciável sobre $\mathbb{M}$, e portanto, técnicas clássicas de otimização não podem ser aplicadas para se resolver o problema

$$\underset{\mathbb{M}}{\text{Max}} \left\{ Z(\mathbb{M}) \right\}.$$

Vários procedimentos têm sido sugeridos na literatura com o objetivo de se fornecer valores aproximados para $\mathbb{M}^*$, tais como: métodos duais ascendentes, por ajuste de multiplicadores, de geração de colunas e otimização por subgradiente Gavishi, em vários trabalhos correlacionados com assunto [GAVI 88], [GAVI 89b] tem usado com sucesso esta última técnica, o que aliado ao fato de ser uma técnica de fácil programação, fez com que fosse a adotada neste trabalho.

O procedimento de otimização por subgradiente [GEOF 74] e [FISH 81] é um procedimento iterativo onde na k-ézima iteração é obtido o vetor $\mathbb{M}^k = [\mu_i^k , \eta_{ij}^k]$. Usando estes valores de multiplicadores, o problema Lagrangeano é resolvido. Baseados nesta solução, os multiplicadores a serem usados na próxima iteração são obtidos pela seguinte fórmula:

$$M^{k+1} = M^k + t_k \nabla(M^k)$$

Onde $\nabla(M^k)$ representa as direções do subgradiente dadas por:

$$\nabla(\mu_i^k) = \lambda_{N_i}^k - \sum_{r\in\mathbb{R}} \lambda_r W_r^k \rho_i^r$$

$$\nabla(\eta_{ij}^k) = \lambda_{B_{ij}}^k - \sum_{r\in\mathbb{R}} \lambda_r W_r^k \delta_{ij}^r$$

e t_k é o tamanho do passo. Foi mostrado que $Z(\mathbb{M}^k)$ converge para $Z(\mathbb{M}^*)$ quando a sequência de t_k satisfaz as seguintes condições:

$\lim t_k \longrightarrow 0$ e $\sum t_k \longrightarrow \infty$. Infelizmente tal sequência não é prática para propósitos computacionais e alternativamente t_k é aproximado pela expressão:

$$t_k = \frac{\bar{Z} - Z(M)}{\| \nabla(M^k) \|^2} . S_k$$

Onde $\bar{Z}$ é um limite superior para $Z(M^*)$ e S_k é um escalar entre 0 e 2 . Como $Z(M^*) \le Z$, qualquer solução viável para o problema do projeto topológico pode ser usada como $\bar{Z}$.

O esquema básico do algoritmo de otimização por subgradiente é apresentado a seguir.

Passo 1- Inicialização

A- Usando-se uma heurística determina-se um valor inicial para $\bar{Z}$;

B- Seleciona-se um valor inicial para M^0

C- Faz-se: - $M^* = M^0$, $k = 0$, contador melhorias = 0,

- contador iterações = 0, $Z(M^*) = 0$, $S_k = 2$

Passo 2- Resolução do problema relaxado

A- Faz-se contador melhorias $\longleftarrow$ contador melhorias + 1

B- Solução do subproblema $P_1(M^k)$

- Rotina de otimização não linear

- Obtém-se λN^k, $Z_1(M^k)$

C- Solução do subproblema $P_2(M^k)$

- Rotina árvore geradora mínima

- Obtém-se λB^k, Y^k, $Z_2(M^k)$

D- Solução do subproblema $P_3(M^k)$

- Rotina caminho mais curto

- Obtém-se W^k, $Z_3(M^k)$

E- Obtém-se $Z(M^k) = Z_1(M^k) + Z_2(M^k) + Z_3(M^k)$

Passo 3- Atualização dos parâmetros

A- Se $Z(M^k) > Z(M^*)$

então faz-se: $Z(M^*) \longleftarrow Z(M^k)$ e $M^* \longleftarrow M^k$

B- Se $\lambda N^k, \lambda B^k, Y^k, W^k$, sao viáveis para o problema P e tem custo menor que $\bar{Z}$

então faz-se: $\bar{Z} \longleftarrow Z(M^*)$

C- Se contador melhorias = contador melhorias limite

então faz-se: $S_k \longleftarrow \frac{S_k}{2}$, $M^k \longleftarrow M^*$

contador melhorias $\longleftarrow$ 0

retorna ao passo 2

D- Critérios de término (uma das seguintes condições)

.1 $S_k < \varepsilon_1$

.2 $\| \nabla(M^k) \|^2 < \varepsilon_2$

.3 $t_k < \varepsilon_3$

$$.4 \qquad \frac{\bar{Z} - Z(M^k)}{\bar{Z}} < \varepsilon_4$$

.5 Contador iterações > contador iterações limite

E- Se alguma condição de término é satisfeita
entao: TERMINA
caso contrário: Passe ao passo 4

Passo 4- Atualização dos multiplicadores

A- Faz-se $M^k \longleftarrow \text{Min}\left\{0,\ M^k + t_k \nabla(M^k)\right\}$

