



INFORMÁTICA 82

XV CONGRESSO NACIONAL DE INFORMÁTICA
II FEIRA INTERNACIONAL DE INFORMÁTICA

ANAIS

004.06
C749a



INFORMÁTICA 82

XV CONGRESSO NACIONAL DE INFORMÁTICA
II FEIRA INTERNACIONAL DE INFORMÁTICA

ANAIIS

DESCRIÇÃO DE HARDWARE E SOFTWARE DA REDE LOCAL, REDPUC

Luiz Fernando Gomes Soares/Daniel A. Menascé/Carlos Henrique Cavalcanti Correa
Fernando Jefferson de Oliveira

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente – Rio de Janeiro, RJ

Palavras-chave: Rede local, protocolo de acesso, protocolo de transporte, interface de acesso a embarramento, núcleo de sistema operacional distribuído.

A rede local REDPUC é composta de um conjunto de módulos processamento-armazenamento interconectados através de dois embarramentos globais com capacidade de 4 Mbps. O controle de acesso ao embarramento é distribuído e os três níveis de protocolo de acesso são implementados em um processador de comunicação, PC, baseado em um microprocessador. Através deste protocolo o PC é responsável pela manutenção da integridade do sistema, controle de erro e controle de fluxo na comunicação entre os processos. Este artigo apresenta uma breve descrição do protocolo de acesso ao embarramento. A seguir é apresentado o hardware do PC e dos módulos de processamento-armazenamento a eles ligados.

A especificação de um núcleo de um sistema operacional para a REDPUC é abordada. O artigo conclui com uma aplicação desta rede local para a realização de uma central de comutação de pacotes.

1. INTRODUÇÃO

Redes locais foram desenvolvidas tanto para viabilizar o compartilhamento de recursos (discos de grandes capacidades, impressoras rápidas, "gateways" para redes geograficamente distribuídas, etc.) quanto na construção de máquinas de arquitetura distribuída, visando uma melhoria no desempenho do sistema.

O uso de múltiplos processadores, principalmente em ambientes de tempo real, é facilmente justificável, dado o poder computacional de baixo custo oferecido pelos microprocessadores e sua crescente melhora em termos de custo/desempenho. Arquiteturas de múltiplos processadores têm aplicação favorecida em sistemas que exigem grande disponibilidade, altos requisitos de vazão, tempos de respostas garantidos e baixo, alto grau de modularidade, e onde as tarefas podem ser executadas de modo concorrente.

Uma rede local é composta por um conjunto de módulos de processadores (MP) interligados por um sistema de comunicação (SC), conforme mostra a figura 1. A mínima interferência na execução de tarefas paralelas vai permitir a obtenção de sistemas com elevado desempenho. A possibilidade de utilização em

larga escala de um pequeno número de elementos de "hardware" e "software" é responsável pelo alto grau de modularidade do sistema. A não existência de elementos sem os quais o sistema pararia totalmente lhe confere grande confiabilidade. Além disso, não existe nenhuma restrição inerente à estrutura que impeça o crescimento do sistema, o que lhe confere a característica de expandibilidade.

O sistema de comunicação, que provê a infraestrutura para a troca de informações entre os MP's, é formado por um arranjo topológico interligando através de enlaces físicos as interfaces de acesso à rede, e por um protocolo de comunicação. As interfaces algumas vezes são compostas por processadores de comunicação (PC) que são os responsáveis pela realização do protocolo interno à rede já mencionado.

As topologias mais comuns são as redes em estrela (1,2 e 3), redes em anel (4,5,6,7,8 e 9), e redes em embarramento global que podem operar por contenção (10,11,12,13,14,15,16 e 17) ou por passagem de "token" (18,19 e 21). Uma descrição de outros tipos de topologias, bem como uma taxonomia para suas classificações pode ser encontrada em (20).

A REDPUC é uma rede local em desenvolvimento no Laboratório de Engenharia e Sistemas de Computação da Pontifícia Universidade Católica do Rio de Janeiro. Ela é uma máquina de arquitetura distribuída utilizando um sistema físico de interconexão composto por dois embarramentos globais seriais, compartilhados pelos diversos módulos processadores (até um total de 254), como mostra a figura 2.

