

Distribuição de tarefas em computação distribuída

Estudo de caso: BOINC

Alex Massao Morinaga, unown.a@gmail.com
Hugo Posca de Vasconcelos, hugo.posca@gmail.com

Orientador: Prof. Dr. Alfredo Goldman vel Lejbman

Novembro de 2009

Computação distribuída

O que é?

Computação distribuída

O que é?

- 1 Em um mesmo computador, processos com memória própria

Computação distribuída

O que é?

- 1 Em um mesmo computador, processos com memória própria
- 2 Computadores diferentes ligados em rede

Computação distribuída

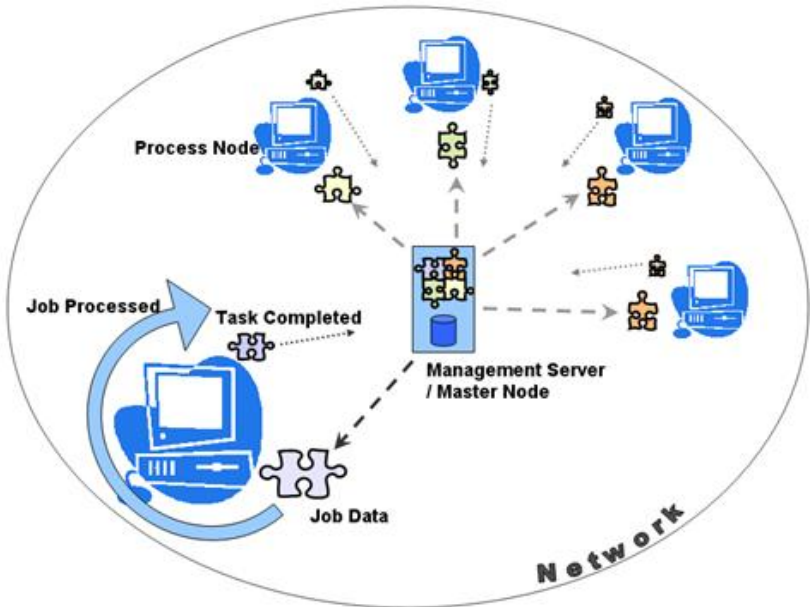
O que é?

- 1 Em um mesmo computador, processos com memória própria
 - 2 Computadores diferentes ligados em rede
- Em ambos os casos, um processo é independente do outro

Computação distribuída

O que é?

- ① Em um mesmo computador, processos com memória própria
- ② Computadores diferentes ligados em rede
 - Em ambos os casos, um processo é independente do outro
 - Comunicação por mensagens



Computação distribuída

Por que utilizar?

Computação distribuída

Por que utilizar?

- A própria natureza da aplicação pode necessitar uma rede de computadores

Computação distribuída

Por que utilizar?

- A própria natureza da aplicação pode necessitar uma rede de computadores
- Opção mais barata do que um computador com o mesmo desempenho

Computação distribuída

Por que utilizar?

- A própria natureza da aplicação pode necessitar uma rede de computadores
- Opção mais barata do que um computador com o mesmo desempenho
- Tolerância a falhas: problema em uma máquina não compromete o sistema inteiro

Computação em grade

- Cada parte da grade pode agir como produtor ou consumidor de recursos
- Não existe anonimidade



Computação voluntária

Voluntários:

- Oferecem recursos computacionais a projetos que têm interesse

Computação voluntária

Voluntários:

- Oferecem recursos computacionais a projetos que têm interesse
- São anônimos

Computação voluntária

Voluntários:

- Oferecem recursos computacionais a projetos que têm interesse
- São anônimos
- Devem confiar nos projetos que estão dispostos a ajudar

Computação voluntária

Projeto:

- Quanto mais voluntários, melhor

Computação voluntária

Projeto:

- Quanto mais voluntários, melhor
- Por isso, projetos precisam fazer “propaganda” para chamar a atenção de várias pessoas

