Análise de tolerância a falhas no protocolo SNMP

André Peres Ingrid Jansch-Pôrto
{peres, ingrid}@inf.ufrgs.br
Curso de Pós-Graduação em Ciência da Computação
Instituto de Informática - UFRGS
Caixa Postal 15064 - CEP 91501-970
Porto Alegre - RS - Brasil

Resumo

Este artigo apresenta as principais conclusões da análise do protocolo de comunicação SNMP - *Simple Network Management Protocol*, em suas características de funcionamento frente à tolerância a falhas e ao seu uso em sistemas de tempo real.

Sistemas de computação tempo real caracterizam-se por apresentar limites de tempo determinados para a realização das tarefas, sendo estes limites impostos pelo ambiente a ser controlado. É importante ressaltar que neste tipo de sistema, em geral a função de controle se revela como um aspecto crítico.

O SNMP é um protocolo de gerência de redes que se caracteriza por possuir nodos gerentes, responsáveis por coletar e analisar informações dos elementos que formam a rede, e nodos agentes, caracterizados como um conjunto de variáveis que representam o estado atual no nodo. Quando o gerente requisita informações aos nodos, realiza sua monitoração; ao alterar estes valores, exerce controle sobre o nodo.

A possibilidade de utilização do protocolo SNMP em aplicações de tempo real com características de tolerância a falhas depende de parâmetros funcionais do mesmo. Este trabalho analisa as características necessárias para a utilização do SNMP neste tipo de sistema a partir de informações teóricas.

Abstract

In this paper, we present the main results of the analysis of the communication protocol SNMP - Simple Network Management Protocol, considering its fault tolerance and real-time systems operational characteristics.

Real-time computational systems have pre-defined deadlines for tasks accomplishment, being the limits imposed by the environment to be controlled. It is important to stand out that, in this kind of system, in general, control appears as a critical element.

The SNMP protocol is a network management protocol which has management nodes, responsible for collecting and analyzing the information of the network elements; and agent nodes, which are a group of variables that represent the current node state. When the manager requests information to the nodes, these ones are being monitored, when changing these values, the nodes are being controlled.

The use of the SNMP protocol in real-time applications with fault tolerance characteristics depends on factors which are related to the way that this protocol will accomplish its functions. This paper analyzes the necessary characteristics for using SNMP in this kind of system considering theoretical information.

1 Introdução

A gerência de redes é responsável por restabelecer o bom funcionamento de uma rede de computadores em situações problemáticas, através da análise dos

diversos elementos que compõem a rede e da alteração de características destes elementos [ROS96].

Um sistema de gerência de redes deve conter cinco componentes: um ou mais nodos gerenciados, cada um contendo um agente; pelo menos uma estação de gerência de rede, contendo uma ou mais aplicações de gerência; entidades capazes de realizar tanto a tarefa de agente quanto a de gerente; um protocolo de gerência de redes, o qual é utilizado para realizar a troca de informações entre os nodos de gerência e os agentes; e informação de gerência nos agentes.

Um nodo gerenciado é um dispositivo qualquer que possui funções em uma rede, como terminais e computadores de grande porte, passando por impressoras, roteadores, *bridges*, *hubs* ou *modems*.

O principal axioma da gerência de redes, segundo Rose [ROS96], é: "O impacto de adicionar gerência de redes em nodos gerenciados deve ser mínimo, refletindo o mínimo denominador comum". Desta forma, tem-se que as aplicações de gerência de redes devem realizar suas funções nesta grande variedade de equipamentos, interferindo o mínimo possível na operação normal destes dispositivos.

Uma estação de gerência corresponde a um servidor executando o protocolo e aplicações de gerência de redes. Cada nodo gerenciado é visto como um conjunto de variáveis que representam informações referentes ao nodo. Ao disponibilizar estas variáveis à leitura, o nodo permite seu monitoramento e, ao alterar seus valores, o nodo está sendo controlado.

Além das operações de leitura (monitoramento) e escrita (controle), existem as operações: transversal, que permite a uma estação de gerência determinar qual o conjunto de variáveis suportadas pelo nodo; e a de interrupção (*trap*), que informa ao nodo de gerência a ocorrência de alguma exceção na estação gerenciada.

