PROTOCOLO DE COMUNICAÇÃO EM UM SISTEMA DISTRIBUÍDO BASEADO NUM BARRAMENTO PARALELO CENTRALIZADO.

CLAUDIO KIRNER (*)

JORGE LUIZ E SILVA (**)

Sumário:

Este artigo apresenta um Protocolo de Comunicação para um Sistema Distribuído Baseado num Barramento Paralelo Centralizado. Inicialmente descreve-se a Arquitetura do sistema em seguida aborda-se o protocolo propriamente dito, algumas questões envolvidas na sua elaboração, e, finalmente, os aspectos do software desenvolvido para sua implementação.

(*) Engenheiro Eletricista(EESC-USP, 1973), Mestre em Engenharia Eletrônica(ITA, 1978), Doutor em Engenharia de Sistemas e Computação (COPPE-UFRJ, 1986), Sistemas Operacionais, Arquitetura, Redes. Prof. Adjunto. DC-UFSCar - C.P. 384. 13560 - São Carlos - S.P. Fone (0162) 71-1100 - Ramal 143.

(**) Bacharel em Ciência da Computação (UFSCar - 1978), Mestre em Ciência da Computação (ICMSC - USP, 1986), Arquitetura, Redes. Prof. Assistente. DC-UFSCar - C.P. 384. 13560 - São Carlos - S.P. Fone (0162) 71-1100 - Ramal 143.

1-INTRODUÇÃO

O Projeto do Sistema Distribuído Baseado num Barramento Paralelo Centralizado [3] é decorrente de um primeiro projeto desenvolvido por um grupo de pesquisa do Departamento de Computação e Estatística da UFSCar, que consistiu na implementação de um Subsistema de Comunicação com conexão ponto a ponto [2] para servir como parte de um suporte para o desenvolvimento de Sistemas Distribuídos. O fator determinante na concepção do primeiro projeto foi única e exclusivamente o interesse acadêmico ,gerando alguns trabalhos de publicação e titulação, entre eles [4] e [7].

Com o conhecimento adquirido, o grupo passou a atuar na àrea de Redes e Sistemas Distribuídos de uma forma mais decisiva, adotando um enfoque que permitisse, ao mesmo tempo, uma especulação científica de vanguarda, e uma proximidade com as necessidades atuais e futuras do mercado.

Sendo assim, a partir de um trabalho que enfatizava uma rede estrela coincidente [5], optou-se pelo desenvolvimento de um barramento paralelo centralizado. Com o auxílio do CNPq, inicialmente através de uma bolsa de iniciação científica, conseguiu-se as condições para o estudo de alguns barramentos e a elaboração, em linhas gerais, da proposta do barramento centralizado e de suas interfaces de acesso [6]; em seguida, através de uma verba, conforme processo n.402081/85-CC, cedida ao projeto sob título de Projeto de Subrede de Comunicação para

aplicações em Tempo Real, coordenado pelo Prof. Claudio Kirner, partiu-se para a implementação de tres processadores de comunicação que, interligados pelo barramento paralelo centralizado "backplane", definem o subsistema de comunicação.

Uma vez implementado o Hardware da Rede, passou-se à implementação do Software de comunicação que permitisse a troca de mensagens a nível de Processador Operador .

Este trabalho descreve o protocolo de comunicação desenvolvido para a comunicação no Sistema Distribuído Baseado num Barramento Paralelo Centralizado.

Inicialmente mostra-se a arquitetura do sistema e o protocolo de comunicação entre os processadores Operadores. Em seguida são levantadas algumas questões envolvidas com a elaboração do Protocolo tais como : endereçamento, tratamento de exceções, formato dos pacotes, etc. Finalmente discute-se a implementação propriamente dita.

2-ARQUITETURA DO SISTEMA

O sistema constitui-se de um subsistema de comunicação composto por vários processadores de comunicação (PC), interligados através de um barramento paralelo disposto num "backplane", e de um conjunto de processadores de trabalho, denominados processadores operadores (PO), conforme a figura 1.

