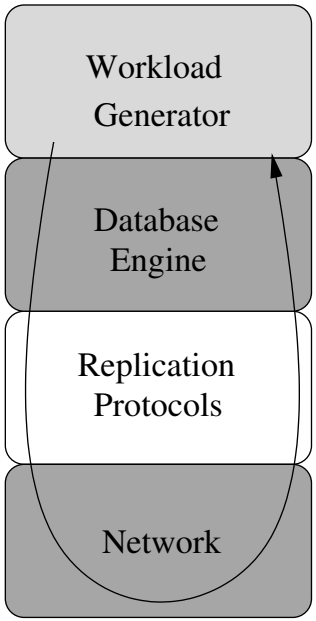
Evaluation of Group-based Database Replication Using Centralized Simulation

Dissertação de Mestrado

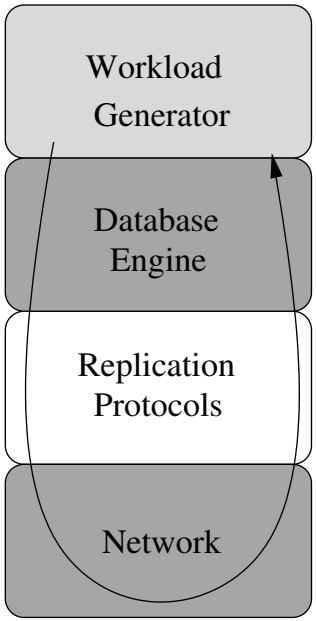
Luís Manuel Oliveira Soares
los@di.uminho.pt

Grupo de Sistemas Distribuídos
Departamento de Informática
Escola de Engenharia
Universidade do Minho

Julho 2006



- Teste e avaliação de protótipos de sistemas distribuídos.
- Replicação de bases de dados com comunicação em grupo.
- Verificar correcção e desempenho dos protocolos de replicação.
- Opções: Benchmarks e ambientes de testes ou simulação.
- Solução: combinar simulação e protótipos reais.



- Teste e avaliação de protótipos de sistemas distribuídos.
- Replicação de bases de dados com comunicação em grupo.
- Verificar correcção e desempenho dos protocolos de replicação.
- Opções: Benchmarks e ambientes de testes ou simulação.
- Solução: combinar simulação e protótipos reais.

Contribuições

- Combinar ambiente de simulação com concretizações reais.
- Criação e validação de modelos de simulação necessários.
- Estudo de protocolos síncronos para replicação de bases de dados.

Simulação Centralizada

- Combina simulação com protótipos reais.
- Execução real baseada em eventos.
- Suporte em simulação discreta por eventos.

Vantagens

- Execução real é reflectida em execução num recurso simulado.
- Oferece: determinismo, observação global, mínima intrusão, baixo custo e flexibilidade.
- Alia benefícios de simulação à possibilidade de teste de protótipos reais.

Desafios

- Reutilizar modelos previamente desenvolvidos e validados.
- Simplificar a integração de modelos com protótipos.
- Contabilização do tempo de execução real e consequente e parametrização de execução simulada de forma transparente.

Soluções

- Reutilizar a especificação do núcleo de simulação *SSF*.
- Distinguir entre processos simulados e processos de execução real.
- Criação de *proxies* que medeiam a interacção entre simulação e código real.
- Proxies realizam mudança de contexto e notificam o núcleo para começar/parar a contabilização de tempo.

