Abordagens de Comunicação em Sistemas Distribuídos Tempo Real

Patrícia Pitthan de Araújo Barcelos[1] Taisy Silva Weber[2]

Universidade Federal do Rio Grande do Sul
Instituto de Informática
Caixa Postal 15064, CEP 91501-970
Porto Alegre – RS, Brasil
{pitthan, taisy@inf.ufrgs.br}

Resumo

Confiabilidade em sistemas distribuídos está deixando de ser uma característica desejável para se tornar uma necessidade. Este fato é ainda mais marcante em se tratando de sistemas distribuídos tempo real, cujo cumprimento das exigências temporais impostas está diretamente relacionado às conseqüências da ocorrência de uma falha. Este artigo descreve algumas das características, exigências e garantias de sistemas distribuídos tempo real sob o ponto de vista de comunicação, enfatizando sua aplicação em um sistema de controle distribuído.

Abstract

Distributed systems reliability are becoming necessary instead of a desirable feature. This fact is more important if we're talking about real time distributed systems, whose execution of temporal requirementes imposed is closely related to the results of a fault. This article describes some real time distributed systems features, requirements and guarantees under communication point of view, emphasizing its application in a distributed control system.

1. Introdução

Quando se deseja controlar o comportamento de um sistema deve-se agir sobre o mesmo em momentos determinados do tempo. Desta forma é possível realizar uma análise dos efeitos produzidos pela ação tomada. Tal situação ilustra o funcionamento de um sistema tempo real. Apesar de redundante, a definição informal de sistema tempo real pressupõe um sistema que muda seu estado em função do tempo, o tempo real.

Atualmente os sistemas tempo real apresentam-se cada vez mais de forma distribuída. Tais sistemas são compostos por um conjunto de nodos interconectados por um sistema de comunicação tempo real. O acesso a esses sistemas de comunicação se dá por meio de protocolos tempo real, os quais caracterizam-se por apresentar um tempo de execução máximo reduzido a um período conhecido.

---

[1] Mestre e Doutoranda em Ciência da Computação (UFRGS), Professora do Instituto de Informática da PUCRS. Áreas de interesse: Sistemas Distribuídos, Redes de Computadores, Tolerância a Falhas.

[2] Doutora em Ciência da Computação (Karlsuhe, 1986), Professora do Departamento de Informática da UFRGS. Áreas de Interesse: Sistemas Distribuídos, Tolerância a Falhas, Arquitetura de Computadores.

Este artigo apresenta aspectos de comunicação em sistemas distribuídos tempo real. São explorados os conceitos básicos de sistemas tempo real, que envolvem sua definição, classificação e paradigmas de implementação. A forma com que os sistemas tempo real endereçam as características de sincronismo na presença ou não de relógios, bem como a maneira pela qual manipulam mensagens frente às exigências temporais impostas também são alvo de análise. Com base neste estudo, é proposta uma aplicação para a qual os aspectos acima mencionados podem ser associados.

O objetivo deste trabalho é apresentar os resultados de uma pesquisa em comunicação em sistemas tempo real enfatizando uma aplicação de controle para a qual é necessária a manipulação de restrições temporais.

2. Sistemas Distribuídos Tempo Real

Kopetz [KOP 93] define um sistema tempo real como um sistema que muda de estado em função do tempo (real). Esta é uma definição genérica e engloba todo e qualquer sistema tempo real, não somente aqueles controlados por sistemas computacionais.

Para Santos [SAN 91], um sistema é caracterizado como tempo real porque as aplicações que ele controla devem ocorrer em um determinado tempo, de forma independente do sistema controlador. Assim, o relógio que controla as ações do sistema tempo real não é o relógio interno do computador, mas sim o relógio que controla a dinâmica dos estados do processo no ambiente em que o sistema está inserido.