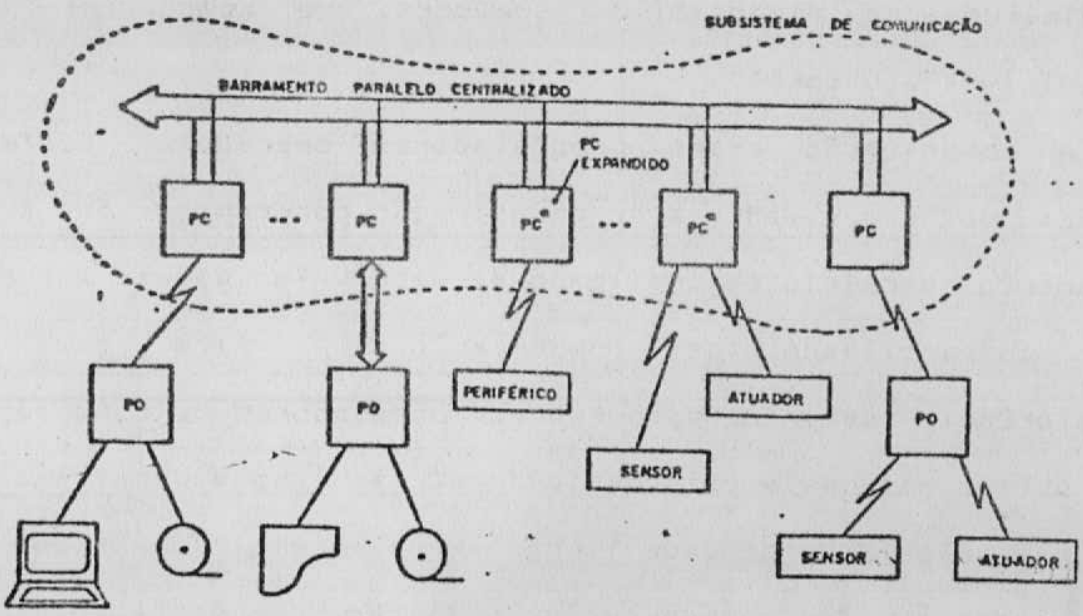

Figura 1 - Configuração Genérica do Sistema Distribuído

O subsistema de comunicação é constituido de um gabinete, contendo uma fonte confiável "no break", o barramento paralelo disposto no painel traseiro, e placas de circuito ligadas ao barramento, cada uma correspondendo a um processador de comunicação. Este conjunto poderá ser colocado dentro de uma sala bem protegida, de onde sairão todas as ligações para os processadores operadores remotos e, eventualmente, para dispositivos periféricos, sensores e atuadores que estiverem ligados a processadores de comunicação expandidos. Cabe salientar que um processador de comunicação expandido conterá , além das funções de um processador de comunicação normal, funções simples de controle de alguns

periféricos, tais como terminal de vídeo e impressora, funções de leitura de dispositivos sensores, e acionamento de dispositivos atuadores.

A comunicação entre processadores operadores ocorrerá através dos processadores de comunicação correspondentes e do barramento paralelo centralizado de alta velocidade, que é o meio compartilhado de comunicação. A alta taxa de transferência de informação entre processadores de comunicação será obtida mais pela existência de várias linhas paralelas, do que pela velocidade de cada linha. Para permitir que todos os processadores de comunicação utilizem o barramento, é necessário que cada um utilize-o no menor tempo possível, sem monopolizá-lo e para isso é necessário a existência de "buffers" de emissão e "buffers" de recepção em cada processador de comunicação. Esses "buffers" (figura 2) servirão para montagem de informação, a ser transferida, ou para a sua desmontagem para posterior utilização, ou retransmissão, em taxas menores, para um processador operador específico.

Um processador operador emissor fará a transferência de informação para um processador operador receptor da seguinte forma. Primeiro, o processador operador emissor transferirá a informação "byte" a "byte" para o processador de comunicação origem, a ele ligado. Esse processador de comunicação irá coletando as partes da informação, conforme a sua chegada, montando-as convenientemente no "buffer" de emissão. Quando a montagem terminar, o processador de comunicação origem acionará