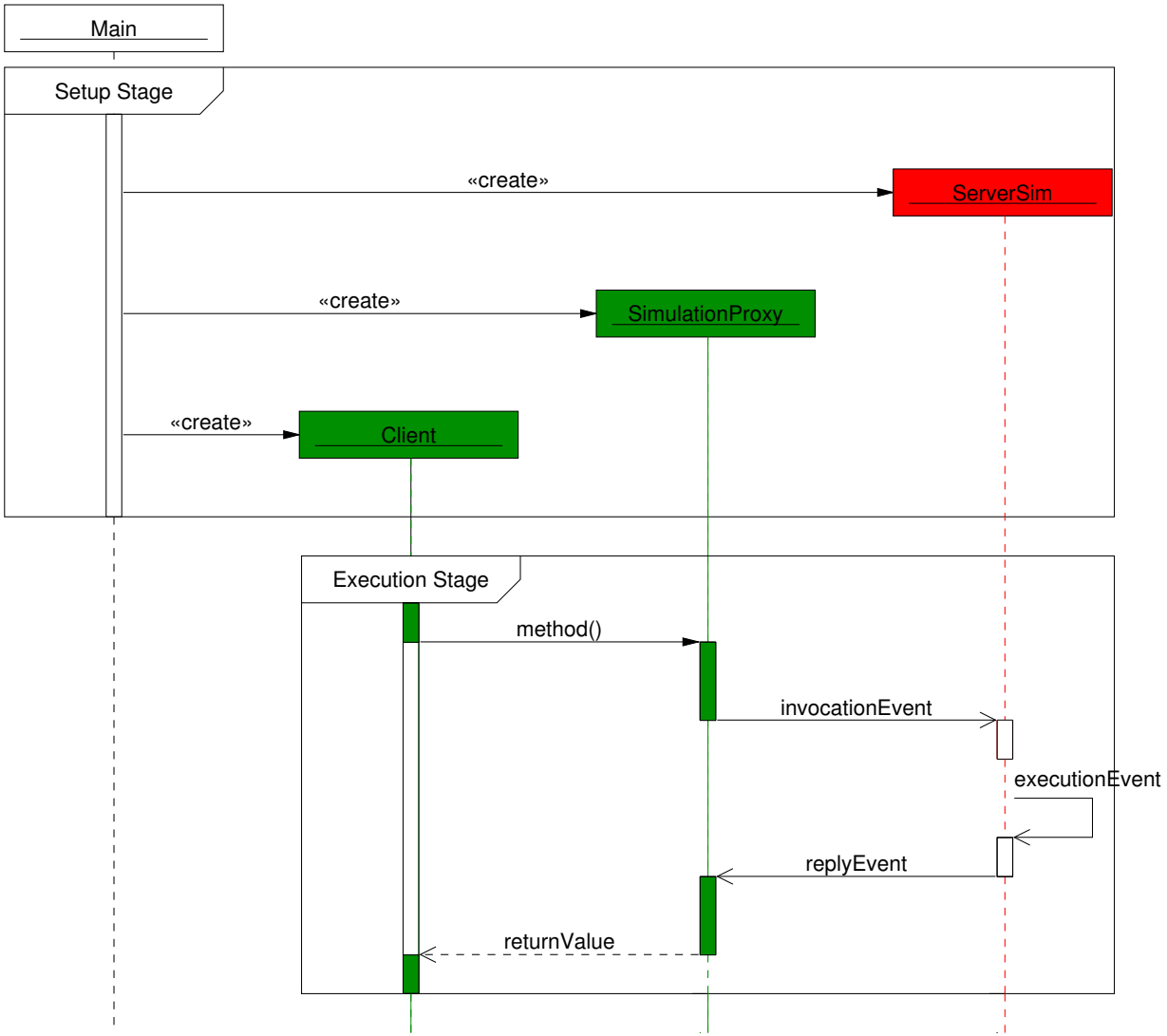


Figura: Proxy de Simulação.