Uma das maiores dificuldades encontradas no desenvolvimento de sistemas tempo real é que eles, normalmente, são caros, pois exigem um profundo conhecimento sobre a aplicação onde atuam. Além disso, são difíceis de testar, uma vez que devem ser verificados diretamente ou por meio de simulação. Em ambos os casos, pequenas alterações no sistema acarretam em uma nova rodada de testes. Sistemas tempo real diferem de sistemas convencionais por apresentarem restrições temporais e tratarem, na maioria das aplicações, com situações críticas. Em tais aplicações, a ocorrência de qualquer tipo de falha, inclusive falha de temporização, pode causar conseqüências catastróficas. Portanto, ao contrário dos sistemas onde há uma separação entre correção e desempenho, em sistemas tempo real, correção e desempenho estão fortemente relacionados [STA 88].

Uma classificação de sistemas tempo real amplamente aceita na literatura é proposta por Kopetz e Veríssimo [KOP 93]. Segundo eles, sistemas tempo real são classificados como sistemas tempo real brandos e sistemas tempo real críticos. Tal classificação leva em consideração o atendimento aos requisitos temporais e a conseqüência das falhas apresentadas pelos sistemas.

Um sistema é caracterizado como tempo real brando se as aplicações que ele controla são capazes de tolerar pequenos atrasos, os quais não implicam em conseqüências desastrosas. Este sistema pode apresentar exigências de alto grau de disponibilidade de utilização e de integridade física no que se refere aos dados. Quando as conseqüências de uma falha conduzem a situações de risco, trata-se de um sistema tempo real crítico. Tais sistemas podem se apresentar livres de falhas ou com falhas mascaradas. Um sistema tempo real crítico é considerado livre de falhas se um ou mais estados, os quais serão acessados em caso de falha, podem ser identificados. Entretanto, se a aplicação não permite a identificação de estados seguros, o sistema deve fornecer um nível mínimo de serviço mesmo no caso de falha,

evitando, com isso, a ocorrência de uma catástrofe. Este sistema é caracterizado por apresentar falha mascarada.

Quanto à abordagem seguida durante a implementação, os sistemas tempo-real podem ser classificados ainda como sistemas tempo real com respostas garantidas (*guaranteed-response*) ou sistemas tempo real de melhor esforço (*best-effort*). Se, a partir de cenários de carga e falhas, cria-se um projeto que torna possível um debate sobre sua adequação sem usar argumentos probabilísticos, diz-se que o sistema possui resposta garantida. Os sistemas tempo real de melhor esforço são projetados especialmente para atuarem em situações onde é fundamental que o controle seja completo, irrestrito e confiável, visando garantir a disponibilidade do sistema, mesmo quando é necessária uma determinada adaptabilidade às mudanças do ambiente, como no caso de controle de tráfego aéreo.

Em sistemas tempo real há ainda que se considerar os paradigmas sob os quais os mesmos são implementados. Quanto à forma de reação às modificações externas, existem dois paradigmas distintos utilizados no projeto de sistemas tempo real: sistemas disparados por evento (*event-triggered*) e sistemas disparados por tempo (*time-triggered*) [KOP 94].

No caso de sistemas disparados por evento, todas as atividades são iniciadas como conseqüência de eventos externos, ou seja, alguma mudança significativa do estado do sistema. A reação a eventos externos deve ocorrer de forma direta e imediata, de forma a não comprometer a característica tempo real do sistema. Tais sistemas são caracteristicamente não previsíveis, uma vez que não é possível determinar antecipadamente quando serão realizadas as ações. Entretanto, alguma previsibilidade pode ser alcançada caso seja possível a limitação da taxa máxima de eventos em função da física do sistema controlado.

Em sistemas disparados por tempo todas as atividades são dirigidas pela progressão do tempo global. Desta forma, todas as tarefas ou ações de comunicação do sistema ocorrem em instantes de tempo pré-determinados. Este paradigma é menos flexível que o anterior, porém, devido a característica de previsibilidade, é mais fácil de ser analisado e testado. Sistemas disparados por tempo exigem sincronismo entre os nodos do sistema distribuído no qual estão inseridos.