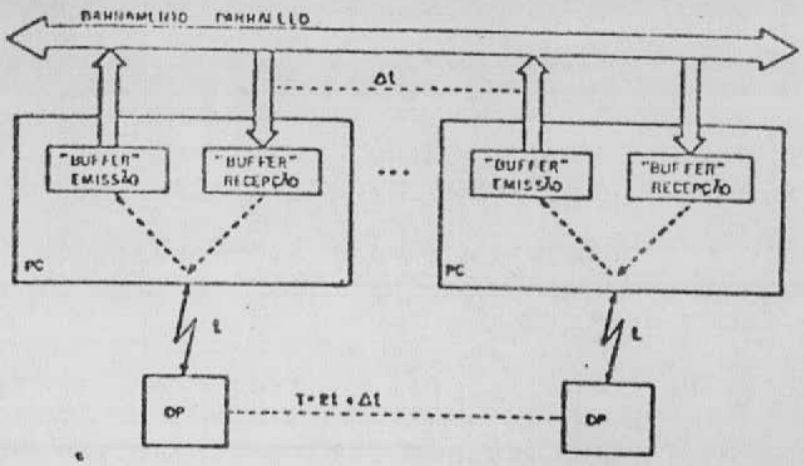

Figura 2 - Comunicação entre 2 Processadores Operadores

o seu circuito de disputa do barramento e, tão logo ganhe acesso, verificará se o processador de comunicação destino está pronto para receber a informação ("buffer" de recepção vazio). No caso do processador de comunicação destino não encontrar-se pronto o processador de comunicação origem desistirá do barramento e aguardará um tempo para nova tentativa. No caso do processador de comunicação destino estar pronto, então haverá transferência da informação, do "buffer" de emissão do processador origem, para o "buffer" de recepção do processador de comunicação destino (figura 2). Este último processador encarregar-se-á de desmontar a informação recebida e transferí-la ao operador, a ele ligado, num padrão de transferência compatível com a ligação.

3-PROTOCOLO DE COMUNICAÇÃO ENTRE OS PROCESSADORES OPERADORES

O Protocolo de comunicação entre os Processadoes Operadores corresponde basicamente ao processo de comunicação descrito na secão anterior.

Uma forma de tornar confiável esse processo foi manter as mensagens em "buffers" situados nos locais de origem, até que ocorressem as condições favoráveis para uma transferência sem riscos de rejeição ou perda. Isto foi implementado através de um protocolo de comunicação baseado em anúncios, ou seja: o anúncio seria utilizado para avisar o destino de que haveria mensagem pendente para ele em condições de ser transferida.

Tendo em vista que o tipo de conexão PC-PO adotada no projeto prevê a interrupção do processador de comunicação pelo processador operador, mas não o contrário, resolveu-se localizar a fila de anúncios no processador de comunicação destino, satisfazendo adequadamente as necessidades da rede e do processador operador destino. Desta maneira, sempre que um processador operador origem quiser transmitir uma mensagem para um processador operador destino, ele deverá armazenar a mensagem num "buffer" local, preparar um anúncio, remetê-lo à fila localizada no processador de comunicação destino, e esperar pelo atendimento de anúncio por iniciativa do processador operador destino.

Para a visualização do protocolo, tomou-se uma comunicação entre dois processos situados em processadores diferentes, onde

um corresponde a um cliente (processo emissor) e outro corresponde ao servidor (processo receptor). O processo emissor executará uma primitiva envia que bloqueará o processo emissor até que o processo receptor receba a mensagem . O processo receptor por sua vez , executará uma primitiva recebe que bloqueará o processo receptor até o recebimento da mensagem, caso ela exista, ou deixará o processo continuar tão logo tome conhecimento da inexistência de mensagem pendente. Haverá mensagem pendente sempre que a fila de anúncios do receptor não estiver vazia.

O protocolo utilizado entre processadores operadores, envolvendo os processadores de comunicação como elementos intermediários e sem levar em conta as várias possibilidades de falhas pode ser vista na figura 3. Cabe citar que, entre os processadores internos envolvidos na comunicação, existe um protocolo de nível mais baixo que será discutido no item cinco.

O funcionamento detalhado do protocolo é descrito em seguida.