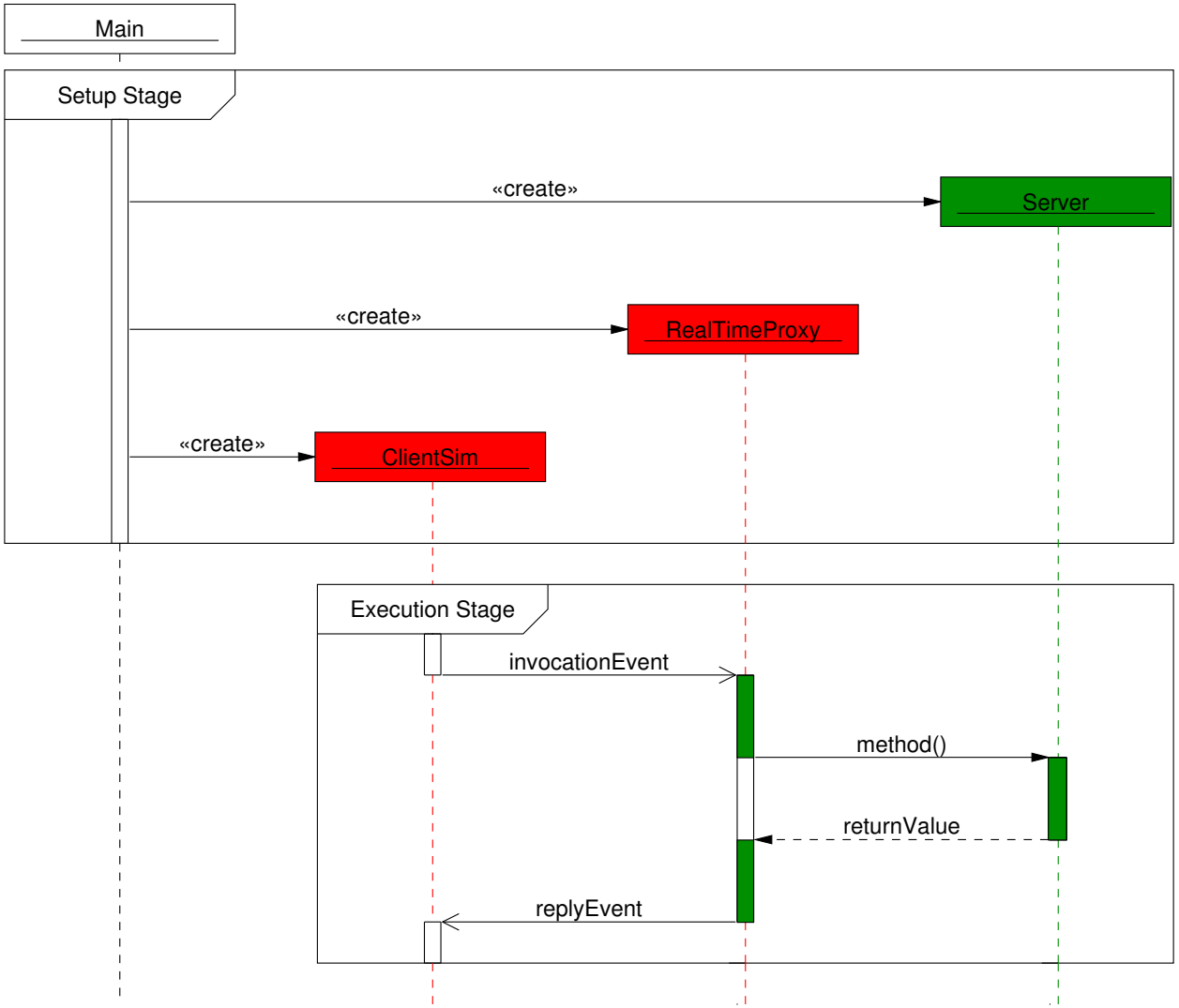
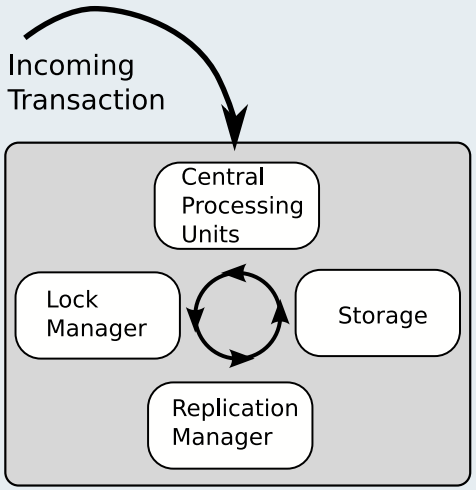


Figura: Proxy de Tempo Real.

Modelo de Rede

- Reutilizados a partir da biblioteca *SSFNet*.
- Componentes de rede, camadas protocolares e camada de aplicação.

