

UMA ANÁLISE COMPARATIVA DAS DIVERSAS PROPOSTAS DE ATENDIMENTO DOS REQUISITOS DE COMUNICAÇÃO PARA SISTEMAS TEMPO-REAL

Keiko Verônica Ono Fonseca *
CPGEEII- CEFET-PR
Av. Sete de Setembro 3165
CURITIBA-PR

Jean-Marie Farines
LCMI-UFSC
Campus Universitário - Trindade
Florianópolis-SC

RESUMO

Este documento apresenta um levantamento dos requisitos de comunicação tempo-real, bem como várias propostas encontradas na literatura para o atendimento destes. Um quadro comparativo baseado em dois requisitos (previsibilidade do sistema de comunicação e garantia de entrega de mensagem) é montado para as propostas estudadas sobre o qual são tecidos comentários sobre as tendências levantadas para o estudo do problema de comunicação tempo-real.

1. Introdução

No contexto de Sistemas Tempo-Real exige-se o atendimento aos requisitos [Aud-90] de:

- Correção Temporal ("Timeliness") entendida como a reação do sistema a estímulos do ambiente dentro de restrições de tempo especificadas;
- Relação de Ordem ("Orderliness") entre as entradas e saídas do sistema;
- Fornecimento do dado o mais válido possível ("Freshness").

Garantir a correção temporal dos Sistemas Tempo-Real é fundamental, em particular no caso de Sistema Crítico onde uma única falha no atendimento às restrições de tempo pode resultar em consequências catastróficas para este. Em consequência disto, a Previsibilidade ("Predictability") passa ser uma das qualidades requeridas para os Sistemas Tempo-Real.

O atendimento destes requisitos num ambiente distribuído exige diversos cuidados na especificação do suporte de comunicação de Sistemas Tempo-Real. Um estudo dos requisitos de comunicação destes sistemas e as abordagens encontradas na literatura para o seu atendimento são apresentados neste documento. O documento se divide da seguinte maneira: inicialmente são apresentados os requisitos de comunicação de sistemas tempo-real e o conceito de comunicação tempo-real. São caracterizadas as mensagens em sistemas tempo-real, bem como os serviços de comunicação tempo-real e seus requisitos de desempenho. O atendimento às restrições de tempo-real é tratado conforme duas abordagens: do ponto de vista da mensagem e do ponto de vista do serviço de comunicação. Por fim, são apresentadas e comentadas diversas arquiteturas e mecanismos propostos na literatura para o atendimento dos requisitos de Sistema de Comunicação Tempo-Real levantados inicialmente.

2. Requisitos de comunicação em Sistemas Tempo-Real

2.1. Definições básicas:

A *Comunicação Tempo-Real* é entendida como "qualquer tipo de atividade de comunicação na qual as mensagens envolvidas possuam restrições de tempo associadas" [Sta-91] ou como aquela "que apresenta serviços de comunicação com desempenho garantido" [Fer-90b].

Neste contexto, as restrições de tempo nas mensagens deverão ser sempre garantidas mesmo em presença de distúrbios tais como sobrecarga ou faltas; esta garantia é entendida como a existência de "uma latência limitada e conhecida de entrega de mensagem" [Ver-91].

* bolsista CAPES-PICD em doutoramento no LCMI/UFSC

2.2. As mensagens em Sistemas de Comunicação Tempo-Real

2.2.1. As restrições de tempo das mensagens

De forma análoga às restrições de tempo em Sistemas Tempo-Real (soft/hard real-time), as mensagens com restrições de tempo podem ser classificadas em [Sta-91]:

Mensagens que exigem garantia ("guarantee seeking"):

São mensagens críticas (essenciais) para a operação correta do sistema tempo-real. O Sistema de Comunicação deve incluir uma garantia de que a atividade referida pela mensagem teve execução aceita e suas restrições de tempo foram obedecidas com certeza.

Para mensagens com características conhecidas a priori (por exemplo, mensagens periódicas) esta garantia pode ser fornecida estaticamente. Para as demais mensagens (de ocorrência não deterministas e que no melhor caso podem ser caracterizadas pela sua probabilidade de ocorrência). A avaliação da garantia é feita no instante de chegada do pedido de seu envio. O resultado desta avaliação torna possível iniciar uma ação corretiva apropriada (por exemplo, uma conexão alternativa).

Mensagens com restrições de tempo mas que não requerem uma garantia quanto ao seu cumprimento ("best effort"):

Neste caso, as restrições de tempo associadas não implicam em alto custo em relação aos benefícios da operação normal ("soft real-time constraints"). O Sistema de Comunicação deve então tentar satisfazer suas restrições minimizando o número de mensagens deste tipo com restrições de tempo violadas.

No que diz respeito à Difusão de Mensagens, a entrega das mensagens para a um grupo ou a todos nós, deve ser garantida aos seus destinatários dentro do prazo máximo determinado pelo usuário do serviço de comunicação.

2.2.2. Outras características das mensagens

Os dados dos processos do usuário:

Em processos dependentes do tempo adota-se usualmente o conceito de **Unidade Atômica de Dados**, onde valor e data de criação são inseparáveis. A informação global do tempo de criação do dado na sua fonte [Rodd-91] é mantida através de um "carimbo" ("Time-Stamping").

A integridade das mensagens:

A **Integridade** física de uma mensagem (em particular aquelas que exigem garantia de entrega) deve ser sempre mantida mesmo em situação de falta no sistema de comunicação.

2.3. especificação dos serviços de comunicação Tempo-Real:

Para a análise dos serviços de comunicação adota-se uma arquitetura multicamada para a rede de comunicação e o modelo cliente-servidor para o fornecimento de serviços de comunicação entre as camadas.

Neste contexto, os serviços habituais de comunicação são dos tipos COS- orientado a conexão (circuito virtual) e CLS- sem conexão (datagrama). Entretanto não existem nestes modelos mecanismos para especificação e garantia dos requisitos de serviços tais como atraso máximo de mensagens, vazão mínima da rede, máxima taxa de erro nas mensagens [Sta-91].

