

Estudo de Padrões de Acesso e Abordagens para Otimização de Acesso a Dados Distribuídos

Vinicius de Freitas Reis
Rodrigo Fernandes de Mello

Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo

II Escola Regional de Alto Desempenho de São Paulo –
ERAD-SP 2011

Motivação

- *Data Grids*;
- Replicação de arquivos:
 - tolerância a faltas, alto desempenho e balanceamento de carga;
- Ambientes globais, com milhares de computadores e arquivos;
- Problema de acesso a dados distribuídos;
- Discrepância entre tempos de acesso locais e remotos;

Objetivo

- Investigar o comportamento de acesso a dados de aplicações reais;
 - prever acessos a arquivos;
 - antecipar a transferência;
 - maximizar a taxa de acertos;
- Propor novas técnicas de replicação de arquivos.

Metodologia

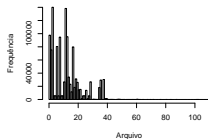
- Técnicas de modelagem de séries temporais foram estudadas;
- Duas técnicas de replicação de arquivos, baseadas na predição de séries, foram propostas;
- Traços de execução de aplicações reais foram convertidos em conjuntos de dados;
- Experimentos realizados:
 - Análise de Determinismo;
 - Análise de Padrões Frequentes;
 - Avaliação das técnicas de replicação.

Análise e Predição de Séries Temporais

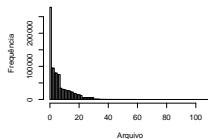
- Modelos estatísticos: AR, MA, ARMA, ARIMA;
- Sistemas Dinâmicos:
 - Teorema de Takens: $(x_t, x_{t-d}, x_{t-2d}, \dots, x_{t-(m-1)d})$;
 - Dimensão de separação: Autoinformação Mútua;
 - Dimensão embutida: Falsos Vizinhos mais próximos;
 - Regressão no espaço fase permite prever observações futuras;
 - Análise de comportamento: *Recurrence Plot*.

Conjuntos de Dados

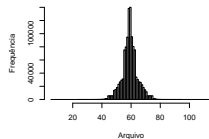
- Traços do repositório: SNIA IOTTA;
 - msn, bs: servidores de dados da Microsoft;
 - lair, deasna, home: servidores NFS;
 - seno, logis: dados sintéticos, com ruído aleatório;
- Variantes:



(a) variante A



(b) variante B



(c) variante C

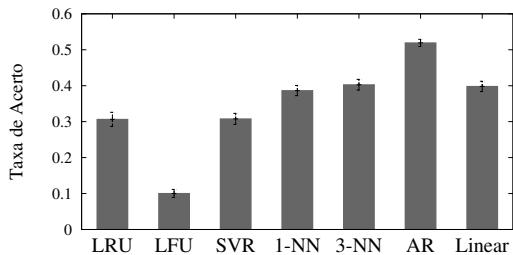
Técnicas de Replicação

- Algoritmos clássicos:
 - LRU;
 - LFU;
- Abordagens de predição;
- Duas técnicas:
 - AR;
 - Sistemas Dinâmicos.

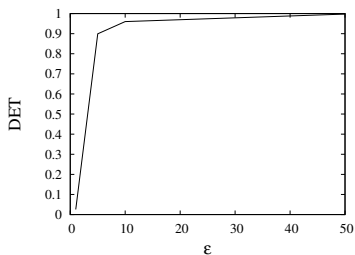
Experimentos

- Dois experimentos de análise:
 - Análise de determinismo (*Recurrence Plot*);
 - Análise de padrões frequentes (*Suffix Arrays*);
- Experimentos de avaliação de taxa de acerto;