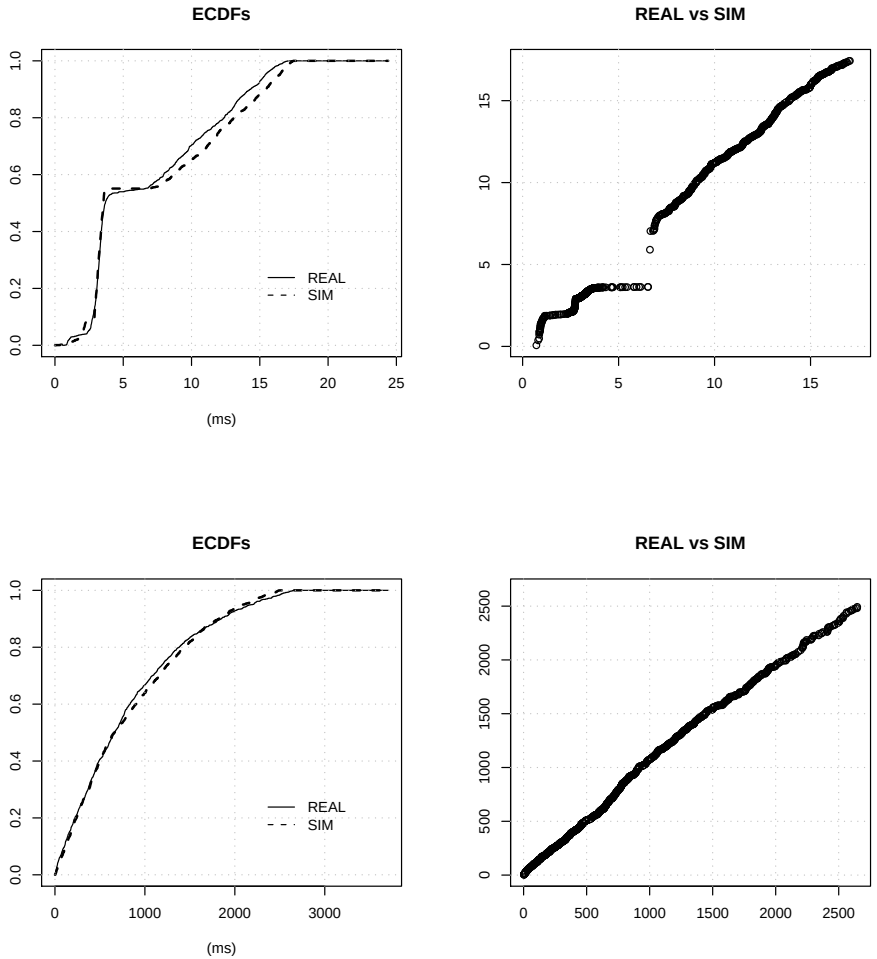
Modelo de Base de Dados



- Transacção: escritas, leituras e processamento.
- Base de dados: colecção de recursos - processadores, controlo concorrência, armazenamento e replicação.
- Recursos modelados como filas de espera.
- Nível de abstracção determinado pelos requisitos do projecto.

Calibração

- Parametrizar o modelo de forma a aproximar à realidade.
- Tendo por base um sistema real instrumentado:
  - CPU: parametrizado com tempos recolhidos.
  - I/O de armazenamento, conforme número de pedidos de acesso a log para escrita.
  - Rede: modelos validados pelos autores. Pequena calibração realizada ao nível do modelo do sistema operativo.
- Decorreu em paralelo com a fase de modelação, fornecendo precioso *feedback*.
- Forte influência na determinação do nível abstracção dos modelos.



Calibração da base de dados (latência e *throughput*).

Protocolos de Replicação Síncronos

- Optimistas vs Conservativos.
- Comportamento em área local (LAN) assim como em larga escala (WAN).
- Diferentes critérios de coerência (Strict Locking, Snapshot Isolation).
- Granularidade na detecção de conflitos.
- Degradação da resposta do sistema na presença de faltas: *jitter* no processamento e perdas de mensagens.