Por outro lado, em Sistemas de Comunicação Tempo-Real, não há como especificar nem garantir restrições de tempo para mensagens individuais.

Uma forma de superar estas limitações é proporcionar aos clientes dos serviços de comunicação tempo-real a possibilidade de especificar os seus requisitos de desempenho e de obter garantias de atendimento dos mesmos na forma de um "Contrato de Garantia" [Fer-90].

As cláusulas deste contrato definem a validade da garantia do cliente e penalidades pelo não cumprimento deste. Elas refletem os desejos do Cliente desde o nível mais alto (homem ou aplicação) até o nível mais baixo (o servidor, isto é, o sistema distribuído inteiro). A implementação do contrato é efetivada pelo cumprimento de condições especificadas por limitantes (deterministas ou estatísticos) dos requisitos de atraso, vazão, confiabilidade e outros [Fer-90]:

os requisitos de atraso

O atraso refere-se ao prazo de entrega da mensagem entre entidades de mesmo nível. Os requisitos de atraso devem ser especificados através de limites estabelecidos em função de um relógio de referência¹. O impacto de relógios não sincronizados sobre os limitantes estabelecidos nestes requisitos é alvo de vários estudos [Ple-90][Rodd-91] e embora de grande importância não serão discutidos neste trabalho.

[Fer-90] propõe duas formas de especificação dos requisitos de atraso: o limitante de atraso (valor máximo para atraso ou a probabilidade de entrega no prazo com sucesso) e o limitante de "delay-jitter" (limite máximo para a variação de atraso ou o limitante inferior de probabilidade de que a variação de atraso esteja no limite definido).

os requisitos de vazão:

A taxa de transferência sem erro (ou com um percentual de erros máximo admissível) determina um valor superior de vazão aceitável denominado aqui de limite de saturação². Duas formas de especificação de requisitos de vazão são o limitante de vazão mínima a ser oferecido pelo servidor (o servidor deve suportar uma vazão maior que este limitante) e o de limitante do valor médio de vazão calculada sobre um dado intervalo de tempo, o que possibilita ao cliente um controle da sua demanda de envio de mensagens.

os requisitos de confiabilidade:

É normalmente especificado pelo cliente através de um limitante inferior de probabilidade de entrega correta de mensagem.

outros requisitos desejáveis:

Outros requisitos desejáveis de serem especificadas em função das necessidades do cliente são:

- a sequenciação e ausência de duplicação que deve definir um limitante máximo para entrega fora de sequência compatível com a capacidade de processamento e de armazenamento temporário de mensagens;
- a recuperação de falhas que especifica o procedimento a ser seguido quando da ocorrência da falta e estabelece um limitante superior de ocorrência de faltas;
- o tempo de "setup" que deve ter um limite especificado em certas aplicações com configurações exclusivas.

Uma outra cláusula do Contrato de Garantia é a de rejeição de serviço não-garantido quando não há como o servidor garantir o desempenho solicitado. O motivo da rejeição pode ou não ser informado ao cliente pelo servidor.

É dever do Cliente fornecer ainda as características do tráfego de entrada esperado sem as quais não há possibilidade de estabelecer a garantia de desempenho de um serviço de comunicação [Ver-91][Fer-90][Fer-90a].

¹ normalmente, os clientes referem-se aos seus próprios relógios

² A vazão real de um sistema é limitada superiormente pela taxa com que a fonte consegue enviar mensagens pelo sistema. A taxa de transferência sem erro (ou com um percentual de erros máximo admissível) ou pelo ~~052~~ entrega de mensagens é função, entre outros, da carga da rede.

3. O atendimento dos requisitos de comunicação tempo-real

O atendimento dos requisitos apresentados é aqui analisado sob dois pontos de vista: o da mensagem (item 3.1) e do serviço de comunicação tempo-real (item 3.2). A primeira abordagem busca traduzir o requisito de latência de entrega da mensagem em procedimentos do protocolo para o seu atendimento. Na segunda abordagem, estes procedimentos são definidos para atender as especificações de desempenho solicitado para o serviço.

São estabelecidas como hipóteses iniciais que o "deadline" da mensagem é fornecido pelo usuário no pedido de serviço de comunicação e que a difusão de mensagens (periódicas ou aperiódicas) é necessária³. Neste e nos próximos itens, o termo mensagem designará mensagem com restrição de tempo associada, a não ser que especificado em contrário.

Procura-se verificar o atendimento de dois requisitos: a capacidade do sistema em atender as restrições de tempo da mensagem e caso contrário propor sua rejeição (previsibilidade) e a entrega da mensagem em tempo hábil (cumprimento da garantia).

3.1. A abordagem do ponto de vista da mensagem

Nesta abordagem, o atendimento do requisito de uma latência limitada e conhecida de entrega da mensagem é suficiente para garantir um comportamento tempo-real para a operação da rede. A maioria das propostas estudadas se preocupa (em termos de tempo máximo de transferência de mensagens) com protocolos deterministas de acesso ao meio, considerando os tempos de execução dos protocolos das demais camadas desprezíveis em relação ao tempo gasto no acesso ao meio.

Os protocolos MAC deterministas devem garantir um tempo limite de acesso ao meio para as mensagens oriundas do LCC. Os mecanismos no MAC podem ser classificados em relação ao tratamento dado à mensagem em função da sua classe de prioridade estática ou dinâmica que é normalmente associada ao "deadline" da mensagem.

3.1.1. Políticas Baseadas em Prioridade Estática:

Nestas políticas a mensagem é recebida com uma classe de prioridade fixa para transmissão. Duas opções são descritas aqui:

Sem preempção de mensagens

Uma vez definida uma ordenação de acesso ao meio e iniciada a transmissão da mensagem não é mais possível truncá-la.

Como exemplos de ordenação da fila de acesso ao meio através do escalonamento de mensagens por classe de prioridade tem-se a geração de uma fila FIFO [Stro-89]; a tabela de rotação da ficha do IEEE 802.4; os mecanismos de alocação e reserva de ficha do IEEE 802.5.