O processo emissor localizado no processador operador, contendo uma mensagem já preparada fará uma chamada requisitando a execução da primitiva envia. Em seguida, o software que trata desta primitiva, considerando os parâmetros da chamada e a própria mensagem, fará a montagem do anúncio e o remeterá num pacote curto de controle para o processador de comunicação existente no local. Ali, o pacote receberá o

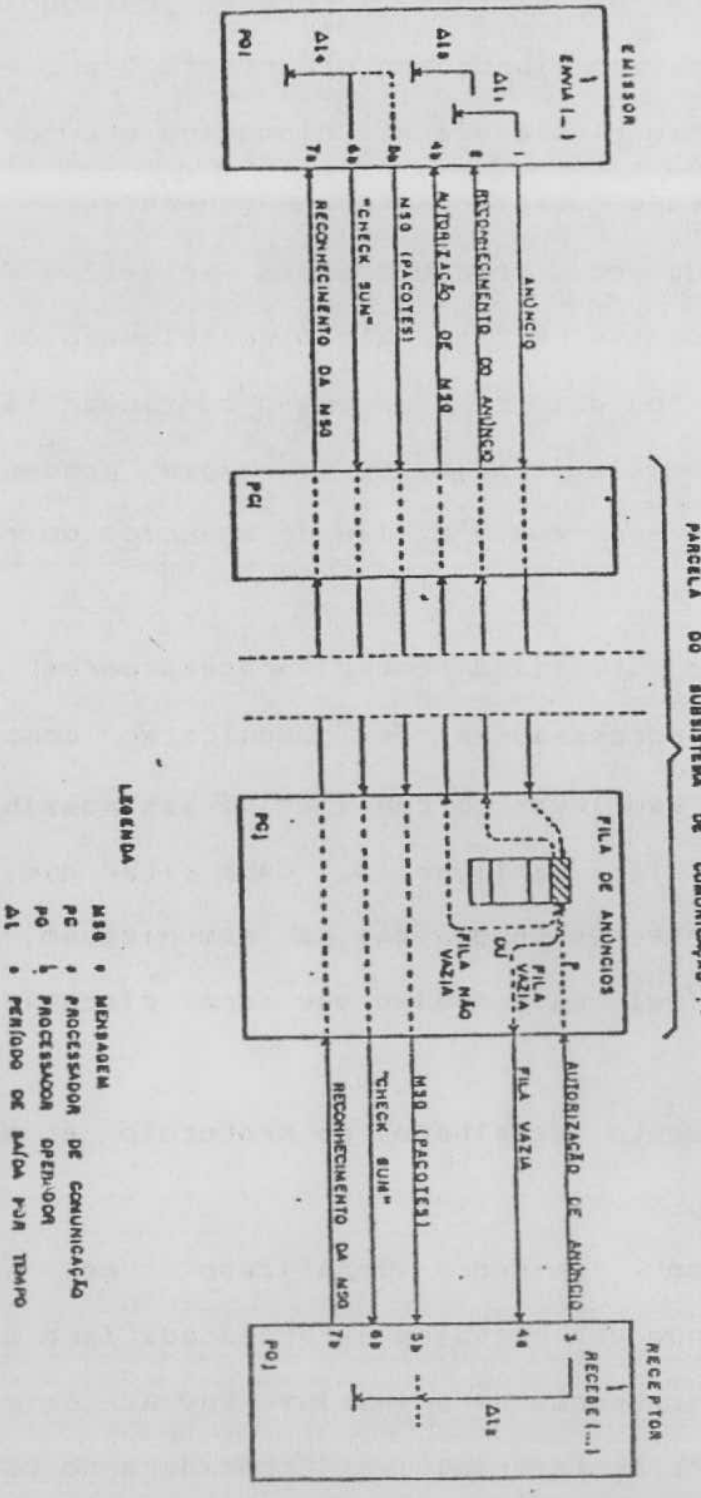

Figura 3 - Protocolo de Comunicação entre dois processadores operadores (usando anúncio)

tratamento complementar por parte do outro software, que neste caso consistirá do encaminhamento adequado do anúncio, visando alcançar o destino. Ao chegar ao processador de comunicação destino, o anúncio será colocado na fila de anúncios ali existente, e um pacote de reconhecimento positivo de anúncio será remetido de volta. Se, por acaso, numa situação inprevista o anúncio não puder ser colocado na fila, então será remetido o reconhecimento negativo de anúncio. Se receber um pacote de reconhecimento positivo de anúncio, o emissor ficará aguardando a chegado de autorização de mensagem, mas se receber um pacote de reconhecimento negativo, o emissor deverá informar o ocorrido ao nível imediatamente superior para as devidas providências.

