CRONO: Um Gerenciador de Máquinas Agregadas para Linux

Marco Aurélio Stelmar Netto, César A. F. De Rose

Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS
Avenida Ipiranga 6681, Telefone: 5133203558 ramal 4463, Fax: 5133203758
stelmar@cpad.pucrs.br, derose@inf.pucrs.br

Introdução

Máquinas agregadas (*clusters*) vêm se tornando uma ótima alternativa em relação aos supercomputadores quando alto desempenho é necessário. Com uma boa relação custo/benefício essas máquinas estão se popularizando em universidades, laboratórios de pesquisa e indústrias. Como os agregados são compostos por diversos nós, e cada um deles com seu próprio sistema operacional, um dos problemas encontrados nesse ambiente é gerenciar esses nós de maneira única e eficiente, possibilitando um melhor aproveitamento dos recursos.

Estão disponíveis diversos sistemas de gerência de agregados (Cluster Management System - CMS), tais como: Computing Center Software [KEL 01], DQS [GRE 93] e Portable Batch System [HEN 95]. Embora esses sistemas sejam muito completos e configuráveis, eles são razoavelmente complexos de instalar, configurar e modificar. Por essa razão, se faz necessário um sistema de gerência de agregados que possua um alto nível de configurabilidade e que seja de simples de instalação, configuração e manutenção.

CRONO é um sistema gerenciador de agregados para o sistema Linux, que vem sendo desenvolvido no CPAD [CPA 02] desde o começo de 2002. Permite gerenciar diversos agregados com políticas de direitos de acesso distintas para cada agregado, suporte para alocação temporal e espacial, e possibilidade gerenciar grupos de usuários, além de outras funcionalidades que serão descritas nas próximas sessões.

Principais Funcionalidades

CRONO disponibiliza dois modos de alocação: espacial e temporal. O primeiro modo é utilizado quando o usuário precisa do uso exclusivo sobre os nós alocados, por exemplo, quando este está medindo o desempenho de sua aplicação. O segundo modo é utilizado em situações onde o usuário está, por exemplo, testando sua aplicação sem se preocupar se existem outros usuários que possam interferir no desempenho da mesma. Este segundo modo de alocação é bastante interessante em ambientes de aprendizado, permitindo que muitos alunos utilizem os nós ao mesmo tempo.

Outra característica do CRONO é sua flexibilidade para definir direitos de acesso. Através de arquivos de configuração, o administrador do sistema pode criar perfis de direitos de acesso e atribuí-los para usuários ou grupos de usuários. Para configurar o ambiente de execução, o CRONO provê *scripts* de pré e pós processamento das requisições. Quan-

do o tempo de um usuário começa, o CRONO executa dois *scripts*: um especificado pelo administrador e outro pelo próprio usuário. No pós processamento, após o término do tempo do usuário, dois *scripts* são utilizados, da mesma forma que no pré processamento. Esse mecanismo é bastante útil para, por exemplo, automatizar a criação de arquivos de máquinas do MPI [GRO 96].

Arquitetura do CRONO

A arquitetura do sistema CRONO é composta por quatro módulos que se comunicam através de *sockets* (TCP/IP) [STE 97]:

- Interface do Usuário (IU) é o conjunto de ferramentas que possibilita o acesso aos recursos dos agregados;
- Gerenciador de Acesso (GA) é responsável pela verificação dos diretos de acesso dos usuários;
- Gerenciador de Requisições (GR) faz o escalonamento das requisições dos usuários e a preparação do ambiente de execução;
- Gerenciador do Nó (GN) é o módulo que roda em cada nó do agregado e sua principal funcionalidade é bloquear o acesso sobre o nó.

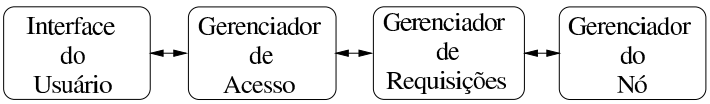

Figura 1: Arquitetura do CRONO.

Os módulos do CRONO são organizados como uma cadeia (Figura 1), isto é, se a Interface do Usuário precisa enviar uma mensagem para o Gerenciador do Nó, a mensagem passará pelo Gerenciador de Acesso e Gereciador de Requisições.

Interface do Usuário

A interface do usuário é responsável por prover interação com o sistema, através da UNIX *shell* (*e.g. bash, tcsh* ou *csh*). Com seis comandos, o usuário pode alocar nós do(s) agregado(s), visualizar informações sobre a(s) fila(s) de requisições, direitos de acesso, liberar os recursos e executar operações definidas pelo administrador diretamente sobre os nós. O CRONO também disponibiliza alguns programas e *scripts* para a criação de arquivos de máquinas, compilação e execução de programas MPI. Para facilitar a execução dos comandos, o CRONO possibilita ao usuário definir argumentos padrão que são passados para esses comandos, visto que alguns argumentos são geralmente os mesmos durante uma sessão. Dentre os argumentos, pode-se citar o número de nós para alocação, tempo de alocação e o agregado utilizado.