3. Comunicação Síncrona x Assíncrona

Protocolos de comunicação tempo real seguem a mesma terminologia adotada pelos sistemas convencionais, onde são encontrados protocolos síncronos e protocolos assíncronos.

Protocolos síncronos são necessários para aplicações tempo real críticas, as quais devem obedecer limites de tempo de resposta mesmo quando falhas de componentes ocorrem. Tais limites são alcançados pela suposição de que atrasos de mensagens entre processadores corretos são limitados e há suficientes caminhos de comunicação redundantes entre os processadores tal que, se todos os caminhos são usados em paralelo para uma difusão, a probabilidade que todos falhem durante a difusão é insignificante [KOP 92].

A suposição de atrasos exige que processadores que implementam um protocolo síncrono sejam controlados por sistemas operacionais tempo real capazes de garantir limites nos atrasos de escalonamento de tarefas. Quando mensagens não são limitadas, isto é, falhas de comunicação podem ocasionar particionamentos, a terminação da difusão em um tempo limitado não pode ser garantida [KOP 92].

Protocolos assíncronos sacrificam terminação para oferecer tolerância a falhas de comunicação, incluindo falhas que conduzem ao particionamento. Por não garantirem limitação no tempo ao difundir uma informação na presença de falhas, tais protocolos não podem ser usados em aplicações críticas, as quais devem garantir que limites temporais sejam sempre encontrados, mesmo na ocorrência de falhas de um componente[KOP 94].

4. Serviços de Comunicação

As semânticas dos serviços de comunicação tempo real podem ser caracterizadas pelas combinações das propriedades de concordância, ordenação e sincronismo. Há essencialmente duas maneiras de classificar protocolos de comunicação de grupo, no que se refere às propriedades do domínio do tempo: protocolos disparados por relógio (*clock-driven*) e protocolos sem relógio (*clockless*).

Protocolos disparados por relógio contam com a existência de um tempo global baseado em relógios ([BAB 85], [CRI 90], [CRI 85]), enquanto os protocolos sem relógio utilizam-se de referências de tempo relativas, os *timers* ([BIR 87], [KAA 89], [VER 90]).

Protocolos disparados por relógio e protocolos sem relógio são classificados em síncronos e assíncronos, respectivamente, embora a primeira terminologia seja preferida, uma vez que existem protocolos síncronos sem relógio. Os protocolos síncronos sem relógio apresentam como característica a presença de limitação nos atrasos das mensagens e capacidade de ordenação temporal das mesmas [VER 90].

5. Mensagens em Sistemas de Comunicação Tempo Real

As aplicações em sistemas de comunicação tempo real podem ser bastante distintas quanto ao tipo de restrição temporal, escala de tempo envolvida, distribuição de controle, âmbito de atuação, etc. Estes fatores determinam diferentes requisitos de comunicação e, portanto, as formas de atendimento a estes requisitos podem exigir várias funcionalidades para o seu suporte de comunicação tais como escalonamento de mensagens, reserva de recursos, controle de tráfego da rede, nem sempre necessárias em um sistema de comunicação convencional [ARV 93].

As mensagens em sistemas de comunicação tempo real são caracterizadas por duas restrições temporais, ocorrência no sistema e exigências quanto a sua integridade física. As mensagens com restrições de tempo podem ser classificadas segundo suas exigências de garantia [ARV 93]: mensagens que exigem garantia (*guarantee seeking*) e mensagens com restrições de tempo, mas que não requerem uma garantia quanto ao seu cumprimento (*best-effort*).