A REDPUC foi concebida para aplicações que requerem uma dispersão do sistema de computação até 1000m e um alto desempenho, o que levou à inclusão de um "hardware" externo aos processadores de aplicação (módulos de expansão ME's somado ao módulo básico MB) para o controle do sistema. O gerenciamento do hardware em tempo de execução em sistemas de múltiplos processadores é difícil de ser conseguido e afasta o "software" de seu objetivo primário que são as aplicações do sistema. Isto justifica a inclusão de sistemas externos de controle de "hardware" de modo a liberar o software de se concentrar em tais problemas. Os mais críticos mecanismos de controle de um sistema distribuído são certamente aqueles envolvidos na comunicação entre processos e entre processadores. Devido a isto como em (11,14 e 19), a REDPUC incorporou à sua interface de acesso ao embarramento um processador de comunicação (PC) para a realização de todo o protocolo de comunicação, ao contrário de outras redes (10,12,13,15, 17 e 18) que optaram por uma interface de comunicação de mais baixo custo.

Os dois embarramentos, em cabo coaxial, operam independentemente, cada um a 2Mbps na primeira versão. De forma a garantir a integridade da rede, o PC pode ser desligado de um ou ambos os embarramentos quando ocorrem condições de falhas irrecuperáveis no módulo processador.

O item 2 do artigo descreve brevemente o protocolo de acesso ao embarramento e a arquitetura de "hardware" do processador de comunicação. O item 3 descreve a arquitetura de um processador de aplicação e apresenta de forma resumida o núcleo de comunicação distribuído da rede. O artigo conclui com uma aplicação da rede local REDPUC para a realização de uma Central de Comutação de Pacotes.

2. O PROCESSADOR DE COMUNICAÇÃO

2.1 Protocolo de Acesso ao Embarramento

O protocolo de acesso ao embarramento (PAE) compreende três níveis independentes entre si: o nível físico, o nível de embarramento e o nível de transporte (22).

O nível físico especifica as características físicas e elétricas do PAE, tais como nível de tensão, conexões físicas e métodos de transmissão. As características importantes deste nível são: transmissão serial bit a bit, síncrona, full duplex em um sistema de comunicação do tipo difusão ("broadcast") usando codificação Manchester para possibilitar o controle distribuído da sincronização nos embarramentos.

O nível de embarramento é responsável pelo controle distribuído do acesso ao embarramento. Cada MP deve receber uma permissão explícita para usar o embarramento antes de começar a transmitir. Esta permissão é um "token" que é passado de MP a MP

numa ordem cíclica. Mesmo quando o MP não tem mensagens a serem transmitidas ele deve enviar uma mensagem vazia (que serve como atestado de vida para fins de reconfiguração) contendo o "token". Um MP que deixa de enviar mensagem durante três ciclos consecutivos é removido de cada lista de controle de ciclo de cada MP. Cada MP mantém uma lista de controle de ciclo que indica o "status" de cada um deles, bem como o do ciclo corrente.

Ao final de cada ciclo existe um intervalo, chamado intervalo de contenção, para ser usado pelos MP's que desejem se inserir no ciclo. Este intervalo pode terminar de três maneiras diferentes: nenhuma transmissão ocorre depois de um período de time-out, um MP transmite com sucesso e é reinserido em cada lista de controle dos outros MP's, ou então mais de um MP transmite e então há colisão. Neste último caso um procedimento de recuperação se inicia, basicamente incluindo todos os MP's fisicamente conectados em todas as listas de controle. Aqueles que estão fisicamente conectados, mas inativos, serão excluídos de cada lista de controle de ciclo nos próximos ciclos. A racionalidade por detrás deste procedimento simples de resolução vem do fato que é esperado que a probabilidade de mais de um MP tentar voltar ao estado ativo durante um mesmo ciclo seja baixa. O procedimento de criação das listas de controle de ciclo é idêntico ao procedimento mencionado acima.

Cada mensagem a nível de embarramento é endereçada a outro MP. Mensagens incorretas (aquelas com FCS incorreto) são descartadas. Mensagens corretas têm o conteúdo de informação passada ao nível de transporte nos MP's de destino. Nos MP's restantes apenas o cabeçalho do nível de transporte é passado a este nível. Este cabeçalho será usado na realização de controle de erro e de fluxo associado ao nível de transporte. Desta forma, a natureza de difusão do meio de comunicação é explorada para o envio de mensagens de controle a processos distintos em uma única mensagem a nível de embarramento.

