

Utilização de Redes Neurais para Gerência de Servidores Virtuais Web

Danilo Souza¹ Iago Medeiros¹

¹Universidade Federal do Pará

20 de Junho de 2013

1 Introdução

- Introdução Geral
- Introdução à Virtualização
- Introdução à Redes Neurais Artificiais (RNA)

2 Trabalhos Relacionados

- Gerência de VM's
- Balanceamento de Carga utilizando RNA

3 Modelo de RNA utilizado

4 Arquitetura Proposta

- Parâmetros Analisados
- Intervenções da API
- Políticas de Reconfiguração

5 Testes

- Avaliação
- Workload utilizando site da NASA

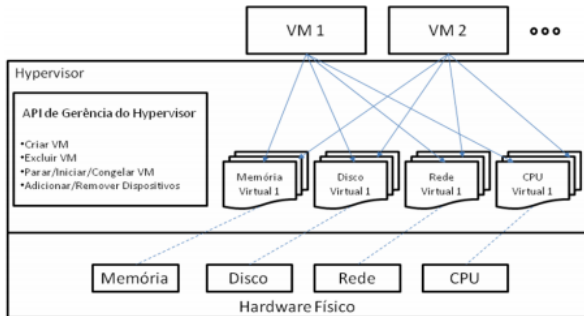
6 Conclusão e Trabalhos futuros

Introdução do Artigo

- Alto consumo de energia em *datacenters* (40% para equipamentos e 60% para infra-estrutura)
- Virtualizar servidores reduz o consumo de energia
 - Gera maior ociosidade
- Normalmente os equipamentos são superdimensionados (ociosidade)
- O autor propõe uma nova política de gerência de servidores *web*

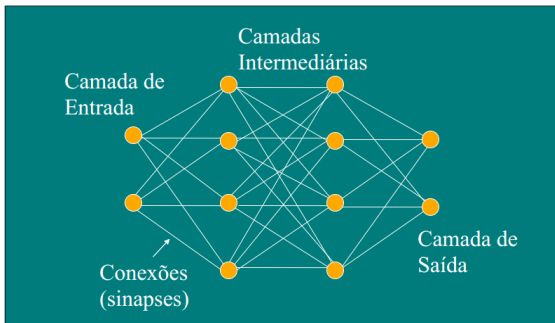
Arquitetura do Xen

- Utiliza para-virtualização
- Possui um *kernel* modificado (*dom0*)
- O *hypervisor* faz o mapeamento de recursos físicos em virtuais
- As VM's rodam no Domínio U (*domU*)



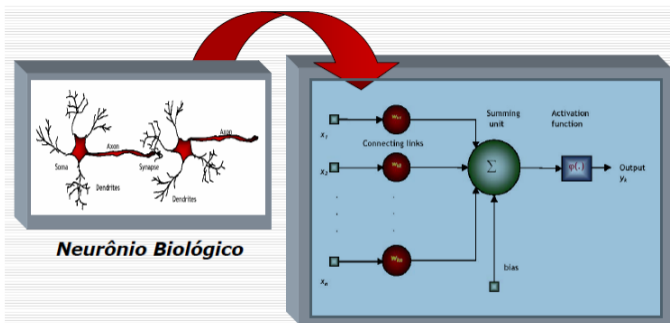
O que é uma RNA?

- São modelos computacionais
 - Adquirir, armazenar e utilizar conhecimento
 - Generalizar determinados sistemas
 - Agrupar e organizar dados
- Compostos por neurônios (unidades de processamento)
- Interligada por sinapses (conexões)
- Agrupados em camadas



Funcionamento da RNA

- Neurônios são utilizados para armazenar informações
- Cada sinapse é caracterizada por um peso
- O somador adiciona as entradas ponderadas
- A função de ativação restringe a saída entre $[0,1]$ ou $[-1,1]$
- O *bias* é utilizado para aumentar ou diminuir a entrada da função de ativação



Gerência de VM's

O autor cita três trabalhos sobre gerência inteligente de VM's

- Utilizando Balanceamento de carga
 - Analisa os recursos dos *hosts* [Megav 2007]
 - Migra as VM's de acordo com a quantidade de recursos disponíveis
- Utilizando um controlador *lookahead* [Kusic et al. 2008]
 - Analisa a quantidade de conexões e o tempo médio de resposta
 - Define a quantidade de *hosts* que devem ser ativados
- Utilizando múltiplos domínios [Ruth et al. 2006]
 - Um *host* e suas VM's formam um domínio
 - O sistema analisa a demanda de CPU e memória das VM's
 - O sistema migra as VM's caso os recursos de um domínio estejam esgotados