Experimentos – seno

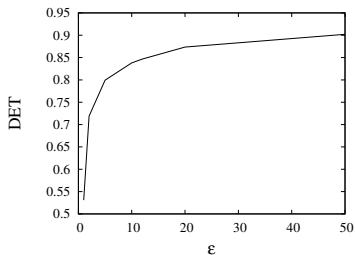
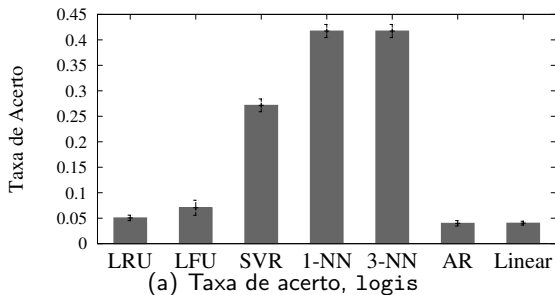


(a) Taxa de acerto, seno

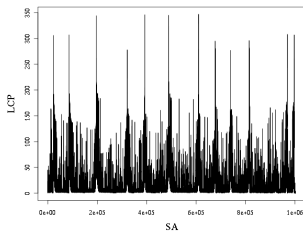


(b) Determinismo, seno

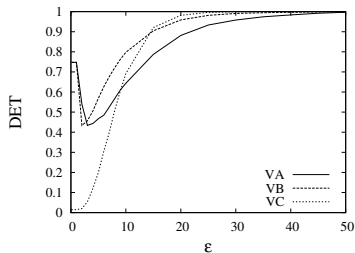
Experimentos – logis



Experimentos – msn

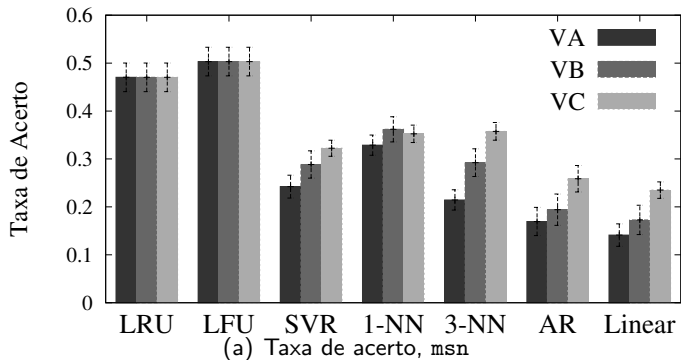


(a) LCP, msn

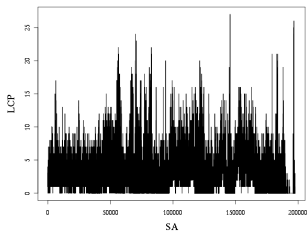


(b) Determinismo, msn

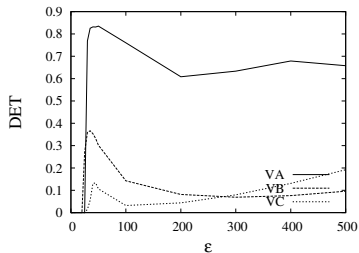
Experimentos – msn



Experimentos – lair

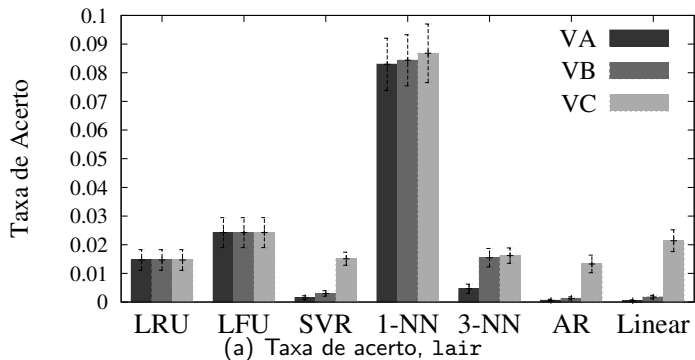


(a) LCP, lair



(b) Determinismo, lair

Experimentos – lair



Contribuições

- Proposta de duas técnicas de replicação de arquivos;
- Análises dos conjuntos de dados reais;
- Estudo da influência dos diferentes mapeamentos de arquivos em séries temporais na taxa de acerto das técnicas.

Trabalhos Futuros

- Variação dos parâmetros w e s ;
- Agrupamento de observações;
- Redução de ruídos;
- Diferentes mapeamentos;
- Outros modelos de predição.

Agradecimentos