Entre outras funções deste nível de protocolo estão: transparência de transmissão, detecção de erro (a recuperação através de time-out e retransmissão é realizada no nível de transporte), manutenção do ciclo mesmo em presença de falhas.

O nível de transporte realiza as conexões entre processos em execução nos diversos MP's. O nome do processo é independente da rede tornando assim a migração de processador e a reconfiguração em caso de falhas bastante simples. Cada conexão é uma associação temporária, com controle de erro e de fluxo, entre processos. Um protocolo de bit alternado modificado é usado em cada conexão. Devido à natureza cíclica da comunicação, um único bit pode ser usado na realização do controle de erro e de fluxo em cada conexão. O endereçamento neste nível é realizado de processo a processo, contudo, endereçamento global e a grupo de processos também é permitido.

2.2 O Hardware do Processador de Comunicação

A figura 3 apresenta um diagrama simplificado do Processador de Comunicação (PC).

O primeiro PC desenvolvido se comunica com o seu computador hospedeiro (módulo de aplicação) via

DMA a uma taxa de 400 k bytes/seg. O processador utilizado foi o Intel 8086. A ligação aos dois embarramentos globais foi realizada através de dois controladores de comunicação 2652-1 da Signetics e de circuitos para sincronismo, codificação e decodificação bifásica. A taxa de transmissão em cada um dos embarramentos é de 2M bps, o que dá uma taxa de transmissão de 4M bps. A ligação dos controladores de comunicação com o processador 8086 é realizada através de uma memória partilhada não ligada à barra principal do sistema 8086, para uma melhora no desempenho (no futuro pretende-se chegar à velocidades de transmissão maiores com o uso de fibras óticas).

Embora, na sua primeira versão, o PC se ligue ao seu computador hospedeiro (um sistema 8088 na primeira versão) via DMA, esta conexão pode ser facilmente realizada serialmente para conexão com outros sistemas, inclusive com sistemas de pouca inteligência, uma vez que o PC pode ser programado. Mais ainda, como em (14), o PC poderá ser ligado a mais de um computador hospedeiro, partilhando assim o custo da interface.

Nas referências (21 e 22) pode ser encontrado a especificação detalhada da arquitetura da software e hardware do Processador de Comunicação.

3. O MÓDULO PROCESSADOR

O Módulo Processador (MP) é formado pelo PC e pelo processador de aplicação (MB + ME na figura 2). O processador de aplicação além de possuir recursos locais de processamento e armazenamento (módulo básico - MB) responsável, entre outras coisas, pelo processamento, pela memória mínima local, e pelo gerenciamento da comunicação entre módulos processadores, possui outros recursos dedicados à realização de subtarefas específicas (módulo de expansão - ME).

Para aplicação específica do REDPUC na realização de uma Central de Comutação de Pacotes (23) foi desenvolvido um módulo básico baseado no processador 8088 da Intel. A figura 4 apresenta um diagrama em bloco de tal módulo.

A memória ROM contém o núcleo de comunicação (NC), que é responsável pela comunicação entre processos residentes ou não no módulo. A memória RAM é o espaço de armazenamento da área de dados do NC.

Toda comunicação realizada através do NC é feita através de troca de mensagens. A forma pela qual os processos usuários comunicam entre si não é influenciada pela alocação dos processos ao módulo. Desta forma o mapeamento dos processos deixa de ser um vínculo do sistema já que as interfaces com o NC são uniformes. Esta característica é fundamental para permitir a modularidade de processos entre módulos, o que por sua vez é fundamental em uma máquina dotada de possibilidade de reconfiguração. Deve-se levar em consideração no entanto, que o custo de comunicação entre processos residentes em um mesmo módulo é menor do que o custo de comunicação que envolve processos em módulos distantes.

Simplificadamente, as operações primitivas de comunicação através do NC são as seguintes:

- 1) Envia mensagem (processo a processo)
- 2) Espera mensagem (de qualquer processo)

- 3) Espera mensagem de tipo específico
- 4) Procura mensagem de tipo específico
- 5) Envia mensagem para grupos (global ou grupo de processos)
- 6) Espera mensagem de tipo específico de grupos (global ou grupo de processos).

Além do NC o MB conterá um pequeno monitor de depuração, para facilitar o desenvolvimento e teste de programas no conjunto MB/ME. A interação com o usuário é realizada através de um periférico do tipo terminal de vídeo. Além do monitor de depuração as ROM's deverão conter rotinas de teste do MB.