O protocolo de gerência de redes SNMP (*Simple Network Management Protocol*) realiza suas funções entre nodos agentes e gerentes, através de uma base de informações de gerência MIB (*Management Information Base*) composta por um conjunto de variáveis. Estas variáveis ficam disponíveis aos gerentes para consulta ou alteração através da troca de mensagens definidas pelo próprio protocolo.

Para a troca de mensagens, o SNMP utiliza-se do protocolo de transporte não orientado à conexão UDP, o qual é recomendado pelos autores do SNMP como o método mais prático de transporte. Em uma mensagem SNMP, estão contidas as informações referentes ao tipo de mensagem que está sendo enviada (leitura, escrita, transversal ou interrupção), além dos valores referentes à base de informações do agente.

Ao receber uma mensagem de consulta, cabe ao agente fornecer as informações da MIB que atendem à requisição do gerente. Caso a mensagem for de escrita, é realizada uma chamada de procedimento remoto RPC (*Remote Procedure Call*) no agente para efetuar as operações correspondentes. Uma alteração de valor em uma MIB pode ser responsável pela reconfiguração do agente, alterando as características da rede. As mensagens transversais são utilizadas pelo gerente para a obtenção da árvore que compõe a MIB, permitindo recuperar as variáveis que estão disponíveis no agente. Já a mensagem de interrupção ocorre quando é identificado um evento extraordinário no agente. Este agente deve, então, informar ao gerente a ocorrência deste evento. Cabe ao gerente realizar as operações necessárias para tratar uma interrupção.

Durante a comunicação, o protocolo SNMP aguarda a resposta sem interromper suas operações, ou seja, enquanto a resposta a uma requisição não é recebida, o SNMP continua realizando outras tarefas. O SNMP possui um período

máximo de entrega de mensagens, sinalizado pela ocorrência de *time-out*; esta circunstância indica o congestionamento da rede, a queda de um nodo ou a perda de mensagens, possibilitando a identificação de problemas na rede ou nos nodos.

Para realizar a verificação da integridade dos dados que estão sendo trocados entre as entidades SNMP, este protocolo realiza apenas a análise de informações referentes à comunidade. A comunidade é definida como um conjunto de caracteres ASCII que serão utilizados para realizar o relacionamento entre entidades SNMP. Esta análise simples dos dados da comunidade, assim como o tráfego desprotegido das informações representam fraquezas no aspecto de segurança do protocolo, permitindo o ataque de um intruso a informações trocadas entre as entidades.

Com o objetivo principal de solucionar alguns problemas, foi criado o SNMPv2. Esta versão do protocolo possui mecanismos adicionais capazes de contornar os principais problemas relativos à segurança e descrição dos dados, tais como funções de privacidade de dados, autenticação e controle de acesso.

Infelizmente, a segunda versão do protocolo SNMP existe apenas em teoria. Seus criadores encontraram divergências em vários aspectos quanto à forma de implementação dos objetivos (principalmente em relação à segurança) deste protocolo, o que impediu a sua finalização. Em 1995, o grupo de trabalho responsável pelo SNMPv2 foi dividido em subgrupos para estudo de alternativas de modelos de segurança diversos. Em agosto de 1996, foi proposta a resolução destes modelos em um modelo definitivo. Março de 1997 marcou o início dos trabalhos no SNMPv3, retomando as atividade dos grupos que trabalhavam nos modelos do SNMPv2, tendo por principal objetivo resolver os problemas relativos à segurança do protocolo [HAR97].

O presente trabalho objetiva apresentar os principais aspectos da análise do protocolo de gerência de redes SNMP, levando em consideração as características desejáveis para este tipo de protocolo nas áreas de tolerância a falhas e sistemas de tempo real. Os aspectos de segurança mencionados não causam efeito nesta análise, tendo em vista que os sistemas de tempo real considerados como principal alvo de interesse no uso do SNMP têm sido utilizados na automação industrial, onde a rede não possui ligações com o exterior, impossibilitando ataques externos. A versão completa da análise compõe a monografia subscrita pelo autor principal deste trabalho [PER97].

