

Identificação de gargalos de disponibilidade e desempenho em sistemas de computação em nuvem

Rubens de Souza Matos Júnior
Orientador: Prof. Paulo Maciel

Agenda

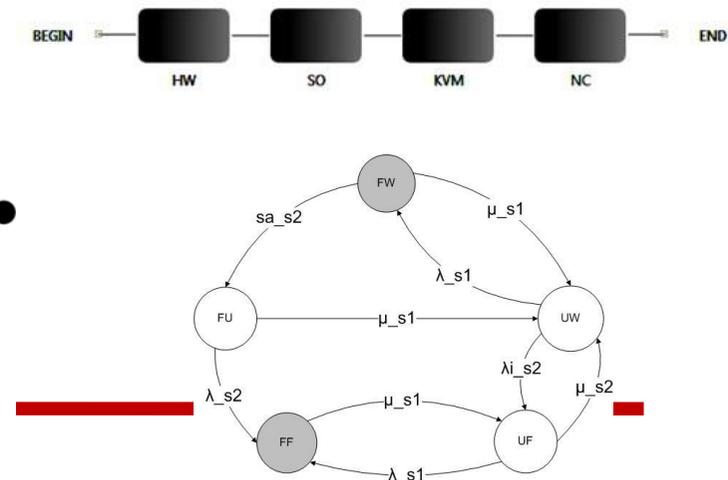
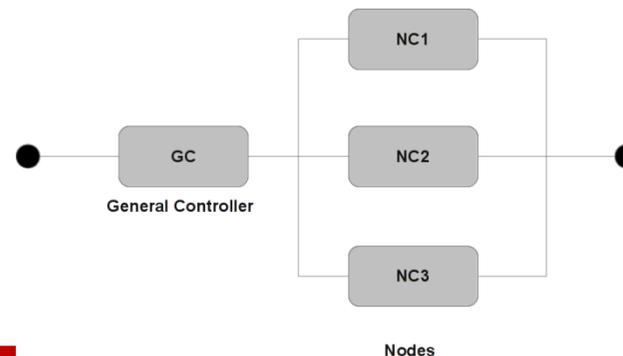
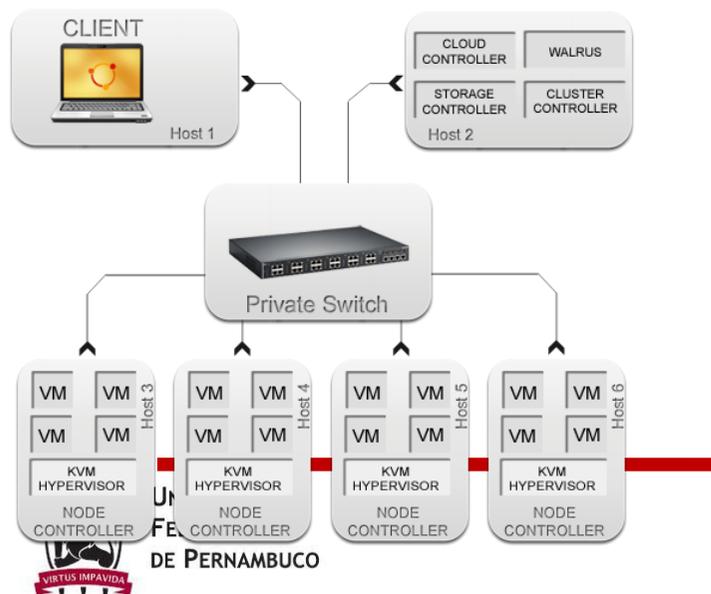


- Motivação
- Objetivos
- Metodologia proposta
- Estudos de caso
- Próximos passos



Motivação

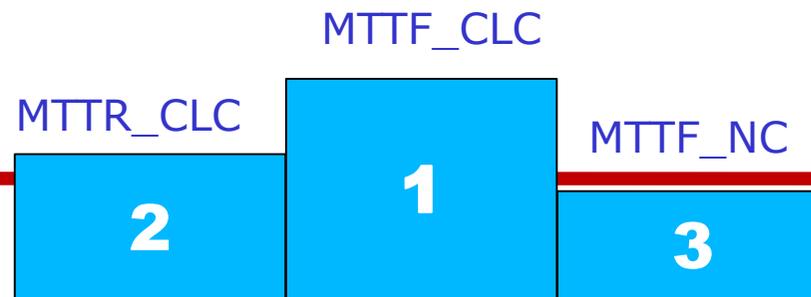
- Computação em nuvem é usada/almejada por várias empresas
- Como **propor melhorias** de desempenho/dependabilidade para sistemas complexos tais como as nuvens computacionais?
 - Muitos componentes de hardware e software
 - Modelos hierárquicos facilitam a descrição desses sistemas e o tratamento de largeness e stiffness.