O dispositivo de temporização possui três relógios temporizadores programáveis. Dois destes relógios são utilizados pelo NC para fornecer um serviço de contagem de tempo aos programas que compartilham o módulo processador. O terceiro contador é utilizado para a realização de um teste de consistência no processador (watch dog timer).

O bloco assinalado DMA além de realizar a comunicação PC-MB permite a introdução de mais controladores de DMA, através dos módulos de expansão, quando necessário.

O bloco "interrupção" é um controlador de interrupções geradas pelo PC, pelo dispositivo de temporização, pelos dispositivos periféricos que compõem o módulo de expansão e pela interface com o usuário (na figura 4, o terminal de vídeo).

Finalmente, o bloco nomeado paridade é um circuito de geração e teste de paridade para a realização do teste de consistência dos dados na via local.

A referência (21) apresenta em detalhes a especificação do Núcleo de Comunicação bem como o "hardware" do módulo básico desenvolvido.

4. A REDPUC COMO UMA CENTRAL DE COMUTAÇÃO DE PACOTES

A especialização da REDPUC para a síntese de uma Central de Comutação de Pacotes vai envolver tanto a programação das tarefas de comutação de pacotes como também a construção de "hardware" adicional (módulos de expansão) de interface. Tal central vem sendo desenvolvida na Pontifícia Universidade Católica do Rio de Janeiro com o nome de CCPUC (23 e 24).

A arquitetura lógica da Central de Comutação é apresentada na figura 5.

Referenciando-se à figura, o processo "N2 X.25" é o responsável pela função de transmissão da Central. Ele é o responsável pelo tratamento dos níveis 1 e 2 do protocolo de acesso à rede geograficamente distribuída, e do protocolo interno à mesma. O processo COM cuida de todas as conexões e desconexões de circuitos virtuais que têm origem ou terminal na central. O processo CV é o responsável pelo controle da transferência de dados em um Circuito Virtual. O processo ENV têm a função de armazenamento temporário dos envelopes em trânsito entre duas centrais. Cada processo ENV é associado a um tronco de saída. A função de coleta de dados é realizada pelos processos ME (medidas e estatísticas) e CF (controle do disco flexível). A interação homem-máquina é realizada pelos processos HM e CT. O processo RECON é responsável pela inicialização da Central, sua reconfiguração automática em caso de falha ou a pedido do operador. O processo

SUP cuida da supervisão da operação da central, da tarifação e da execução dos comandos do operador do centro de controle da rede. Finalmente o processo ICCR é quem realiza o protocolo de comunicação com o centro de controle da rede.

A figura 5 apresenta também, em linhas pontilhadas, o mapeamento dos processos aos módulos agrupando processos N2 X.25 em módulo P2, processos CV's e ENV's em módulos P3, e os processos restantes no módulo de supervisão PS. A figura 6 apresenta a arquitetura da central.

Cada módulo processador será composto do PC e do MB apresentados nos itens 2 e 3, além de módulos de expansão ME. A figura 7 dá um exemplo de um módulo de expansão do P2 para o controle das linhas de acesso e internas à rede.

5. CONCLUSÃO

O artigo procurou apresentar o projeto de uma máquina de arquitetura distribuída de alta confiabilidade, concluindo com a sua aplicação específica na síntese de uma Central de Comutação.

Modelos analíticos e de simulação foram desenvolvidos para avaliar o desempenho do protocolo de acesso ao embarramento e serão apresentados em tempo devido em outro artigo. Apenas como um exemplo a figura 8 apresenta uma curva de retardo versus desempenho.

As hipóteses utilizadas no modelo e em particular na curva apresentada foram:

- O processo de chegada de mensagens é Poisson em cada MP
- O carregamento é balanceado, i.e., a taxa média de chegada de mensagens em cada MP é igual a λ mensagens/seg.
- As mensagens tem comprimento exponencialmente distribuído com média igual a $1/\mu$ bits.
- Existe apenas um embarramento com capacidade igual a C bps.
- A mensagem vazia tem um comprimento constante igual a $1/\mu$ bits.

O retardo no eixo vertical está normalizado com relação ao tempo médio para se transferir uma mensagem de $1/\mu$ bits. O desempenho do embarramento é medido como a fração da capacidade total do embarramento usada na transmissão de mensagem não vazia.