2 Propriedades básicas

Denardin [DEN97] estudou as propriedades necessárias ou características desejáveis para aplicações em tempo real, considerando diversos níveis de implementação na computação (sistemas operacionais, linguagens e protocolos). Foi feita a análise teórica de alguns protocolos, levando em consideração as seguintes características: detecção de erros (corrupção de mensagens, perda de mensagens ou de nodos) entrega de mensagens, flexibilidade, comportamento previsível do protocolo, sincronismo e tolerância a falhas. Este trabalho propõe utilizar as características pré-selecionadas por Denardin para realizar a análise do protocolo SNMP, as quais são brevemente reportadas a seguir com os elementos obtidos naquela referência.

- Detecção de erros

Vários sistemas de tempo real realizam o controle do ambiente (em aplicações críticas, por exemplo) sem intervenção humana. Estes sistemas devem realizar a identificação da ocorrência de erros de maneira ágil, permitindo que o sistema reaja ao erro antes que este seja propagado ao ambiente. Estes protocolos necessitam oferecer suporte adequado para detectar a presença de erros, tais como: a corrupção de

mensagens (mensagens nas quais há perda parcial ou alteração do conteúdo), mensagens perdidas (perda de mensagens previstas no protocolo, ocorridas durante a comunicação); perda do nodos (queda de um ou mais nodos da rede).

- Entrega com tempo determinado

Sistemas de tempo real caracterizam-se pela delimitação do intervalo de tempo para o recebimento de uma informação do ambiente, sua computação e o fornecimento (quando necessário) de respostas ao ambiente. Da mesma forma, os protocolos de comunicação a serem integrados a este tipo de aplicação, devem possuir condições de assegurar intervalos de tempo máximos para realizar a comunicação entre os nodos da rede. Os prazos de entrega de mensagens devem ser conhecidos e limitados.

- Flexibilidade

Esta característica indica se o protocolo é suficientemente flexível para ser utilizado em aplicações diversas ou se foi desenvolvido com limitações a uma arquitetura de rede ou aplicação específica.

O excesso de flexibilidade pode entrar em conflito com a capacidade de detecção de erros em protocolos de comunicação tempo real. A detecção de erros é atingida mais facilmente quando opera sobre um modelo de comportamento rígido, enquanto que esta rigidez traz limitações à flexibilidade.

- Comportamento previsível

A previsibilidade de um sistema de tempo real torna possível a determinação antecipada de seu comportamento. Com previsibilidade permite definir o momento de transmissão de uma mensagem, e o intervalo de tempo necessário à realização deste processo.

Esta característica engloba o comportamento do sistema em relação ao tempo de entrega de mensagens.

- Sincronismo

Esta característica corresponde à análise do suporte oferecido pelo protocolo para sincronização entre os relógios dos nodos. A sincronização é realizada através da comparação entre o instante de tempo em que uma mensagem chega ao nodo destino e o tempo previsto anteriormente para esta chegada.

- Tolerância a falhas

A tolerância a falhas avalia se o nodo é capaz de continuar com suas operações, embora com alguma degradação, após a ocorrência de uma falha. Esta característica é voltada para aspectos adicionais à detecção de erros, tais como a capacidade de realizar atividades para avaliação e confinamento dos danos, recuperação e tratamento dos erros.

As falhas de comunicação: temporais, caracterizadas por atrasos provenientes de uma sobrecarga do sistema; de omissão, que ocorrem devido a erros durante a transmissão; e de particionamento da rede, provenientes de defeitos no meio físico, devem ser consideradas por sistemas de tempo real distribuídos que possuem características de tolerância a falhas.

É importante ressaltar que, assim como comentado com relação ao par flexibilidade - detecção de erros, algumas características apresentadas podem ser conflitantes entre si, ou seja, um protocolo pode apresentar fortemente alguma característica que implique na redução ou inexistência de outra. Um outro conflito possível, por exemplo, é entre a utilização de técnicas que implementem tolerância a falhas e/ou detecção de erro, ocasionando atrasos na entrega de respostas do sistema, prejudicando conseqüentemente a entrega das mensagens nos prazos-limite do sistema.

3 Análise das características do SNMP

Tendo em vista o emprego do protocolo SNMP em operações de gerência em redes de computadores e em sistemas de automação industrial, ele também se torna um alvo de interesse para análise de acordo com a ótica apresentada por Denardin. Esta análise não traz resultados definitivos, mas pode ser considerada como base para a especificação e implementação de mecanismos que venham a corrigir deficiências detectadas pelo procedimento.