Por outro lado , o processo receptor, em qualquer instante e independentemente do emissor, poderá fazer uma chamada da primitiva recebe. O software que trata desta primitiva remeterá então um pacote de autorização de anúncio para o processador de comunicação com o qual está ligado. Aí, o software existente no processador de comunicação, consultará a fila de anúncios alí localizada, verificando se existe ou não anúncios. Se a fila estiver vazia, um pacote de controle, indicando fila vazia, será remetido de volta ao processador operador, caso contrário, o anúncio será transformado num pacote de autorização de mensagem ,e transferido ao processador de comunicação ligado ao processador operador fonte ; alem disso também enviará um pacote ao processador operador destino, informando que foi liberado um anúncio para que este tome as decisões necessárias

para a recepção da mensagem.

Ao receber uma autorização de mensagem, o emissor que estava em estado de espera fará a decomposição da mensagem em diversos pacotes e os remeterá ao emissor. Durante esta tarefa, o "check sum" será calculado, e remetido por último, finalizando a trasnferência da mensagem. O receptor, recebendo os pacotes, remontará a mensagem, calculando também o "check sum" para poder compará-lo com o que vier do emissor. Se a comparação resultar em igualdade, então o receptor enviará ao emissor um pacote de reconhecimento positivo de mensagem, encerrando esta comunicação, caso contrário, haverá a emissão de reconhecimento negativo de mensagem, que será informado ao nível imediatamente superior.

A figura 3 também contém uma indicação de saída por tempo ("time-out") que, se acontecer , também será informado ao nível superior. A saída por tempo ficará latente sempre que a contagem de tempo a partir de um evento for disparada. A saída por tempo ocorrerá quando o tempo esgotar-se, ou será neutralizada quando um evento aceitável aparecer antes dela.

A maior parte das ações relacionadas com o protocolo são originadas e tratadas pelo software localizado nos processadores operadores. Caberá ao software, situado nos processadores de comunicação, cuidar na maior parte das vezes da transferência de pacotes de forma a permitir que fluam da origem para o destino. No entanto, percebe-se diretamente pelo protocolo que o anúncio e a autorização de anúncios, embora

originados nos processadores operadores, são tratados a nível de processadores de comunicação.

4-QUESTÕES ENVOLVIDAS COM A IMPLEMENTAÇÃO DO PROTOCOLO

Na implementação deste protocolo foram analisados aspectos relacionados diretamente com a comunicação entre processos num sistema distribuido.

Existem vários mecanismos e muitas variações na forma de executar troca de mensagens [4], dependendo de como são realizados o sincronismo, o endereçamento, a utilização de "buffers", etc.

Os mecanismos são aqueles que utilizam primitivas básicas como envia e recebe, dispostas convenientemente nos processos de maneira a permitirem a sincronização e comunicação. O comportamento dos processos emissor e receptor, ao executarem as respectivas primitivas, dependerá do tipo de sincronização e do endereçamento adotado.

Na etapa de sincronização, as primitivas podem ser bloqueantes (síncronas) e não bloqueantes (assíncronas). Se uma primitiva é bloqueante significa que o processo que a estiver executando ficará bloqueado até que a operação correspondente seja executada por completo. Se uma primitiva é não bloqueante, o processo que a estiver executando continuará em andamento tão logo ela seja concluída localmente.

No protocolo implementado, optou-se pelo mecanismo de sincronização Envia bloqueante-Recebe bloqueante, devido a sua simplicidade de implementação.

As primitivas envia e recebe podem ser definidas de forma resumida como :

envia (destino, mensagem, outros parâmetros);
recebe (origem, mensagem, outros parâmetros).

Um aspecto importante nessa estrutura de comunicação refere-se à técnica usada para a identificação do emissor/receptor da mensagem, o que é possível através de esquemas específicos de endereçamento.