BIBLIOGRAFIA

- (1) SAFER, S.I.; MISHELEVICH, D.J.; FOX, S.J.; SUMMEROUR, V.B. "NODAS - The Network-Oriented Data Acquisition System for the Medical Environment", Proceedings, National Computer Conference, 1977.
- (2) RAWSON, Eric G.; METCALFE, Robert M.; "Fibernet: Multimode Optical Fibers for Local Computer Networks". IEEE Transactions on Communications, July 1978.
- (3) BIRZELE, Paul; THJNSCHMIDT, Hans; "A Local Distributed Microcomputer Network Based on an Optical Bus System with Decentralized Communication Control". Proceedings of the Second International Conference on Distributed Computing Systems, April 1981.
- (4) FARBER, D.J.; "A Ring Network", Datamation, vol. 21, nº 2, February 1975.
- (5) OH, Young; LIU, Ming T.; "Interface Design for Distributed Control Loop Networks", National

Telemetry Conference, Los Angeles, 1977.

(6) FRASER, A.G.; "Spider - An Experimental Data Communications System", International Conference on Communications, 1974.

(7) MAFNER, E.R.; NENADAL, Z.; TSCHANZ, M. "A Digital Loop Communication System". IEEE Transactions on Communications, June 1974.

(8) SHIMIZU, Egmont Y.; RUGGIERO, Wilson V.; MOSCATO, Lucas A.; Componentes Básicos de uma Máquina de Arquitetura Distribuída: Identificação e Implementação. Anais do VII Seminário Integrado de Software e Hardware, Brasil, 1980.

(9) JAFARI, H.; LEWIS, T.; "A New Loop Structure for Distributed Microcomputing Systems". Proceedings of the First Rocky Mountain Symposium on Microcomputers, August 1977.

(10) METCALFE, Robert M.; BOGGS, David R. "Ethernet: Distributed Packet Switching for Local Computers Networks". Communications of ACM, nº 7, volume 19, July 1976.

(11) WITTIE, Larry D.; "Micronet: A Reconfigurable Microcomputer Networks for Distributed System Research" Simulation, volume 31, nº 5, November 1978.

(12) STRITTER, Edward P.; SAAL, Harry J.; SHUSTEK, Leonard J.; "Local Networks of Personal Computers": Proceedings Compcon Spring 81, February, 1981.

(13) TOKORO, Mario; TAMARU, Kiichiro, "Acknowledging Ethernet". Digest of Papers, Compcon Fall 77, 1977.

(14) BASS, Charlie; KENNEDY, Joseph; DAVIDSON, John M.; "Local Network Gives New Flexibility to Distributed Processing" Electronics, September, 25, 1980.

(15) BENHANOU, Eric; ESTRIN, Judy; "Design Goals for Z-net: A Commercial Local Network of Microcomputers" Proceedings, Compcon Springer 81, February 1981.

(16) CARPENTER, Robert J.; SOKOL, Joseph; ROSENTHAL, Robert; "A Microprocessor-Based Local Network Nodes". Proceedings, Compcon Fall 78, September 1978.

(17) GERHARDSTEIN, L.H.; SCHROEDER, J.O.; BOLAND, A.J.; "The Pacific Northwest Laboratory Minicomputer Network" Proceedings, Third Berkeley Workshop on Distributed Data Management and Computer Network, 1978.

(18) NICKENS, Don O.; GENDUSO, Thomas B.; SU, Y.W. Stanley; "The Architecture and Hardware Implementation of a Prototype Micronet". Proceedings of the Fifth Conference on Local Computer Networks, October 1980.

(19) JENSEN, E. Douglas; "The Honeywell Experimental Distributed Processes - An Overview". Computer, January 1978.

(20) ANDERSON, George A.; JENSEN, E. Douglas, "Computer Interconnection Structures: Taxonomy, Characteristic and Example", ACM Computing Surveys, 7, nº 4, December 1975.

(21) GOMES SOARES, Luiz Fernando; MENASCÉ, Daniel A.; "Arquitetura de uma Máquina Distribuída", Relatório Técnico P1.2 (PUC-TELEBRÁS). PUC/RJ, outubro de 1981.

(22) GOMES SOARES, L.F.; MENASCÉ, D.A.;