Com preempção de mensagens:

Neste caso, a mensagem pode ser truncada, sendo retransmitida posteriormente. Nesta truncagem é gerado o CRC próprio para a parte truncada e o restante da mensagem é absorvido e fica armazenado para posterior transmissão. Existe um mecanismo que altera a prioridade para a mensagem truncada de modo que esta tenha a sua transmissão completada desde que possível.

A seguir, são apresentadas algumas das propostas baseadas em protocolos das camadas inferiores (LLC e MAC) com alocação estática de prioridade.

³ Em aplicações de controle assume-se também que o tamanho das mensagens críticas no tempo seja pequeno em relação ao das mensagens não críticas (menor ou igual ao menor ciclo de tarefa periódica do sistema) [Sta-90].

3.1.1.1. A Política de fila de mensagens garantidas

No contexto de comunicação, uma rotina com garantia de atendimento à restrições de tempo é um procedimento que, para cada mensagem que chega, determina se é possível ou não inserir a próxima mensagem na fila de transmissão de mensagens garantidas. Se for possível, a mensagem é dita garantida, caso contrário ela é rejeitada e cabe as camadas superiores tomarem medidas cabíveis neste último caso. Estas rotinas são baseadas em protocolos com escalonamento de mensagens. [Fer-90] e [Sta-91] adotam esta política.

3.1.1.2. A Política de interrupção e preempção de mensagens

Esta política [Cher-90] baseia-se na interrupção de transmissão de mensagens de baixa prioridade que devem ceder o meio de transmissão a uma mensagem de prioridade mais alta de outro nó. A alocação estática de prioridades não garante o atendimento da restrição de tempo da mensagem interrompida. No limite, se todas as mensagens forem de mesma prioridade (estática), tem-se o desempenho de um protocolo MAC determinista.

3.1.1.3. A política de chaveamento de protocolos

[Etk-89] propõe um gerenciamento do tráfego da rede para ativação de dois tipos de protocolos de acesso ao meio - um para mensagens sem restrição temporal e outro para mensagens com restrição temporal. Um limiar de intensidade de tráfego é estabelecido e quando atingido um chaveamento automático entre dois protocolos ocorre simultaneamente em todos os nós.

3.1.2. Política baseada na Alocação Dinâmica de Prioridade:

A classe de prioridade original da mensagem é alterada dinamicamente conforme uma política estabelecida a priori (por exemplo: folga, "deadline", endereço) [Kan-91]. Uma fila de mensagens é mantida em cada nó e constantemente escalonada em função dos "deadlines" das mensagens que chegam para transmissão. A disputa do meio é realizada através de um "poll number" gerado em função do "deadline" e da prioridade da mensagem. Se a estação não consegue acessar o meio para transmitir uma mensagem cujo "deadline" está para ser alcançado, o "poll number" pode ser alterado dinamicamente até um limite especificado para cada estação. A alteração da prioridade se faz em função da folga (prazo restante para a transmissão sem que se perca o "deadline") da mensagem. Em casos de folgas idênticas, o desempate é feito via um identificador único da mensagem (em um campo denominado "unicidade"). Um tempo máximo de acesso ao meio é garantido através de um número limite de alterações de prioridades.

3.1.3. A garantia de atendimento as restrições da mensagem do ponto de vista de tolerância a faltas

A garantia de atendimento as restrições da mensagem pode ainda ser analisada através da identificação, análise e minimização dos fatores associados a um comportamento faltoso no sistema de comunicação. O comportamento faltoso de um sistema de comunicação é caracterizado em [Ver-91] como aquele que apresenta:

- i. falhas de tempo (por exemplo, atrasos devido a sobrecarga);
- ii. falhas de omissão (por exemplo, devido a erros de transmissão);
- iii. partições da rede (por exemplo, devido a falhas no meio físico).

As condições suficientes e as políticas para o atendimento do requisito de latência limitada já citada são resumidas tabela 1.

Um modelo de rede de comunicação abstrata [Ver-91] é utilizado para verificar quais os atributos necessários em uma rede real para satisfazer estas condições. Este modelo é uma interface em uma arquitetura multicamada que sob a subcamada LLC abstrai particularidades físicas dos protocolos das camadas MAC, PHY e entidades de gerenciamento. As restrições impostas pelo modelos são que as faltas de omissão são pouco frequentes e ocorrem em "bursts" causados por falhas em um só componente.

Condição	estratégias de atendimento
Forçar atraso limitado desde o pedido até a transmissão do quadro	implementadas a nível de rede local, garantem o envio da mensagem em tempo limitado (mesmo que esta não chegue ao destino): - políticas de alocação de prioridades - escalonamento de mensagens - controle de carga oferecida à rede
Garantir a entrega da mensagem transmitida	implementadas a nível usuário, garantem que a mensagem transmitida é entregue ao usuário mesmo que seja necessário retransmiti-la devido a erros de omissão: - redundância temporal - redundância espacial - mascaramento de erros de omissão
controlar períodos de suspensão temporária da comunicação entre nós ou subconjunto de nós (inacessibilidade)	implementação de (i) recuperação de um partição causada por falhas elétrica ou mecânica do meio físico obtida através de redundância deste meio ⁴ ; (ii) garantia de que os períodos de inacessibilidade ⁵ possuem um limite aceitável e baixo para os requisistos de serviços de comunicação (ex. limitação de tempo dos procedimentos de recuperação de ficha, anel lógico, falha no barramento); (iii) mascaramento dos períodos de inacessibilidade ao usuário

tabela 1 - condições suficientes para o atendimento da latência limitada

O atendimento da condição 3 requer a determinação do limite superior de tempo dos períodos de inacessibilidade de cada rede e sua posterior utilização nos cálculos de tempo de pior caso de execução de protocolos. O mascaramento dos períodos de inacessibilidade pode ser feito por "congelamento" dos temporizadores quando uma ocorrência de inacessibilidade é detetada⁶ e por transformação do período de inacessibilidade em um erro de omissão através do ajuste dos temporizadores do protocolo.