O endereçamento pode ser: implícito, correspondendo ao uso de ligações previamente estabelecidas entre os processos do sistema; explícito, quando o endereçamento é baseado na utilização de nomes, endereços, e outros atributos do processo, explicitamente referenciados nas primitivas de comunicação; simétrico, que caracteriza-se por exigir que as primitivas utilizadas numa comunicação contenham a indicação recíproca da origem e do destino da mensagem; e assimétrico, que exige que exista somente a identificação do destino da mensagem.

O endereçamento utilizado no protocolo foi definido como sendo implícito na conexão do Processador Operador com o Processador de Comunicação , explícito na comexão entre

Processadores de Comunicação e entre Processadores Operadores, e finalmente assimétrico com relação à primitiva de comunicação.

Alem do endereçamento especificado acima , existe um modo de endereçamento por difusão , que é um endereço comum a um conjunto de processadores com objetivo de difundir uma mensagem comum a esses processadores, também implementado aqui.

Um dos aspectos também analisados na implementação deste protocolo foi a utilização de "buffers", com o objetivo fundamental de controlar situações de congestionamento e o fluxo de informação na rede.

Mais especificamente com relação ao controle do fluxo de informação , para evitar que um receptor receba mensagens mais rapidamente do que possa consumir, uma primeira abordagem consistiria em dotar o processador, que contém o processo receptor , com "buffers" em quantidade suficiente para absorver todas as mensagens que porventura fossem dirigidas simultaneamente para ele. Uma outra alternativa seria manter a mensagem em "buffers" situados nos seus locais de origem sob controle dos processadores respectivos, e enviar aos processos receptores somente os resumos dessas mensagens, aqui denominados anúncios. A terceira possibilidade seria encaminhar as mensagens, correndo o risco de serem aceitas ou não.

Por não envolverem perda de mensagem, as duas primeiras funcionam bem, inclusive com cargas elevadas, enquanto que a terceira funciona muito bem com cargas baixas mas sofre um alto índice de degradação em situações de cargas elevadas.

Na implementação do protocolo escolheu-se a segunda opção que exigiu a utilização de anúncios.

Finalmente, nesta análise , foram considerados os problemas relacionados com as falhas na comunicação. Isso porque um sistema distribuído deve tolerar falhas físicas e lógicas de forma que, ocorrendo algumas delas, o sistema continue operando mesmo em situação de degradação. Nesse sentido, o sistema de comunicação deve possuir um esquema de detecção de erros capaz de permitir que somente as mensagens corretas sejam encaminhadas aos seus destinos, ou quando necessário, abandoná-las em função do tempo de reposta estabelecido por um limite de tempo ("time-out") para a realização da comunicação.

5-IMPLEMENTAÇÃO DO PROTOCOLO

O software gerado para a implementação do protocolo compreende basicamente as rotinas ENVIA E RECEBE suportadas por um software de nível inferior denominado PROGRAMA SUPERVISOR, que controla o fluxo de informação na rede.

O Programa Supervisor será instalado em cada processador de comunicação individualmente, de tal forma que o controle do fluxo de informação fique distribuído entre os processadores de comunicação ligados ao barramento.

Este software utiliza uma estrutura de informação baseada em filas e "buffers" de mensagens, conforme figura 4.