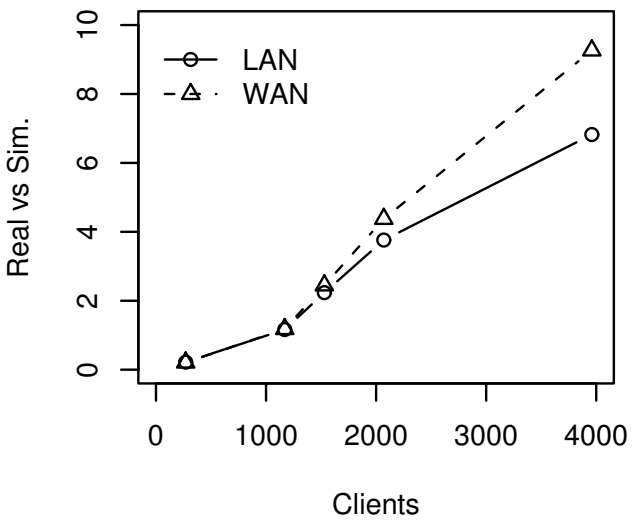
Desempenho

Protocol	LAN			WAN		
	TPM	LAT	ABS	TPM	LAT	ABS
CONS-G	Low	>> 10 ms	0%			
DBSM-G	High	≈10 ms	≈5%			
CONS-g	High	≈10 ms	0%	High	≈160 ms	0%
DBSM-g	High	≈10 ms	≈0%	High	≈160 ms	≈5%
PGR	High	≈10 ms	≈0%	High	> 200 ms	≈8%
CONS-SI	Low	>>10 ms	0%			
DBSM-SI	High	≈10 ms	≈0%	High	≈160 ms	≈3%

Síntese dos resultados.

Correcção da terminação do protocolo optimista na presença de faltas.

- Latência aumenta 10 a 20% quando mensagens são perdidas.
- O número de insucessos não é proporcional ao aumento da latência.
- O maior número de insucessos ocorre quando existem perdas aleatórias e granularidade fina.



Ratio entre tempo de simulação e tempo real.

Desempenho do Simulador

- Simular o sistema real em LAN leva menos tempo que o necessário para a execução real.
- No cenário WAN, apenas um ligeiro acréscimo de tempo seria necessário.

Conclusões

- Implementação de um ambiente de Simulação Centralizado.
- Biblioteca de modelos de simulação de bases de dados com calibração realista.
- Plataforma permite estudo/teste de protocolos de replicação de bases de dados.
- Custos dos testes reduzido.
- Controlo das variáveis envolvidas nos testes bastante flexível.

Trabalho Futuro

- Modelos de simulação de primitivas de controlo de concorrência: suporte para multi-threading.
- Concretizar algumas interfaces do mundo real no lado da simulação centralizada.
- Expôr uma interface de injecção de faltas mais amigável: suporte para redes com processos faltosos (P2P).

- Evaluating database replication in ESCADA (Extended Abstract). *In Proc. of the Workshop on Dependable Distributed Data Management (SRDS-WDDDM'04)*. 2004.
- Revisiting Epsilon Serializability to improve the Database State Machine (Position Paper). *In Proc. of the Workshop on Dependable Distributed Data Management (SRDS-WDDDM'04)*. 2004.
- Testing the dependability and performance of GCS-based database replication protocols. In *Proceedings of The International Conference on Dependable Systems and Networks (DSN'05)*. 2005.
- Group-based replication of on-line transaction processing servers. In *Dependable Computing: Second Latin-American Symposium (LADC'05)*. 2005.
- Experimental performability evaluation of middleware for large-scale distributed systems. In *7th International Workshop on Performability Modeling of Computer and Communications Systems (PMCCS'05)*. 2005.
- Revisiting 1-copy Equivalence in Clustered Databases. *Proceedings of the 2006 ACM Symposium on Applied Computing (ACM SAC'06)*. 2006.
- Evaluating Certification Protocols in the Partial Database State Machine. *Proceedings of ARES 2006 - The First International Conference on Availability, Reliability and Security (DILSOS'06)*. 2006.
- Group-Based Replication and 1-SR. Submitted to publication.
- A Pragmatic Protocol for Database Replication in Interconnected Clusters. Submitted to publication.
- A Flexible Replication Protocol for Demanding Transactional Workloads. Submitted to publication.