Balanceamento de Carga utilizando RNA

O autor cita três trabalhos referentes ao uso de RNA's para balanceamento de carga

- Controlar frequências de operação de CPU's [Souza et al. 2008]
 - Utiliza DFVS (*Dynamic Voltage and Frequency Scaling*) para controlar o *clock* das CPU's
 - Prove economia de energia mantendo o nível do serviço
 - Foram mapeadas frequência e taxa de ocupação das CPU's para associação com consumo de energia
- Classificar estados de um servidor *web* utilizando RNA [Simula et al. 2008]
 - Mapas auto-organizáveis com nove dimensões de entrada
 - Foram medidos uso de CPU, pacotes recebidos, blocos lidos/escritos do disco, etc.
 - A rede apresentou 4 estados de ocupação do servidor
 - Criação de uma política de tolerância a falhas baseada nesses estados

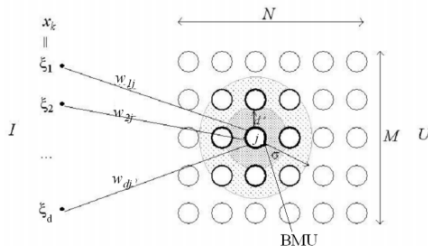
Balanceamento de Carga utilizando RNA

- Classificar uso de memória utilizando RNA [Lin 2006]
 - Utiliza 3 indicadores de consumo de memória como entrada
 - A rede indicou 6 estados de consumo de memória
 - Um módulo atuador foi projeto com base nos valores desses estados
 - Este módulo substitui o *OOM Killer* nativo do *Linux*

Mapas Auto-Organizáveis

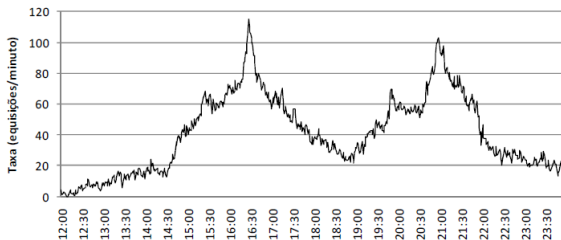
Faz analogia com a região do córtex cerebral

- Aloca regiões específicas para atividades específicas
- O grau de ativação dos neurônios diminui conforme se distancia da região de ativação inicial
- RNA com duas camadas (Entrada I e saída U)
- Geralmente são mapas $N \times M$



A RNA do Artigo

- 3 nós na interface de entrada (indicadores do cluster)
 - Tempo de resposta
 - Potência consumida
 - Taxa de requisições
- Uma mapa de 10x10 neurônios (captar o máximo de padrões possível)
- Treinamento *offline* usando *workload* de acesso ao site da Copa do Mundo de 1998



Introdução
oooo

Trabalhos Relacionados
ooo

Modelo de RNA utilizado

Arquitetura Proposta
ooo

Testes
oo

Conclusão e Trabalhos futuros

Introdução
○○○○

Trabalhos Relacionados
○○○

Modelo de RNA utilizado

Arquitetura Proposta
●○○

Testes
○○

Conclusão e Trabalhos futuros

Parâmetros Analisados

Introdução
oooo

Trabalhos Relacionados
ooo

Modelo de RNA utilizado

Arquitetura Proposta
o●o

Testes
oo

Conclusão e Trabalhos futuros

Intervenções da API

Introdução
oooo

Trabalhos Relacionados
ooo

Modelo de RNA utilizado

Arquitetura Proposta
oo●

Testes
oo

Conclusão e Trabalhos futuros

Políticas de Reconfiguração

Introdução
oooo

Trabalhos Relacionados
ooo

Modelo de RNA utilizado

Arquitetura Proposta
ooo

Testes
oo

Conclusão e Trabalhos futuros

Introdução
○○○○

Trabalhos Relacionados
○○○

Modelo de RNA utilizado

Arquitetura Proposta
○○○

Testes
●○

Conclusão e Trabalhos futuros

Avaliação

Introdução
oooo

Trabalhos Relacionados
ooo

Modelo de RNA utilizado

Arquitetura Proposta
ooo

Testes
o●

Conclusão e Trabalhos futuros

Workload utilizando site da NASA

Introdução
oooo

Trabalhos Relacionados
ooo

Modelo de RNA utilizado

Arquitetura Proposta
ooo

Testes
oo

Conclusão e Trabalhos futuros