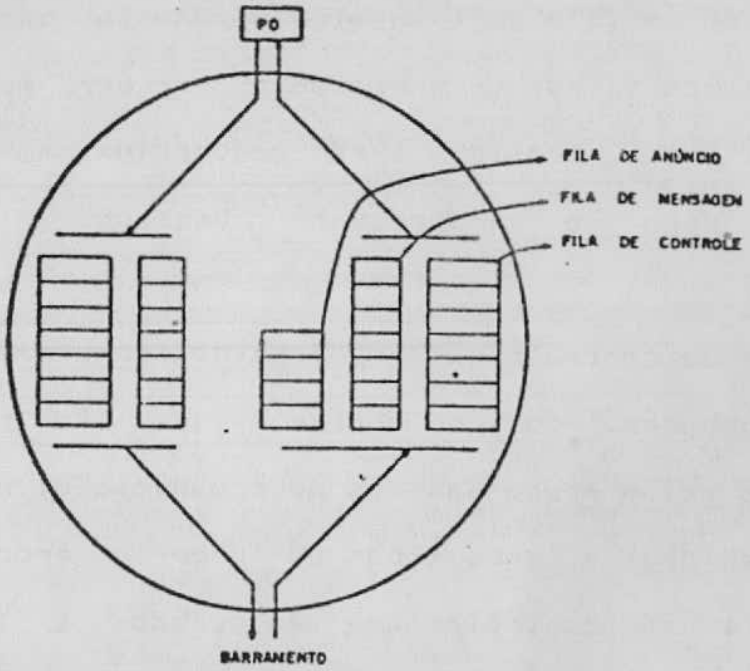

Figura 4 - Estrutura das Filas no Processador de Comunicação

Na estrutura acima tem-se duas filas de recepção e transmissão de mensagens, uma associada ao processador operador e outra associada ao barramento de comunicação; duas filas de recepção e transissão de controle, da mesma forma uma associada ao processador operador e outra ao barramento de comunicação; e uma fila de anúncios. As filas de recepção e transmissão de mensagens podem armazenar apenas um pacote de mensagem, enquanto que as filas de controle e de anúncio podem armazenar pacotes de controle.

As filas de mensagens serão utilizadas tanto para

recepção quanto para transmissão das mesmas. Desta forma, a fila de recepção e transmissão ligada ao processador operador receberá uma mensagem do processador operador e em seguida será convertida em fila de transmissão para o barramento. A fila de mensagem ligada ao barramento receberá mensagens do barramento e em seguida será convertida em fila de transmissão para o processador Operador, e assim sucessivamente.

As filas de controle são tratadas de forma diferente das filas de mensagens, pois há controles que são recebidos e retransmitidos pelos processadores de comunicação, há controles que são recebidos e consumidos no proprio processador de comunicação, e há controles que são gerados e transmitidos a partir de um processador de comunicação específico, além de ser uma fila capaz de armazenar vários pacotes simultaneamente.

As filas de controle, portanto, foram implementadas para armazenar pacotes , e cada fila foi dividida em dois conjuntos de elementos; no caso da fila associada ao processador operador, um conjunto para recepção do processador operador e outro conjunto para transmissão no barramento; no caso da fila associada ao barramento, um conjunto para recepção do barramento e outro conjunto para transmissão ao processador operador. Os conjuntos são dinâmicos, portanto variável no seu limite interno, porém as filas possuem um limite máximo de capacidade.

A fila de anúncios é uma fila de controle específica para

receber anúncios, mas não é tratada como as filas de controle descritas acima pois é de uso exclusivamente interno ao processador de comunicação.

A cada fila estão associados variáveis de estado que descrevem a cada momento informações relacionadas com as filas como: liberadas, não liberadas, transmissão, recepção, etc.

O programa supervisor, portanto, faz uso das filas para o controle do fluxo de informação no subsistema de comunicação, servindo-se para tanto de um conjunto de pacotes de mensagens específicos para a comunicação a nível de programa supervisor.

Há dois tipos de pacotes que circulam pela rede: os pacotes curtos de controle, e os pacotes longos de informação. Um pacote curto de controle é formado unicamente por um cabeçalho de seis campos, enquanto um pacote longo de informação contém, além desse cabeçalho, um campo adicional de informação. O campo de informação será responsável pela transferência da mensagem ou de parte dela entre dois processadores operadores.

Os campos de um pacote, referentes ao cabeçalho são os seguintes:

DESTI.	ORIG.	TIPO	OPERA.	TAMAN.	N.PAC.	inf.

Os campos DESTINO E ORIGEM são utilizados para endereçar ou identificar os pontos da rede formados pelo par PC-PO. Ao chegar no ponto destino, o pacote poderá ser consumido no PC ou remetido ao processador operador, dependendo do seu conteúdo.

O campo TIPO é utilizado para caracterizar a função do pacote. Ele indicará o uso do pacote para transportar: anúncio, reconhecimento positivo de anúncio, reconhecimento negativo de anúncio, reconhecimento positivo de mensagem, reconhecimento negativo de mensagem, autorização de anúncio, autorização de mensagem, indicação de fila de anúncios vazia, indicação de que foi liberado um anúncio, 'check sum', e mensagem.