A primeira classe de mensagens, mensagens que exigem garantia (*guarantee seeking*), se refere a mensagens críticas, ou seja, essenciais para a operação correta do sistema tempo real. O sistema deve incluir uma garantia de que a atividade referida pela mensagem teve execução aceita e suas restrições de tempo foram obedecidas com certeza. A segunda classe de mensagens, mensagens com restrições de tempo que não exigem garantia quanto ao seu cumprimento (*best-effort*), corresponde às aplicações nas quais as restrições de tempo associadas não implicam em alto custo com relação aos benefícios da operação normal. Neste caso, o sistema de comunicação deve tentar satisfazer as restrições minimizando o número

demensagens com restrições de tempo violadas. Um percentual de perdas ou atrasos é admitido para este tipo de mensagem.

6. Aplicação da Comunicação em Sistemas Tempo Real

Os aspectos enfocados neste artigo denotam as características usualmente necessárias em sistemas distribuídos tempo real. Entretanto, a apresentação de tais características implica na determinação de uma aplicação que envolva as mesmas. Uma das aplicações para a qual são necessárias restrições temporais corresponde a um sistema de controle distribuído de tráfego ferroviário. Tal sistema pode ser classificado como sistema tempo real de melhor esforço (*best-effort*).

A aplicação de controle de tráfego ferroviário baseia-se em um experimento desenvolvido em Newcastle por Cecília Rubira [RUB 94], cuja implementação foi modelada com orientação a objetos, entretanto não considera restrições temporais e leva em conta apenas falhas de ambiente. Como o nosso objetivo não está concentrado em modelagem, a idéia é utilizar esta sugestão de aplicação para implementar comunicação de grupo com restrições temporais.

O sistema de controle distribuído de tráfego ferroviário proposto supõe uma malha ferroviária composta pelos seguintes elementos:

- vias, as quais por definição são todas unidirecionais, formadas por trilhos, sendo que seções de trilhos podem ser compartilhadas por mais de uma via;
- cruzamentos controláveis, os quais possuem chaves que recebem o sinal de rota;
- trens inteligentes, não somente com capacidade de enviar e receber informações do sistema, mas também com capacidade de ser controlados pelo sistema ou, em caso de colapso total do sistema de controle, procurar um estado seguro;
- sistema distribuído de contole de vias;
- sensores de fluxo nas vias, os quais informam a presença ou não de trem no trilho.

Cada seção de trilhos é controlada por um grupo de nodos (processadores do sistema). Além disso, os grupos devem comunicar-se para avisar a passagens de trens pela fronteira entre as seções. Neste sentido são necessários protocolos de comunicação de grupo, as quais suportem restrições temporais, como os apresentados itens 3 e 4.

Os objetivos do sistema de controle são evitar colisões e descarrilamento de trens e fornecer informações aos trens sobre rotas alternativas. O modelo de falhas para este sistema pode ser composto por falhas de ambiente e falhas do sistema. Situações que ilustram falhas de ambiente são vias bloqueadas, chaves e sensores com defeito, e falta de comunicação com trens. Falhas do sistema podem ocorrer quando não há comunicação entre sistema e chaves ou sensores, servidores ou controladores não respondem a ativações do sistema e falha na comunicação entre os nodos.

O tráfego ferroviário pode ser considerado um sistema de controle relativamente fácil de ser representado, uma vez que as vias férreas apresentam grau de liberdade limitado e a taxa de defeitos de ambiente é baixa. É interessante observar que não pode ser assumido o modelo de falha simples (ou seja apenas uma falha de cada vez) devido a grande quantidade de elementos no sistema. Para cada componente que pode apresentar defeito, deve ser analisado qual o comportamento sob falha assumido. Além disso, deve ser analisado também o impacto de falhas de temporização no comportamento do sistema.

7. Conclusão

Uma característica quase sempre presente em sistemas tempo real é a previsibilidade, que indica a capacidade de se prever uma violação de garantia ou uma incorreção temporal. Esta característica torna possível um tratamento de exceções que minimize as conseqüências de uma falha, desde que, obviamente, esta previsão aconteça em tempo hábil para iniciá-lo.