B- Faz-se $k \longleftarrow k + 1$

C- Faz-se Contador iterações ⟵ contador iterações + 1

D- Retorna ao passo 2

5 - Conclusões

A relaxação Lagrangeana descrita na seção anterior fornece um limite inferior para a solução ótima, desconhecida, do problema de projeto topológico de uma interligação de redes de computadores. Para este problema é difícil estabelecer a priori um limite teórico para o "gap" entre a solução ótima e a melhor solução Lagrangeana. Com a finalidade de ser capaz de avaliar empiricamente a qualidade destas soluções, bem como de se estabelecer uma metodologia de projeto, várias heurísticas, que fornecem soluções viáveis para o problema, têm sido desenvolvidas. O "gap" entre a melhor solução viável encontrada e o melhor limite inferior estabelecido é uma estimativa da distância desta solução viável para o valor ótimo e também por outro lado, da distância do melhor limite inferior para a solução ótima.

Em complementação ao trabalho aqui apresentado quatro heuristicas foram desenvolvidas. Duas baseadas somente em condições estruturais do problema, isto é, basicamente no fato do tráfego entre duas redes quaisquer ter que ser suportado por todas as redes e ligações existentes no caminho entre estas duas redes. A idéia por trás destas duas heuristicas é diminuir este caminho de modo que um número menor de redes e pontes tenham que suportar este tráfego externo. As outras duas são baseadas em procedimentos de troca de arcos a partir de uma solução viável inicial.

No presente estágio estas heurísticas, bem como a rotina de otimização por subgradiente estão sendo programadas em Turbo Pascal 5.0, já tendo sido testado um pequeno conjunto de exemplos, sem que nenhuma observação conclusiva possa ser feita.

Bibliografia

BACK 88 Backes F. ; **Transparent Bridges for Interconnection of IEEE 802 LANs**; IEEE Network; Vol. 2, No 1, pp 5-9, Janeiro 1988.

HART 88 Hart J.; **Extending the IEEE 802.1 MAC Bridge Standard to Remote Bridges**; IEEE Network; Vol. 2, No 1, pp 10-15, Janeiro 1988.

SINC 88 Sincoskie W.D. ; Cotton C. J.; **Extended Bridge Algorithms for Large Networks**; IEEE Network; Vol. 2, No 1, pp 16-24, Janeiro 1988.

BERT 87 Bertsekas D.; Gallager R.; **Data Networks** ; Prentice-Hall Inc; Englewood Cliffs, New Jersey.

KLEI 75 Kleinrock L.; **Queueing Systems, Vol II - Computer Aplications** ; John Wiley & Sons.

GERL 75 Gerla M.; **Aproximations and Bounds for Topological Design of Distributed Networks** ;Proceedings for the Fourth Data Communicatios Symposium; Quebec, pp 4.7-4.15, Outubro 1975.

BERN 83 Bernard G.; **Interconection of Local Computer Networks: Modeling and Optimization**; IEEE Transactions on Software Engineering Vol SE-9, No 4, Julho 1983.

LELL 85 Lellis L.L.; **Projeto e Avaliacao de Redes Locais e Interligadas**; Tese de Doutorado PUC-RJ, Dezembro 1985.

ALTI 89 Altinkemer K.; **Parallel Savings Heuristics for Designing Multi Center Tree Networks**; Working Paper, Purdue University, West Lafayette ; 1989.

GAVI 86a Gavish B.; **A General Model for the Topological Design Of Computer Networks**; GLOBECOM-86 pp 1584-1588, 1986.

GAVI 88 Gavish B.;**Topological Design of Computer Communication Networks**; Working Paper # 88-12 Vanderbilt University, Nashville, Tennessee; 1989.

GAVI 89a Gavish B.; Neuman I.; **A System for Routing and Capacity Assignment in Computer Communications Networs** ; IEEE Transactions on Communications Vol. 37, No 4, Abril 89.

GAVI 89b Gavish B.; Altinkemar K.; **Backbone Network Design with Economic Tradeoffs** ; Working Paper # 89-06 Vanderbilt University, Nashville, Tennessee; 1989.

BOAV 79 Boaventura P.O. ; **Teoria e Modelos de Grafos** ; Editora Edgard Blücher LTDA, São Paulo; 1979.

GEOF 74 Geofrion A.M.; **Lagrangean Relaxation And Its Uses In Integer Programming**; Mathematical Programming Study, Vol 2, pp 82 - 114, 1974.

FERR 89 Ferreira F$^{\underline{o}}$ V.J.M.; **Projeto Topológico Para Interligação de Redes Locais**; Dissertação de Mestrado, em elaboração, COPPE/UFRJ.

FISH 81 Fisher M.L.; **The Lagrangian Relaxation Method For Solving Integer Programming Problems**; Management Science, Vol 27, No 1, pp. 1-18; 1981.