3.1 Detecção de erros

A detecção de erros pelo protocolo UDP é feita através da verificação do *checksum* contido em seus pacotes. O controle de mensagens corrompidas é realizado através da garantia de que caso o pacote UDP não for recebido por completo, ou se o cálculo de *checksum* não conferir com o seu conteúdo, a mensagem será descartada pelo protocolo; logo, apenas as mensagens corretas (completas e com cálculo de *checksum* correto) serão consideradas pelo protocolo SNMP. Apesar de identificar a corrupção da informação contida no pacote, nenhuma notificação será gerada ao protocolo SNMP.

As mensagens perdidas, particionamento da rede, assim como a perda de nodos, são identificadas pela aplicação de gerência através da ocorrência de *time-outs*. Caso a resposta a uma requisição não seja entregue no prazo determinado, a aplicação de gerência irá realizar as operações previstas de retransmissão de mensagens ou reconfiguração da rede.

Através da análise das informações recebidas pelos agentes, a aplicação de gerência é capaz de identificar uma série de problemas referentes ao nodo analisado podendo, assim, realizar operações para solucionar ou contornar estes problemas. É através destas ações que o SNMP permite às demais aplicações que utilizam a rede permanecerem em funcionamento.

3.2 Entrega com tempo determinado

O protocolo SNMP aguarda pelas respostas de suas requisições sem interromper suas operações. Este protocolo possui um intervalo de *time-out* definido pela aplicação, que determina o período máximo de tempo de espera pelo transporte de uma mensagem.

Através da utilização deste tempo máximo previsto para a troca de mensagens, é possível à aplicação determinar a ocorrência das situações de comunicação que excedam o prazo determinado, em decorrência de congestionamento da rede, perda de algum nodo ou perda da mensagem.

A determinação do tempo máximo para a troca de mensagens depende da configuração e dos equipamentos utilizados na rede. Com a construção de uma rede de dispositivos e a análise do tempo na troca de mensagens entre eles, é possível construir um sistema previsível que utilize o protocolo SNMP, respeitando os limites de tempo impostos por sistemas de tempo real. Este sistema irá identificar a ocorrência de atrasos na transmissão de mensagens entre os nodos através do *time-out*, possibilitando a identificação de falhas no sistema de tempo real.

3.3 Flexibilidade

O protocolo SNMP tem suas funções objetivamente dirigidas à gerência de redes de computadores, através da consulta e alteração de valores em bases de dados específicas (as MIBs) que se encontram distribuídas pelos nodos da rede. Não existe qualquer restrição explícita quanto à utilização do protocolo SNMP para outras aplicações.

Em relação ao conflito existente entre a detecção de erros e a flexibilidade dos protocolos, temos que a detecção de erros está baseada na forma simples como o

SNMP realiza suas funções, o que permite a previsibilidade de seu comportamento e a identificação de qualquer situação divergente com relação a esta previsão. Já a flexibilidade, é garantida pela maneira como são realizadas as operações de consulta e alteração de valores, possibilitando a utilização do protocolo SNMP em diversos tipos de sistemas.

3.4 Comportamento previsível

Apesar de possuir um meio de transporte não orientado à conexão, o SNMP tem condições de identificar a ocorrência de problemas de perda de mensagens e nodos, através da utilização do tempo máximo de transporte de mensagens. Além disso, a forma simples como este protocolo foi implementado, e a capacidade de receber informações referentes aos equipamentos que formam a rede, permite ao protocolo prever o comportamento destes equipamentos.

O protocolo possui condições de analisar o estado em que se encontram os nodos pertencentes à rede, podendo compor uma visão geral do estado da rede. Isto possibilita a identificação de problemas e a reconfiguração necessária para contorna-los.

3.5 Sincronismo

Apesar de não utilizar sincronismo na sua definição devido à necessidade de acréscimo de tarefas aos nodos agentes para realizar este tipo de controle, o protocolo SNMP apresenta condições para a implementação desta característica.

Para adicionar sincronismo nos nodos que utilizam o protocolo SNMP na gerência de redes de computadores, é necessário o acréscimo de funções de sincronismo aos nodos agentes da rede. Entretanto, o acréscimo de funções nos nodos agentes deve ser controlado para evitar que as funções básicas destes nodos sejam prejudicadas pela gerência.