3.2. a abordagem do ponto de vista do serviço:

3.2.1. O modelo COS (serviço orientado a conexão)

A grande maioria das propostas de comunicação tempo-real baseadas em COS concentram-se no estabelecimento de estratégias para problema do tempo de estabelecimento de conexões específicas para mensagens urgentes, nos requisitos de desempenho destas conexões tempo-real, e na tradução dos requisitos de tempo do usuário em requisitos de desempenho dos serviços e protocolos das camadas inferiores.

3.2.2. O modelo CLS (serviço sem conexão)

O modelo CLS é simples e eficiente em termos de tempo (a transmissão se faz em um único ciclo), mas também menos confiável que o COS. Uma de suas vantagens é a facilidade de se realizar difusão de mensagens sobre a rede através do serviço de datagramas. O problema da confiabilidade deste serviço limita o seu uso para as mensagens do tipo "guarantee-seeking".

3.2.3. A integração de modelos de serviços COS e CLS

Protocolos para comunicação tempo-real que integram ambos modelos de serviços são propostos por [Sta-91]. Basicamente integram o serviço de datagrama tempo-real (RTDG) e o modelo de canal virtual tempo-

⁴ O modelo proposto em [Ver-90] não considera as falhas no meio físico porque assume um chaveamento entre meios físicos (meio físico redundante)

⁵ Um estado inacessível de um componente é definido se: pode ser feito conhecido aos usuários do componente; os seus limites (duração, taxa de ocorrência) fazem parte da especificação do componente; a violação destes limites implica em falha permanente do componente.

⁶ O hardware ou software da rede devem ser capazes de detetar e sinalizar esta ocorrência.

real com filas garantidas (RTCV). O desempenho dos protocolos integrados é inferior ao dos protocolos dedicados COS e CLS do mesmo autor.

Em resumo, as principais vantagens e desvantagens de ambos os modelos são apresentadas na tabela 2:

	VANTAGENS	DESvantagens
COS	<ul style="list-style-type: none"> - garantia de recebimento da mensagem; - negociação de parâmetros durante o tempo de estabelecimento de conexão; 	<ul style="list-style-type: none"> - aumento do tráfego da rede, - múltiplas associações concorrendo aos recursos da rede, - falta de suporte para difusão de mensagens. - consome tempo para o estabelecimento da conexão
CLS	<ul style="list-style-type: none"> - menor sobrecarga da rede que o COS por não estabelecer conexões, - o tráfego da rede depende somente do quantidade de mensagens, - há suporte para difusão global e a grupos, 	<ul style="list-style-type: none"> - a falta de garantia de recebimento, - a falta de controle de fluxo.

Tabela 2 - Vantagens e Desvantagens dos modelos de serviço COS e CLS

A escolha por um tipo ou outro de serviço pode ser influenciada pelo modelo de sistema adotado no controle (distribuído ou não) conforme discutido no item 4.

3.2.4. As estratégias para fornecimento de serviços tempo-real

Os modelos propostos para fornecimento de serviços tempo-real são baseados em serviços COS sobre um canal lógico estabelecido entre as entidades de comunicação. [Ple-90] apresenta na sua proposta de arquitetura procedimentos de estabelecimento de conexão tempo-real a nível das camadas Transporte e Rede, [Sta-90][Arv-90][Arv-91] faz o mesmo para as camadas MAC e LLC e [Fer-90] apresenta um modelo genérico de canal tempo-real.

O principal problema abordado nos protocolos que implementam COS é o tempo gasto no estabelecimento de conexões e a garantia que a conexão estabelecida mantém-se conforme a especificação negociada. As soluções propostas partem do princípio que a configuração dos serviços é estática, isto é, as associações são estabelecidas antes do controle de processos começar, bem como fornecidos todos os parâmetros da conexão utilizada para comunicação tempo-real. Em [Rodd-91] a questão da configuração dinâmica ou reconfiguração não são consideradas facilidades obtidas através de comunicação tempo-real.

3.2.4.1. O modelo proposto em [Sta-91]:

Os RTCOS (COS tempo-real) e RTCLS (CLS tempo-real) são propostos em [Sta-91]. No RTCOS estabelece-se uma conexão tempo-real somente após o sistema de comunicação verificar se podem ser atendidos os requisitos de desempenho especificados pelo cliente. Estes serviços são acessíveis via pontos de acesso do LLC. As políticas sugeridas para implementação destes serviços ao nível LLC estão resumidas na tabela 3.

técnicas de garantia da conexão tempo-real	estratégia adotada
alocação de prioridade ⇒	escalonamento de mensagens estaticamente definidas
RTCV (circuito virtual tempo-real) ⇒	escalonamento de mensagens (hipótese de pior caso)
reserva de recursos ⇒	baseada em conhecimento global de todos os pedidos de conexão do sistema

Tabela 3- Garantia do RTCOS no nível LLC

O controle de fluxo e de erros pode ser do tipo redundância temporal ou espacial (normalmente aplicadas em RTCLS). Os RTCOS ao nível LLC devem ser suportados por RTCOS a nível MAC. A tabela 4 resume algumas políticas sugeridas para implementação destes serviços a nível MAC.

As RTCLS/MAC são apropriadas para mensagens tipo "best-effort" e propostos para a sua implementação protocolos baseados em FIFO quando todas as mensagens possuem a mesma folga ("laxity") e protocolos de janela com escalonamento por MLF- "minimum laxity first" ou protocolo RTDG- Real Time Datagram [Arv-90][Arv-91][Sta-91].

conexão tempo-real MAC	estratégia adotada
Resolução de prioridades para escolha da mensagem de maior prioridade para acesso ao meio ⇒	- atrasos de deferência - comprimento de preâmbulos - "forcing headers" - protocolo de janela
RTCV- garantia de acesso em tempo limitado ⇒	- protocolos MAC deterministas
protocolos de reserva ⇒	- protocolos de difusão confiável

Tabela 4 - Garantia do RTCOS ao nível MAC

3.2.4.2. O modelo proposto por [Fer-90]

Um *canal tempo-real* é um canal lógico baseado em conexões simplex de rota fixa escolhida no momento do seu estabelecimento e com recursos necessários reservados para garantir um atraso limitado e conhecido para entrega de pacotes [Fer-90].