Em um ambiente distribuído, como no caso da aplicação de controle ferroviário, a obtenção de uma garantia de correção temporal e de previsibilidade exige diversos cuidados na especificação e configuração do suporte de comunicação. A configuração do suporte requer uma escolha criteriosa dos parâmetros do sistema de comunicação, que envolve a determinação de protocolos e serviços de comunicação. Além disso, a configuração do suporte deve considerar a carga do sistema de comunicação e a arquitetura de comunicação adotada para o sistema.

A modelagem de aplicação exposta servirá de base para a construção do ambiente de teste para o sistema de controle, o qual possivelmente será implementado a partir de uma rede convencional, composta por PC's e sistema operacional QNX, caracteristicamente tempo real.

Comunicação em sistemas distribuídos tempo real tem sido alvo de diversas pesquisas [ARV 93] [FET 97] [FON 93] [KOP 93] [KOP 94] [KOY 88]. Entretanto, por se tratar de um assunto cujo interesse vem aumentando gradativamente, há uma série de tópicos relacionados ainda não explorados.

8. Referências Bibliográficas

[ARV 93] ARVIND, K. et. al. *A Local Area Network Architecture for Communication in Distributed Real Time Systems. In*: Advances in Real Time Systems, IEEE, Los Alamitos, California, 1993.

[BAB 85] BABAOGLU, O.; DRUMMOND, R. *Streets of Byzantium: Network Architectures for Fast Reliable Broadcast*. IEEE TOSE, v.11, n.6, 1985.

[BIR 87] BIRMAN, K.; JOSEPH, T. *Reliable Communication in the Presence of Failures*. ACM TOCS, NY, v. 5, n. 1, Feb. 1987.

[CRI 85] CRISTIAN, F. et. al. *Atomic Broadcast: From Simple Message Diffusion to Byzantine Agreement. In*: XV FTCS (Fault Tolerant Computer Systems), Ann Arbor, USA, 1985.

[CRI 90] CRISTIAN, F. *Synchronous Atomic Broadcast for Redundant Broadcast Channels*. Journal of Real Time Systems, NY, v. 2, n. 3, Sep. 1990.

[FET 97] FETZER, C.; CRISTIAN, F. *Real Time Systems*. Distributed Computing, Berlin, 1997.

[FON 93] FONSECA, K. O.; FARINES, J-M. *Uma Análise das Diversas Propostas de Atendimento dos Requisitos de Comunicação para Sistemas Tempo Real em Sistemas de Manufatura*. XI SBRC,1993.

[KAA 89] KAASHOEK, M. F. et. al. *An efficient reliable broadcast protocol*. Operating System Review, v. 2, 1989.

[KOP 92] KOPETZ, H. *Sparse Time versus Dense Time in Distributed Real Time Systems. In*: International Conference on Distributed Computing Systems, Yokohama, Japan, 1992.

[KOP 93] KOPETZ, H.; VERÍSSIMO, P. *Real Time and Dependability Concepts. In*: Mullender, S. J. (Ed.) Distributed Systems. NY, 1993.

[KOP 94] KOPETZ, H. *Protocol for Real Time Systems*. Computer, NY: IEEE, Jan. 1994.

[KOY 88] KOYMANS, R. et. al. *Paradigms for Real Time Systems. In*: Symposium on Formal Technicques in Real Time and Fault-Tolerant Systems, Warwick, Gran Bretanha, 1988.

[SAN 91] SANTOS, J. *Redes Locales en Tiempo Real*. Nova Friburgo: EBAI, 1991.

[RUB 94] RUBIRA, C. M. F. *Structuring Fault-Tolerant Object-Oriented Systems Using Inheritance and Delegation*. PhD Thesis, University of Newcastle upon Tyne, 1994.

[STA 88] STANKOVIC, J. A.; RAMAMRITHAM, K. *Hard Real Time Systems*. NY: IEEE, 1988.

[VER 90] VERÍSSIMO, P.; MARQUES, J. *Reliable broadcast for fault-tolerance on local computer networks. In*: IX Symposium on Reliable Distributed Systems, Huntsville: IEEE, 1990.