3.6 Tolerância a falhas

O gerente possui uma visão do estado dos nodos agentes, através da requisição de informações destes nodos. Isto possibilita a identificação de problemas nos dispositivos gerenciados e da ocorrência de congestionamento em determinado ponto da rede, além de outros problemas específicos de configuração dos equipamentos que compõem a rede os quais serão analisados e contornados pela aplicação de gerência. Desta forma, cabe ao SNMP identificar os problemas da rede.

Quando um gerente detecta um nodo com problemas, cabe à aplicação de gerência realizar a reconfiguração da rede de maneira a mascarar ou confinar o erro encontrado. Tem-se então que a aplicação de gerência é a responsável por realizar operações para solucionar os problemas nos sistemas que utilizam SNMP, considerando as características do transporte de mensagens realizadas por este protocolo.

O protocolo utiliza o *time-out* como identificador de falhas temporais e omissão, mas a decisão de retransmissão de mensagens cabe à aplicação.

4. Conclusões

O gerenciamento de redes de computadores é uma forma de acrescentar confiabilidade. A função da aplicação de gerência é a de manter a rede funcionando adequadamente. A aplicação mantém uma visão global da rede, cuidando para que cada componente execute suas funções corretamente.

A gerência então realiza funções de tolerância a falhas para o funcionamento correto da rede, utilizando o conceito de mascaramento de falhas, quando necessário.

Quando não é possível à aplicação realizar suas funções de reconfiguração, é esperado que o usuário receba de maneira imediata a notificação da falha encontrada.

Segundo [BIR96], um protocolo que implemente um canal de comunicação confiável deve garantir, em princípio, que mensagens perdidas serão retransmitidas e que mensagens fora de seqüência serão reordenadas e enviadas na ordem original. O controle de fluxo e mecanismos que inibem o remetente, quando o volume de dados torna-se excessivo, também são comuns em protocolos para transporte confiável. O protocolo UDP não pode ser considerado como um protocolo confiável, devido às suas características de transmissão de mensagens.

A escolha desta forma de transporte pelos autores do SNMP deve-se ao objetivo principal da aplicação de gerência através deste protocolo: realizar operações de análise e alteração da configuração da rede inclusive quando esta se encontra nos piores estados, utilizando o mínimo de recursos e causando o mínimo impacto sobre as funções normais dos nodos.

Desta forma, tem-se que a melhor maneira de se efetuar gerência de redes com confiabilidade, é através do desenvolvimento de uma aplicação com técnicas de tolerância a falhas que utilizem a simplicidade da implementação do modelo do protocolo SNMP de maneira a não sobrecarregar os nodos envolvidos no gerenciamento, e considerando a falta de recursos de confiabilidade existentes no protocolo de comunicação UDP.

A construção de transmissão confiável pode ser alcançada através da utilização das informações de identificação das mensagens (para garantir a ordem), e os controles de perda de mensagens através da verificação dos tempos máximos de resposta. Estas operações devem ser realizadas através de mensagens SNMP, na aplicação de gerência.

O protocolo SNMP apresenta implementação simples, e características de flexibilidade e previsibilidade que tornam sua utilização viável nos mais diversos tipos de sistemas. Sistemas de tempo real podem realizar suas funções utilizando-se das características de troca de mensagens, estrutura de dados e simplicidade de modelo e funcionamento do protocolo SNMP.

5. Referências

[BIR96] Birman, Kenneth P. *Building Secure and Reliable Network.* Applications. Manning Publications Co. Greenwich, CT, 1996.

[DEN97] Denardin, Fernanda Kruel. *Software tolerante a falhas para aplicações tempo real.* Dissertação (mestrado) CPGCC-UFRGS, 1997.

[PER97] Peres, André. *Análise de tolerância a falhas no protocolo SNMP.* Trabalho Individual CPGCC-UFRGS, 1997.

[ROS96] Rose, Marshall T. *The Simple Book - An Introduction to Networking Management.* Revised Second Edition. Englewood Cliffs: Prentice-Hall, Inc. Upper Saddle River, NJ. 1996.

[HAR97] Harrington, David. *The Evolution of Architectural Concepts in the SNMPv3 Working Group.* The Simple Times - The Quarterly Newsletter of SNMP Technology, Comment, and Events. Volume 5, N.1 Dezembro, 1997. Disponível por www em http://www.simple-times.org.