Este canal é estabelecido para garantia de um atraso limitado de mensagens em uma rede WAN de ambientes de chaveamento de pacotes orientada a conexão. A rede suposta pode possuir topologia arbitrária consistindo de várias subredes de vários tipos e implementar chaveamento de pacotes mas com a restrição que o enlace ("store-and-forward") estabelecido entre a fonte e o destino do pacote possua um limitante finito e conhecido para cada pacote. Os atrasos de entrega de pacotes são muito difíceis de controlar uma vez que é permitido aos pacotes seguirem por diferentes rotas entre a sua origem e o seu destino. O método é resumido na tabela 5.

Problema	Solução proposta
redução do tempo de estabelecimento da conexão	- procedimento de estabelecimento em uma única troca de mensagens - aumento da velocidade dos seus algoritmos de estabelecimento de conexão
resolução de prioridades	- escalonamento do tipo EED- Earliest Due Date - O escalonamento prioriza a fila de mensagens do canal determinista ⁷ , em relação às demais filas (dos canais estatísticos e outros)
controle de fluxo	- baseado em taxa ao invés de janelas. - controle de taxa de envio de pacotes do emissor baseado no conhecimento das características do receptor e do caminho estabelecido no canal. - Testes no estabelecimento do canal verificam se o receptor será capaz de aceitar pacotes oferecidos pelo emissor na taxa declarada

Tabela 5 - o modelo de canal tempo-real

7 Em função do tipo de limitante (determinista ou estatístico) os canais estabelecidos podem ser deterministas ou estatísticos.

Os problemas decorrentes de nós cujas taxas estão acima das verificadas nos testes e que podem causar atrasos indesejáveis nos pacotes que trafegam pelo canal tempo-real, são resolvidos através de um controle de taxa distribuído [Fer-90]. Este controle, nos nós de maior carga, aumenta o "deadline" dos pacotes gerados acima da taxa declarada de modo que não interfiram seriamente na operação dos demais canais⁸.

O canal é estabelecido através de testes e tentativas de alocação de recursos em cada nó atravessado pelo pedido de requisição. O último nó (o destinatário) envia a mensagem de estabelecimento que durante o caminho de volta confirma os recursos reservados. Se em todos os nós os testes⁹ forem favoráveis a ligação é estabelecida. Caso algum teste falhe, o resultado é retornado ao enviador ou a algum nó intermediário que possa enviar a mensagem por algum outro caminho.

3.2.4.3. Outras propostas

[Ple-90] distingue para a camada de transporte o estabelecimento de uma conexão tempo-real através de uma declaração de conexão ao invés de negociação para distinguir claramente que o nó servidor não negocia a conexão e sim declara uma conexão e seus parâmetros ao nó cliente. O nó cliente não pode recusar a declaração; se os mecanismos declarados são incompatíveis com o seu funcionamento, ele sinaliza e se retira do sistema, ou seja, a sua configuração está errada.

No caso particular de sistemas de comunicação voltados para a automação da manufatura, modificações em termos de novos objetos e serviços são propostos para o protocolo de Aplicação MMS (RTMMS [Rodd-91] e Perfil de notificação periódica [Ple-90]). Ambos autores complementam o MMS com serviços que possibilitam o envio de mensagens regularmente em intervalos periódicos (dados amostrados a uma taxa constante). As extensões do MMS definem novos objetos (por exemplo, variáveis de estado com carimbo de tempo, lista de variáveis de estado, objeto de configuração, condições de eventos). É proposta uma extensão no objeto VMD (RTVMD [Rodd-91]) que fornece suporte para informação de tempo físico durante a operação do equipamento.

4. O caso particular de sistemas de controle em Sistemas de Manufatura

Outras propostas procuram minimizar os problemas que possam ser causados pela não garantia de entrega de mensagens em sistemas de controle, em especial aqueles voltados para a manufatura. As soluções propostas envolvem outros aspectos além do sistema de comunicação em trabalhos baseados em mensagens tipo "best-effort". Neste contexto, [Rodd-91] e [Etk-89] propõem modelos de sistema de controle distribuído com nós o mais independentes possível como solução para implementação de um sistema de comunicação tempo-real. O sistema de controle seria composto de um conjunto de nós autônomos onde cada nó realizaria seu "check-out" de funcionamento e acionaria os procedimentos de emergência em caso de faltas.

[Rodd-91] propõe um ambiente orientado a estados ao invés de eventos. A comunicação se faria por protocolos de difusão com transferência constante de variáveis de estado e implicaria em tráfego constante na rede através de serviços de datagramas. Esta política manteria uma visão global do sistema, sem manter mecanismos complexos de recuperação de mensagens. Uma perda de mensagem não seria crítica ao sistema e ativaria processos de diagnóstico em que o estado do sistema poderia ser inferido a partir de informações armazenadas e ativaria mecanismos de emergência. A referência global de tempo seria necessária para controle da validade das mensagens na rede (para retirada de mensagens inválidas, por exemplo). Quando necessário, o reconhecimento de mensagens seria feito pelo usuário ou serviço convencional MMS tempo-real.

Uma abordagem denominada "global" é proposta por [Etk-89] para ambientes de manufatura. Basicamente, discute-se a transferência de parte das funções de controle do sistema para os processadores de comunicação. Dois conceitos suportam esta abordagem: a eliminação de mecanismos de divisão de tempo (para reduzir o custo de chaveamento de contexto em termos de sobrecarga e complexidade, eliminando ou reduzindo a preempção de tarefas em execução) e a separação de responsabilidade de gerenciamento de tarefas entre os processadores de controle e de comunicação.