BOINC



Berkeley **O**pen **I**nfrastructure for **N**etwork **C**omputing

O que é

- Infra-estrutura para computação distribuída

O que é

- Infra-estrutura para computação distribuída
 - Computação em grade
 - Computação voluntária

O que é

- Infra-estrutura para computação distribuída
 - Computação em grade
 - Computação voluntária
- Aplicação mais comum: mestre-escravo

O que é

- Infra-estrutura para computação distribuída
 - Computação em grade
 - Computação voluntária
- Aplicação mais comum: mestre-escravo
- Desenvolvido sob LGPL (GNU Lesser General Public License)

O que é

- Infra-estrutura para computação distribuída
 - Computação em grade
 - Computação voluntária
- Aplicação mais comum: mestre-escravo
- Desenvolvido sob LGPL (GNU Lesser General Public License)
- Desenvolvido originalmente para o projeto SETI@home (Search for Extraterrestrial Intelligence)

O que é

Atualmente com

- 62 Projetos afiliados

O que é

Atualmente com

- 62 Projetos afiliados
 - Malariaccontrol.net
 - SETI@Home
 - Einstein@home
 - PrimeGrid
 - ClimatePrediction.net
 - NQueens@Home

O que é

Atualmente com

- 62 Projetos afiliados
 - Malariaccontrol.net
 - SETI@Home
 - Einstein@home
 - PrimeGrid
 - ClimatePrediction.net
 - NQueens@Home
- Aproximadamente 327.000 voluntários ativos

O que é

Atualmente com

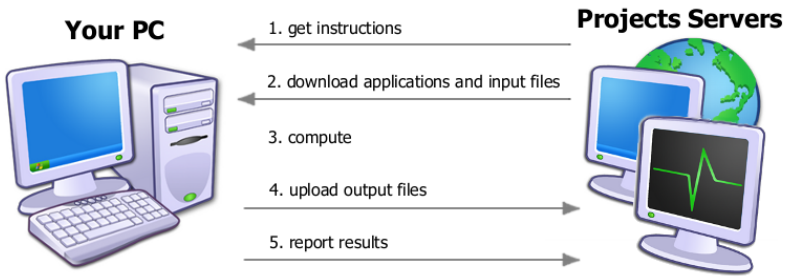
- 62 Projetos afiliados
 - Malariaccontrol.net
 - SETI@Home
 - Einstein@home
 - PrimeGrid
 - ClimatePrediction.net
 - NQueens@Home
- Aproximadamente 327.000 voluntários ativos
- Aproximadamente 586.000 computadores ativos

O que é

Atualmente com

- 62 Projetos afiliados
 - Malariaccontrol.net
 - SETI@Home
 - Einstein@home
 - PrimeGrid
 - ClimatePrediction.net
 - NQueens@Home
- Aproximadamente 327.000 voluntários ativos
- Aproximadamente 586.000 computadores ativos
- Média a cada 24 horas: 2,7 PetaFLOPS
(IBM RoadRunner: 1,5 PetaFLOPS)

Como funciona



http://boinc.berkeley.edu/wiki/How_BOINC_works

Sistema de créditos

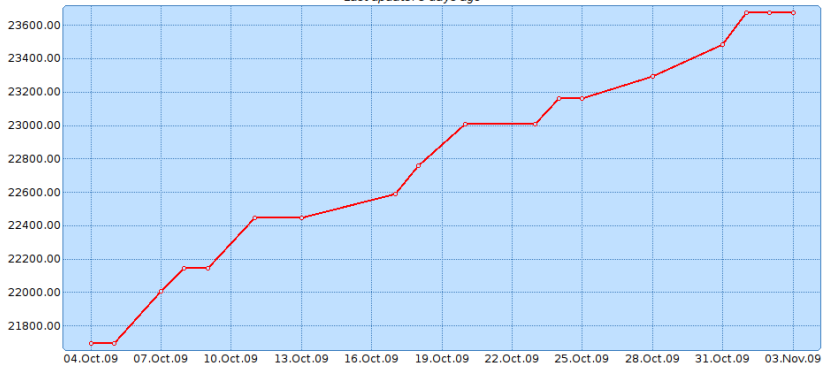
- Créditos são dados para demonstrar o quanto um usuário participou do projeto