Gerenciador de Acesso

O Gerenciador de Acesso (GA) é o módulo do CRONO responsável por receber as requisições dos usuários da IU e validá-las, antes de repassar para o Gerenciador de Requisições, se necessário. O GA pode gerenciar vários agregados, podendo utilizar direitos de acesso distintos para cada um deles. O CRONO permite ao administrador atribuir direitos de acessos para usuários e grupos de usuários. Para isso, o GA utiliza três arquivos: o `groups` que define os grupos de usuários, o `accessrights.defs` que define os direitos de accesso e o `accessrights.users` onde são atribuidos os direitos de acesso.

Os direitos de acesso são definidos pelo tempo máximo e quantidade máxima de nós que um usuário pode utilizar, tanto para alocações quanto para reservas para uso futuro. Ainda há a possibilidade de definir direitos de acesso baseados em períodos do dia e dias da semana.

Gerenciador de Requisições

O Gerenciador de Requisições (GR) tem como objetivo escalonar as requisições dos usuários e preparar o ambiente de execução. O CRONO tenta aproveitar os recursos disponíveis que seriam desperdiçados usando algoritmo FIFO, mas sem prejudicar os usuários que estejam da fila de espera. Ou seja, se um usuário espera ser atendido num determinado horário previsto pelo escalonador, este será atendido no máximo naquele horário, podendo ser atendido antes caso algum usuário que esteja a sua frente libere os recursos antes do tempo requisitado.

Em alguns casos, os usuários estão apenas depurando suas aplicações e conseqüentemente não precisam de acesso exclusivo aos nós alocados. Em outros, os usuários precisam ter acesso exclusivo aos nós, ou por estarem analisando desempenho, ou pelo fato da aplicação precisar de alto poder computacional. Portanto, se faz necessário prover dois modos de alocação: a temporal e a espacial respectivamente.

A execução de operações ao início (pré processamento) e término (pós processamento) do tempo de um usuário é uma característica que facilita a configuração do ambiente de execução. Pode ser, por exemplo, criar arquivos de máquinas para o MPI e disparar *scripts* que enviam *e-mail* para o usuário sobre o início e término do seu tempo.

Além do pré processamento no início do tempo do usuário, o CRONO envia uma mensagem para o usuário através do seu terminal (`tty`) avisando que os recursos já estão disponíveis. Como os usuários podem estar usando mais que um terminal, o CRONO usa o arquivo `utmp` (arquivo que possui informações sobre os usuários logados no sistema operacional) para descobrir o terminal com menor ociosidade e envia a mensagem para este terminal. Isso é feito pois a probabilidade que o usuário esteja lendo esse terminal é maior.

Gerenciador do Nó

O Gerenciador do Nó (GN) é responsável por fornecer e alterar informações dos nós do agregado, através de requisições feitas pelo GR. A principal funcionalidade deste

módulo é não permitir que usuários façam acesso indevido aos nós sem estarem devidamente autorizados. Para controlar o acesso sobre os nós, o CRONO altera arquivos de configuração do Linux. Como algumas distribuições do Linux utilizam o PAM (`Pluggable Authentication Modules`) e outras não, a alteração desses arquivos deve ser diferenciada. Nas distribuições onde é utilizado o PAM, o CRONO altera apenas o arquivo `login.access` e nas distribuições que não utilizam o PAM, são alterados os arquivos `login.access` e `hosts.equiv`. Esses arquivos são alterados sempre ao início e término do tempo de um usuário. Além do controle de acesso dos nós, o GN é utilizado para executar operações diretamente sobre os nós. O administrador pode criar operações que os usuários geralmente precisam executar, como por exemplo, interromper processos que esteja executando nos nós. Isso faz com que o usuário não precise criar *scripts* ou conectar-se em cada máquina e executar essa operação.

Considerações Finais

O CRONO é uma alternativa em relação aos sistemas mais complexos como o CCS, PBS e DQS para a gerência de máquinas agregadas de pequeno e médio porte, especialmente por ter um alto grau de configurabilidade e possuir um código de menos que 7000 linhas, enquanto que o CCS por exemplo possui mais de 120.000 linhas.

O CPAD está utilizando o sistema CRONO desde janeiro deste ano para gerenciar quatro máquinas agregadas, atendendo diversos perfis de usuários e grupos de usuários. O CRONO está bastante estável e seu código está disponível sob a licença GPL no site `http://www.cpad.pucrs.br/crono`.

Referências

[GRE 93] GREEN, T. P. et al. DQS, **A Distributed Queueing System**. Annual Review of Scalable Computing. Mar. 1993.

[HEN 95] HENDERSON, R.L. et al. **Portable Batch System: Requirement Specification**. NAS Technical Report, NASA Ames Research Center, Apr. 1995.

[KEL 01] KELLER, A. et al. **Anatomy of a Resource Management System for HPC Clusters**. Annual Review of Scalable Computing, v.3, 2001.

[GRO 96] GROPP, W. et al. **A high-performance, portable implementation of the MPI message passing interface standard**. Parallel Computing, v.22, 1996;

[STE 97] STEVENS, W. R. **UNIX Network Programming Vol 1: Networking APIs - Sockets and XTI**. Prentice-Hall, 1997.

[CPA 02] CPAD - Centro de Pesquisa em Alto Desempenho http://www.cpad.pucrs.br, 2002.