⁸ Deve ser alocado estaticamente um espaço de armazenamento temporário de pacotes para este fim.

⁹ Os testes referem-se à disponibilidade de alocação de banda passante aos enlaces, espaço em "buffers" nos nós e capacidade de processamento. Um novo canal não pode alterar as garantias de desempenho definidas para os canais já estabelecidos.

5. Comentários finais:

Os requisitos importantes para a implementação da comunicação tempo-real são os que implementam a **garantia de entrega da mensagem** ou do serviço que a implementa. Sob este enfoque foram analisadas diversas propostas de comunicação tempo-real encontradas na literatura. Uma comparação destas propostas procura classificar e relacionar as diversas técnicas utilizadas e apresentada na tabela 7. Cabe aqui observar que não são aqui discutidos os diversos resultados referentes à ajustes nos parâmetros existentes em protocolos já normalizados [Rze-89][Mar-88][Stray-88][Yan-90]. O interesse deste trabalho voltou-se basicamente em novas propostas.

5.1. as abordagens de estudo do problema:

As propostas de comunicação tempo-real baseiam-se no fornecimento de uma *garantia* de atendimento seja das restrições de tempo da mensagem (ponto de vista da mensagem- item 2.1) ou dos requisitos de serviço (ponto de vista do serviço- item 2.2). No primeiro caso, a restrição de tempo (C1) é associada ao dado (mensagem), enquanto que no segundo, uma ação definida (serviço) é associada a restrição de tempo (C2).

É possível provar que a restrição de tempo em C2 é satisfeita se a restrição C1 associada ao dado transmitido também o é [Tho-90], e portanto, é razoável associar ambos os tipos de restrições. Sob este enfoque, a implementação da comunicação tempo-real é em grande parte independente da abordagem adotada. Distingue-se duas técnicas para a transferência de mensagens com restrições de tempo: as modificações nos procedimentos dos protocolos de comunicação e as alterações em ambos, no modelo de controle e no modelo do sistema de comunicação.

5.2. a questão da Previsibilidade:

A previsibilidade é função direta da monitoração dos prazos associados às mensagens e traz informações valiosas para o sistema em termos de tratamento de faltas. Sob este enfoque a atomicidade da mensagem (o dado em si e a data de sua criação) é muito importante para o sistema de comunicação. As diversas propostas de arquitetura estudadas baseiam suas predições de atendimento às restrições das mensagens em termos do tempo máximo de acesso ao meio e procuram implementar funcionalidades que permitam o cálculo de folgas e alterações de prioridades. [Kan-91], [Sta-91] e [Fer-90][Fer-90a] discutem os diversos aspectos do problema de rejeição de pedidos de entrega de mensagem por questões de garantia.

Nas propostas estudadas, a previsibilidade das atividades de comunicação tempo-real praticamente só foi considerada em termos do determinismo do protocolo de acesso ao meio. No entanto, o tempo limite de transferência de mensagens deve ser considerado entre os processos de aplicação! A questão do determinismo das camadas superiores foi citada em [Rze-89] que constatou em pequenas redes com condições de carga leve um tempo de acesso ao meio em torno de 20% do tempo utilizado no processamento do protocolo de Transporte no MAP. Esforços para redução desta proporção são pouco relatados [Rodd-91][Fer-90][Fer-91] e [Ple-90].

5.3. o problema do tempo de transferência de mensagens:

Há consenso que o processo de estabelecimento de conexão é prejudicial para o envio de mensagens com restrições críticas de tempo. Conexões a nível de enlace de dados e MAC são descritas ou recomendadas nas propostas estudadas, no entanto, não está clara a influência do uso ou não de conexões nas camadas superiores. A questão do impacto do modelo de serviço adotado não conclui onde (nem se) devem ser mantidas as conexões em todas as camadas. [Fer-90a] argumenta ser muito difícil ou impossível construir canais tempo-real sob serviços sem conexão. Várias propostas vão neste sentido: utilização SAPs específicos para uso dos RTVCs [Sta-91]; SAPs para protocolo LLC tipo 3 sob conexão de transporte [Ple-90] e outras [Arv-91][Arv-91a][Ram-91].

O uso de serviços sem conexão (datagramas) são sugeridos por [Rodd-91] em um ambiente orientado a estados e por [Sta-90][Sta-91] com o protocolo RTDG, mas sempre pressupondo mensagens não críticas no tempo. A tabela 6 apresenta uma visão de algumas das propostas estudadas em termos de tipos de serviço.

Em relação aos mecanismos das camadas inferiores baseados em MAC determinísticos, a política de fila garantida tem o inconveniente de fornecer garantia só às mensagens que podem ser incluídas na fila. Uma mensagem importante e com um "deadline" próximo que chega ao sistema é perdida se não for possível incluí-la na fila de mensagens garantidas, mesmo que as mensagens já na fila sejam de prioridade menor. Uma vez estabelecida a fila de acesso, não há como refazê-la. No entanto, se existe uma previsão de não atendimento (rejeição de pedido) com sinalização às camadas superiores mecanismos de emergência para tratamento deste problema podem ser ativados.

A política de interrupção torna o acesso ao meio das mensagens urgentes independente do tráfego na rede de mensagens de baixa prioridade. A grande desvantagem é que esta política não garante o determinismo do protocolo para as mensagens interrompidas. Deve ser implantado algum mecanismo para "fome de ficha". Não há a atomicidade da mensagem no sentido de que esta pode ser truncada. Da mesma forma, se a mensagem truncada possuir "carimbo" de tempo de criação, mecanismos devem garantir este dado nos seus fragmentos.

A alocação dinâmica de prioridades associada a um protocolo MAC possui sempre um limite de tempo com a disputa pelo meio resolvida em caso de empate pela unicidade do "poll number". Mecanismos de rotação dos números únicos ou outro critério devem ser adotados para que o sistema não priorize sempre as mesmas estações. Apesar de interessante, a alocação dinâmica de prioridades não é considerada em outros protocolos MAC e/ou LLC. Foram poucas as informações encontradas sobre dados referentes a estimação do tempo de reescalonamento de mensagens e/ou de modificação de prioridades e sua influência sobre o desempenho da rede.