Sistema de créditos

- Créditos são dados para demonstrar o quanto um usuário participou do projeto
- Eles são fornecidos após o usuário devolver o resultado para o servidor, e este confirmar o resultado

Sistema de créditos

- Créditos são dados para demonstrar o quanto um usuário participou do projeto
- Eles são fornecidos após o usuário devolver o resultado para o servidor, e este confirmar o resultado
- Este processo pode demorar alguns dias

User Total*Project: Einstein@Home**Account: Hugo**Team:**Last update: 5 days ago*

Segurança

Alguns cuidados relacionados à computação voluntária:

Segurança

Alguns cuidados relacionados à computação voluntária:

- Falsificação de respostas e créditos

Segurança

Alguns cuidados relacionados à computação voluntária:

- Falsificação de respostas e créditos
 - Sem replicação
 - Com replicação
 - Com replicação adaptativa

Segurança

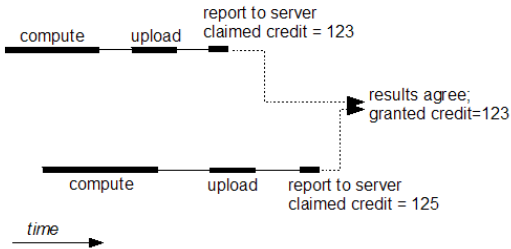
Alguns cuidados relacionados à computação voluntária:

- Falsificação de respostas e créditos
 - Sem replicação
 - Com replicação
 - Com replicação adaptativa
- Distribuição de executáveis maliciosos

Segurança

Alguns cuidados relacionados à computação voluntária:

- Falsificação de respostas e créditos
 - Sem replicação
 - Com replicação
 - Com replicação adaptativa
- Distribuição de executáveis maliciosos
- Falha de servidores de dados



Replicação

Segurança

- Roubo de contas por ataque ao servidor ou pela rede

Segurança

- Roubo de contas por ataque ao servidor ou pela rede
- Roubo de arquivos de entrada/saída

Segurança

- Roubo de contas por ataque ao servidor ou pela rede
- Roubo de arquivos de entrada/saída
- Abuso dos projetos aos participantes

Segurança

- Roubo de contas por ataque ao servidor ou pela rede
- Roubo de arquivos de entrada/saída
- Abuso dos projetos aos participantes
 - Intencional
 - Acidental

Linguagens aceitas

Originalmente, aceita programas escritos em C/C++. Com certos programas, também pode aceitar:

- FORTRAN
- Java (Windows apenas)
- Python

Programando

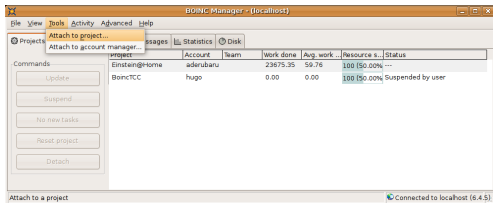
Um programa precisa utilizar as bibliotecas fornecidas pelo BOINC para poder funcionar corretamente.

As bibliotecas permitem:

- Checkpointing
- Criar seções críticas que não podem ser interrompidas pelo cliente
- Marcar o progresso do aplicativo
- Contabilizar os créditos

Cliente

- Roda na maioria dos sistemas operacionais
- Fácil instalação
- Fácil de se afiliar a um projeto
- Executa os programas em 'nice'
- Altamente configurável



Servidor

Para se criar um projeto que utilize o BOINC precisamos:

- PHP
- Apache
- MySql
- Código fonte do BOINC

File Edit View Bookmarks Widgets Tools Help

REPLACE WITH PROJECT NAME

http://localhost/BoincTCC/

REPLACE WITH PROJECT NAME

About REPLACE WITH PROJECT NAME

XXX is a research project that uses Internet-connected computers to do research in XXX. You can participate by downloading and running a free program on your computer.