O campo OPERAÇÃO apresenta vários significados. Por exemplo, quando associado a um anúncio, ele estará indicando a operação a ser realizada pelo processador destino que poderá ser o servidor...

O campo TAMANHO tem sua aplicação no pacote de anúncio, permitindo ao processador operador destino fazer a previsão e controle do 'buffer' necessário para receber a mensagem por ocasião de seu atendimento. No caso do pacote de 'check sum', este campo conterá outra parcela do mesmo.

O campo N. DE PACOTE é utilizado para numerar os pacotes de mensagem emitidos, permitindo a remontagem da mensagem independentemente da ordem de recebimento dos mesmos, mas também pode ser utilizado como N. DE SEQUÊNCIA para indicar a insistência de atendimento de um determinado anúncio, ou o abandono do anúncio corrente e atendimento de um novo.

Alem do programa supervisor, existem rotinas que fazem a recepção, tanto na conexão com o processador operador quanto na conexão com o barramento de comunicação. A recepção é feita por interrupção nos dois casos.

No caso da conexão com o processador operador esta sendo implementado um software semelhante ao HDLC [8], com uma adaptação para troca de informações por "bytes", uma vez que a conexão processador de comunicação e processador operador foi implementada através de um canal serial comum.

Os detalhes da implementação tanto da rotinas ENVIA E RECEBE quanto do PROGRAMA SUPERVISOR serão publicados oportunamente.

6-CONCLUSÃO

Como citado anteriormente, foram implementados tres processadores de comunicação, alem do software das primitivas e do programa supervisor, especificados em PASCAL.

O projeto permite muitas variações em todos os seus níveis: a nível de hardware, pode-se sofisticar o barramento centralizado, o processador de comunicação e as interfaces de conexão com outras redes; a nível de software básico pode-se conceber vários conjuntos de primitivas, apropriadas a aplicações diferentes; da mesma maneira, a nível de sistema operacional e a nível de aplicação, o sistema poderá ir de situações bastante simples até as mais complexas.

É intenção do grupo atuar nesse projeto até o nível de sistemas operacionais, abrangendo a área de aplicações em tempo real.

BIBLIOGRAFIA

[1] ACAMPORA, A.S. & HLUCHYJ, M.G. - "A New Local Area Networks Architecture Using a Centralized Bus". IEEE Communications Magazine, august, 1984, pg.12-21.

[2] KIRNER, C. - "Suporte para Desenvolvimento de Sistemas Distribuídos : uma implementação". In: Anais do IV Congresso da Sociedade Brasileira de Computação - XI SEMISH, Viçosa, MG, SBC, jul. 1984. p.109-121.

[3] KIRNER, C. & MARQUES, E. - "Suporte para Desenvolvimento de Sistemas Distribuídos Baseado num Barramento Paralelo Centralizado". In: Anais do V Congresso da Sociedade Brasileira de Computação - XI CLAI - XII SEMISH, Porto Alegre , RS, SBC, jul. 1985. p. 423-435.

[4] KIRNER, C. - "Desenvolvimento de Suporte Básico para Sistemas Operacionais Distribuídos". Tese de Doutorado, Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, RJ, 1986. 232p.

[5] LINDSAY, D.C. - "Local Area Networks: Bus and Ring Vs. Coincident Star". Computer Communication Review, july/october, 1982, p. 83-91.

[6] MARQUES, E. - "Projeto de um Barramento Compacto para Sistemas Distribuídos". Publicação interna, Departamento de Computação e Estatística - UFSCAR, 1984.

[7] SILVA, J.L. - "Análise e Projeto de um Subsistema de Comunicação para uma Rede de Computadores". Tese de Mestrado, Instituto de Ciências Matemáticas de São Carlos, USP, São Carlos, SP, 1986. 110p.

[8] TANENBAUM, A.S. - Computer Networks. Englewood Cliffs, N.J. Prentice-Hall, Inc., 1981. 517p.

[9] TEIXEIRA, C.A.C. & TREVELIN, L.C. - "Protocolo de Acesso e Superação de Falhas em Redes com Topologia Estrela Coincidente ". In: Anais do II Simpósio sobre Redes de Computadores. Campina Grande, Paraíba, UFPb, abr. 1984. p.16.1-16.15.