A proposta de [Ver-91] baseada no modelo de rede abstrata do mesmo autor cita as políticas a serem utilizadas no atendimento às condições levantadas na tabela 1 sem especificar qual a mais adequada.

Modelo de Serviço	propostas
COS	-[Ple-90] simplificação de procedimentos de estabelecimento de conexão para mensagens críticas no tempo -[Fer-90] canal tempo-real -[Sta-91][Sta-90][Arv-91] circuito virtual tempo-real
CLS	-[Rodd-90]: difusão de mensagens (orientação a estados) -[Sta-91]: RTDG
COS/CLS	[Stan-91]: protocolos integrados (diferenciação dos tipos de mensagens)

Tabela 6 - as propostas em termos de modelo de serviço

5.4. a questão dos modelos de sistema

O modelo baseado numa visão global do sistema, através de transferência regular de variáveis de estados entre os nós [Rodd-90], é ainda dependente de protocolos de difusão confiável e de sistemas de gerenciamento e manutenção de múltiplos bancos de dados. Não achamos válido dizer que a perda de uma mensagem não seria crítica ao sistema, pois sua importância é dependente da aplicação do usuário e da sincronização de tarefas e dos dados. No entanto, a discussão sobre o gerenciamento de processos distribuídos e a necessidade de tomar decisões de controle baseadas em conhecimento global do sistema deve ser estudada.

A abordagem global de [Etk-89] discute o papel do sistema de comunicação no ambiente de manufatura em relação às tarefas de controle dos processos de aplicação e coloca em questão a tradicional separação de atividades de comunicação e controle. Este questionamento leva em conta que futuramente a maioria dos sistemas a serem controlados serão sistemas dinâmicos, onde as mensagens serão dinâmicas e com diversos tipos de restrições de tempo (padrões ricos de comunicação) [Sta-91]. Prevê-se restrições individuais de tempo independentes de outras mensagens tendo como causa os requisitos de cooperação de software de controle e de sistemas operacionais distribuídos tempo-real (por exemplo: escalonamento distribuído). Esta previsão deve ser levada em conta de alguma forma nas propostas de comunicação tempo-real. Modelos de controle de execução de

tarefas críticas no tempo em ambiente de manufatura devem ser estudados, bem como em modelos que permitam uma estimação da relação custo-benefício da alocação de maior "inteligência" ao sistema de comunicação.

proposta	Garantia de Entrega em tempo limitado	Previsibilidade	Atomicidade da Mensagem	Mecanismos	Observações
[Sta-90] [Sta-91]	sim	sim	sim	fila garantida RTVC RTDG	-chaveamento entre protocolos -protocolos integrados
[Kan-90]	sim "polled bus"	não	sim	alocação dinâmica de prioridades -"poll number"	
[Cher-90]	sim (*) token-ring	não	não	Interrupção e pre-empção de mensagens	somente para mensagem com prioridade máxima
[Etk-89]	sim token-bus	não	sim	chaveamento entre protocolos	sistema distribuído "inteligente"
[Rodd-91]	não difusão	não	sim	datagramas (difusão)	-RTMMS -controle baseado em estados
[Fer -90] [Fer-90a]	sim	sim	sim	-escalonamento de mensagens - controle de taxa distribuído	canal tempo-real
[Ver-91]	sim	sim	?	- redundância temporal ou espacial - controle dos períodos de inacessibilidade	- modelo de rede abstrata
[Plein-90]	sim token-ring	sim	sim	-simplificação da conexão -escalonamento de mensagens no MAC	

Tabela 7 - Quadro comparativo das propostas estudadas

5.5. conclusões:

As seguintes questões conduziram a investigação apresentada neste trabalho:

Quais são os requisitos de comunicação de um Sistema Tempo-Real?

Quais são os fatores que influenciam durante a execução dos protocolos no atraso das mensagens?

Onde devem ser implementados os mecanismos de garantia de cumprimento das restrições temporais das mensagens?

Quais são estes mecanismos?

Como prever se uma garantia pode ser fornecida e se esta não puder ser fornecida qual procedimento tomar?

Não há ainda consenso sobre onde deve se concentrar o tratamento das mensagens- nas camadas superiores ou inferiores de uma arquitetura multicamada. Basicamente, são atacados quatro pontos nas propostas estudadas:

- o tempo máximo de acesso ao meio (através de propostas de protocolos MAC),
- o aumento da "inteligência" dos protocolos (cálculo da folga, modificação de prioridades, escalonamento, previsibilidade nos protocolos MAC/LLC),
- a diminuição do tempo de execução dos protocolos superiores (Transporte, Rede e Aplicação),
- a inclusão de mecanismos que possibilitem a difusão e controle distribuído de taxa de mensagens.

Há consenso sobre o uso de COS sobre canais virtuais dedicados à comunicação tempo-real para mensagens críticas no tempo e para as mensagens não críticas são sugeridos CLS. No caso de COS, trabalha-se na otimização dos procedimentos de estabelecimento da conexão, na identificação de parâmetros a serem negociados e no mecanismos de resolução de prioridades e alocação de recursos. Nota-se uma preocupação crescente em se adequar a rede para ambos tipos de serviços de modo a atender o tráfego heterogêneo previsto para aplicações em multimídia [Zhe-92][Fer-90a].

A especificação de requisitos de desempenho para os serviços de comunicação e o problema de tradução destes para as diversas camadas é um problema ainda em aberto. Não há ainda um perfil definido em termos de uma arquitetura multicamada sobre onde devem ser estabelecidas ou não as conexões. Sobre a hipótese de que as características de carga ofertada é conhecida e pode ser controlada supõe mecanismos de gerenciamento de rede não descritos nas propostas estudadas.