XXX is based at [describe your institution, with link to web page]

- [Link to page describing your research in detail]
- [Link to page listing project personnel, and an email address]

Join REPLACE WITH PROJECT NAME

- [Read our rules and policies](#)
- This project uses BOINC. If you're already running BOINC, select [Attach to Project](#). If not, [download BOINC](#).
- When prompted, enter **http://201.92.183.32/BoincTCC/**
- If you're running a command-line or pre-5.0 version of BOINC, [create an account](#) first.
- if you have any problems, [get help here](#).

Returning participants

- [Your account](#) - view stats, modify preferences
- [Teams](#) - create or join a team
- [Certificate](#)
- [Applications](#)

Community


- [Profiles](#)
- [User search](#)
- [Message boards](#)
- [Questions and Answers](#)
- [Statistics](#)
- [Languages](#)

News

Sample news item title
March 1, 2004, 13:50 GMT
Sample news item

Another item title
March 2, 2004, 11:50 GMT
Another item

News is available as an [RSS feed](#)



Depois que o servidor estiver instalado precisamos:

- Criar Work Units
 - Arquivos XML que descrevem o trabalho e seus resultados
 - Os dados a serem computados
- Os programas que serão utilizados no projeto para verificar as repostas, analisar os dados, etc.

Exemplo de Work Unit

```
<file_info>
  <number>0</number>
</file_info>
<workunit>
  <file_ref>
    <file_number>0</file_number>
    <open_name>in</open_name>
  </file_ref>
  <min_quorum>1</min_quorum>
  <target_nresults>1</target_nresults>
</workunit>
```

Exemplo de Work Unit

Exemplo de Work Unit

```
<file_info>
  <name><OUTFILE_0/></name>
  <generated_locally />
  <upload_when_present />
  <max_nbytes>10000</max_nbytes>
  <url><UPLOAD_URL/></url>
</file_info>
<result>
  <file_ref>
    <file_name><OUTFILE_0/></file_name>
    <open_name>out</open_name>
  </file_ref>
</result>
```

Exemplo de Work Unit do resultado

O que fizemos

- Instalamos um servidor do BOINC
- Criamos um programa em C
 - Realizava alguns cálculos
 - “Queimava” ciclos de CPU para vermos a execução de forma gradual
- Distribuimos por alguns computadores
- Recebemos e analisamos as respostas

```
static void burnCPUtime() {  
    double a = 0;  
  
    for (int i = 0; i < 1<<17; i++)  
        for (int j = 0; j < 1<<17; j++)  
            a += 0.000001;  
}
```

Função utilizada para queimar ciclos da CPU

Conclusão

- Utilizar o cliente do BOINC enquanto se realizam tarefas cotidianas é quase imperceptível

Conclusão

- Utilizar o cliente do BOINC enquanto se realizam tarefas cotidianas é quase imperceptível
- Sua estrutura permite que usuários facilmente se afilem a um projeto, sem necessidade de configurações avançadas, o que permite que projetos tenham um grande alcance de público

Conclusão

- Utilizar o cliente do BOINC enquanto se realizam tarefas cotidianas é quase imperceptível
- Sua estrutura permite que usuários facilmente se afilem a um projeto, sem necessidade de configurações avançadas, o que permite que projetos tenham um grande alcance de público
- Permite que universidades, pequenas e médias empresas aproveitem todos os computadores disponíveis em suas redes, sem muitos custos adicionais, para realizarem tarefas que necessitem de muito poder computacional

Conclusão

- Utilizar o cliente do BOINC enquanto se realizam tarefas cotidianas é quase imperceptível
- Sua estrutura permite que usuários facilmente se afilem a um projeto, sem necessidade de configurações avançadas, o que permite que projetos tenham um grande alcance de público
- Permite que universidades, pequenas e médias empresas aproveitem todos os computadores disponíveis em suas redes, sem muitos custos adicionais, para realizarem tarefas que necessitem de muito poder computacional
- Devido à grande quantidade de informações que podem ser geradas e obtidas é recomendável ter uma equipe dedicada ao funcionamento do projeto

Agradecimentos



IME - USP