Inúmeros são os esforços sobre protocolos MAC deterministas (ajustes e determinação de parâmetros em função da carga). Fica clara também a importância de mecanismos associados para o atendimento das restrições de tempo e da previsibilidade tais como resolução de prioridades e escalonamento de mensagens. No entanto, ainda não está claro quais as propostas de MAC deterministas e de mecanismos associados mais adequados para a comunicação tempo-real (em especial, o impacto da inclusão destes mecanismos sobre o tempo de acesso ao meio).

As tendências identificadas neste trabalho para o problema da comunicação tempo-real estabelecem duas frentes:

- na sistematização da especificação e avaliação do atendimento de requisitos de comunicação e de modelos com conexões dedicadas a comunicação tempo-real, e
- nas propostas de arquiteturas e protocolos que atendam as propriedades e requisitos estabelecidos a priori. O segundo tópico envolve ambas soluções, analíticas e práticas (simulações e medições em protótipos).

Sugere-se uma análise crítica da primeira tendência através de modelos que permitam avaliar melhor as propostas de serviços de comunicação tempo-real, bem como de um modelo único para avaliação e comparação de implementações das diversas propostas existentes.

6. Bibliografia

- [Arv-91] K.Arvind, J.Stankovic, K.Ramamritham: "Window MAC Protocols for Real-time Communication Services"; COINS Technical Report 90-127; University of Massachusetts at Amherst; January 1991.
- [Arv-91a] K.Arvind, J.Stankovic, K.Ramamritham: " Reservation Approach to Real-time Communication Services"; COINS Technical Report 90-128; University of Massachusetts at Amherst; 1991
- [Aud-90] Audsley, Bhattacharyya, Burns, Fohler, Kantz, Kopetz, McDerrmid, Schutz, Zainlinger; "Timeliness - Sumary and Conclusions" - Specification and Design for Dependability -ESPRIT BRA Proj 3092 May 92 vol2.
- [Cher-90] V.Cherkassky, H.Lari-Najafi, N.L.Lawrie, D.Masson, D.W.Pritty: "Performance of a new LAN for Real-time traffic"; Computer Communications; vol.13; n5; June 1990; pp.259-266
- [Etk-89] J.Etkin; J.Zinky: "Development Life Cycle of Computers Networks: The Executable Model Approach"; IEEE Transactions on Software Eng.; vol.15; n9; pp.1078-1089; September 1989
- [Fer-90] D.Ferrari: "Client Requirements for Real-Time Communicatio Services"; IEEE Communications Magazine; November 1990
- [Fer-90a] D.Ferrari, D.C.Verma: "A Scheme for Real-time Channel Establishment in Wide-Area Networks"; IEEE Journal of Selected Areas in Communications, vol.8, n°3, April 1990.

- [Kan-90] Kang G.Shin, C-H Hou: "Analysys of Three Contention Protocols in Distributed Real-Time Systems" ; Proceedings of Real-Time Systems Symposium, IEEE Computer Society, Lake Buena Vista Florida, december 1990
- [Kan-91] Kang G.Shin: "Real Time Communications in a Computer-Controlled Workcell"- IEEE Transactions on Robotics and Automation, vol.7, n 1, February 1991
- [Mar-88] M.Marathe, R.A.Smith: "Performance of a MAP Network Adapter; IEEE Network, vol.2, n^o3, May 1988
- [Ple-90] Patrick Pleinevaux : "Architecture de Communication Pour Temps Reel Strict" - Ecole Polytechnique Federale de Lausanne- 1990.
- [Ram-92] K.Ramamritham, J.Stankovic: "Time-constrained Communication Protocols for Real-time Communication Services"; Technical Report ; University of Massachusetts at Amherst; 1992
- [Rodd-91] M.G.Rodd, I.Izikowitz, G.F.Zhao: " RTMMS - An OSI-based Real-Time Messaging Systems"- The Journal of Real-Time Systems, n3, 1991
- [Rze-89] H. Rzechak, A.E.Elnakhal, R.Jaeger: "Analysis of Real-Time Properties and Rules for Setting Protocol Parameters of MAP Networks"- The Journal of Real-Time Systems, n 3 , 1991
- [Sta-90] K.Arvind, K.Ramamritham, J.Stankovic: "Integrated Protocols for Real-Time Virtual Circuit and Real-Time Datagram Services", IEEE RTS'90.
- [Sta-91] K.Arvind, Ktithi Ramamritham, John A.Stankovic: "A Local Area Network Architecture For Communication In Distributed Real-Time Systems"- The Journal of Real-Time Systems- 1991
- [Sta-92] J.Stankovic, K.Ramamritham, E.M.Nahum: "Predictable Interprocess Cmmunication for hard real-time systems"; Technical Report ; University of Massachusetts at Amherst; 1992
- [Stray-88] W.T.Strayer, A.C.Weaver: "Performance Measurements of Data Transfer Services in MAP"; IEEE Network, vol.2, n^o3, May 1988
- [Str-89]- J.K.Strosnider, T.Marchok, "Responsive Deterministic IEEE 802.5 Token Ring Scheduling", Journal of Real-Time Systems, 1989, pp.133-158.
- [Tho-88] J-P Thomesse, J-L Delcuvellerie: "Reflexions des utilisateurs sur les réseaux locaux industriels"; Congrès AFCET Automatique -1988 ; Grenoble - France
- [Tho-90] J-P Thomesse: "Time-Critical Applications and Time Critical Communications"; ISA/SP50-1990-398, april 1990
- [Ver-91] P.Verissimo, J.Rufino, L.Rodrigues: "Enforcing Real-Time Behaviour on LAN-based Protocols"; Proceedings of the IFAC Workshop on Distributed Computer Control Systems; September 1991, Áustria
- [Yan-90] Yann-H.Lee, L-T. Shen: "Real-time Communication in Multiple Token Ring Networks; Proceedings of Real-Time Systems Symposium, IEEE Computer Society, Lake Buena Vista Florida, december 1990
- [Zhe-92] Q.Zheng, K.G.Shin: "Real-time Fault-tolerant Communication in Computer Networks"; Research Summary; University of Michigan . 1992