Motivação



- Análise de sensibilidade:
 - Métodos para verificar o quanto as mudanças nos parâmetros de entrada irão afetar as saídas (resultados) de um sistema/modelo.
 - Técnica essencial para **detecção de “gargalos”** de desempenho/dependabilidade
 - Variação dos parâmetros, um por vez
 - Análise diferencial
 - Análise de correlação/regressão
 - Design of Experiments: Full-factorial, 2k-Factorial, ...



Objetivos específicos



- Principais objetivos:
 - Criação de **modelos** abrangendo **nível de aplicação e de infraestrutura** (Software executando em IaaS)
 - Prover **métodos e ferramentas** para **análise de sensibilidade** automatizada dos modelos hierárquicos
 - Incorporação de métodos de **A.S. em algoritmos de otimização** de infraestruturas/serviços de nuvem



Focos do trabalho



- Disponibilidade de infraestruturas de nuvem (IaaS)
 - Nuvens privadas (Eucalyptus)
 - Mobile cloud
- Desempenho/disponibilidade de aplicações executando na nuvem
 - Servidores web com mecanismos de recuperação de desastres
 - Composições de web services
- Desempenho de alocação de VMs (auto-scaling)

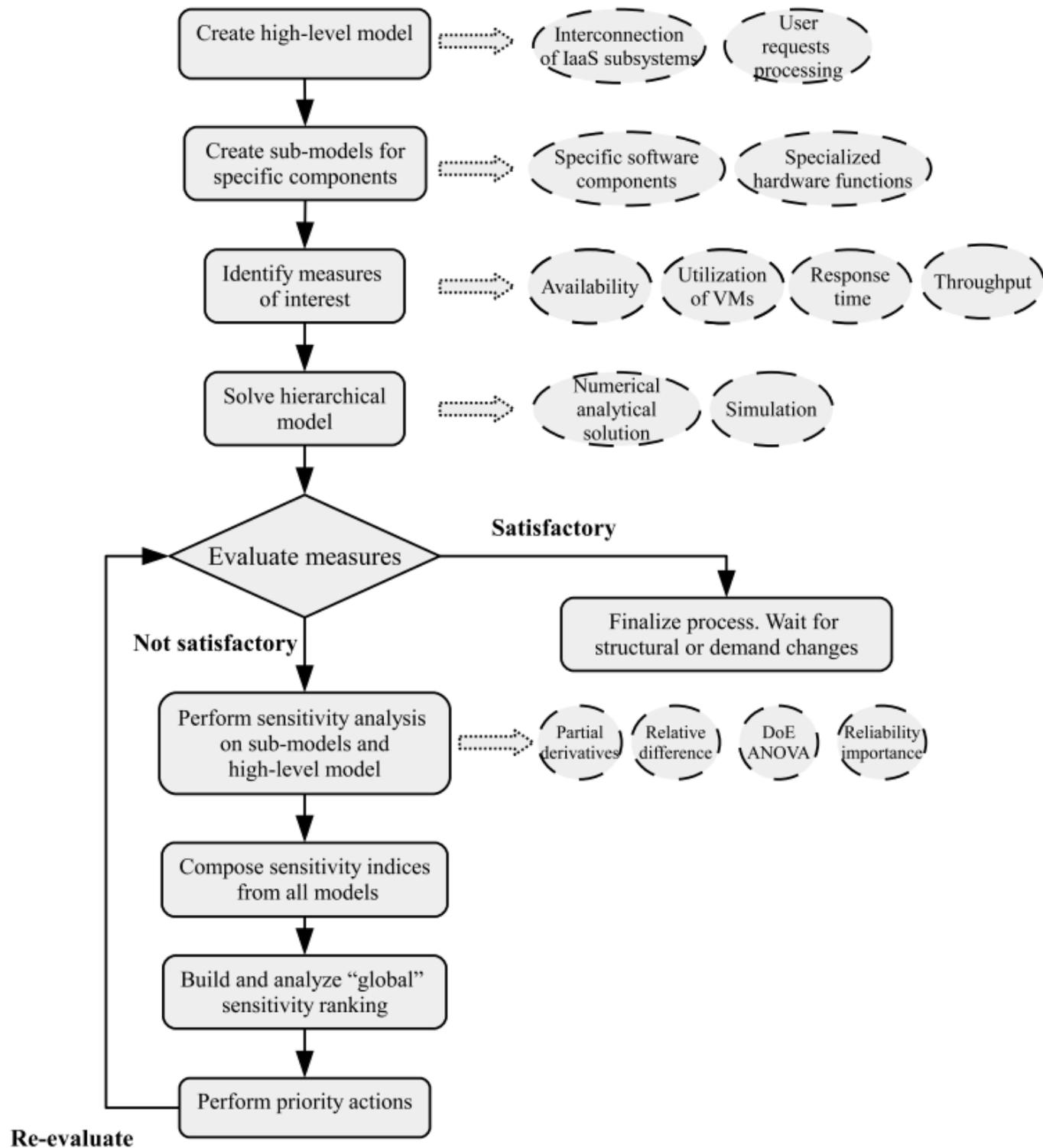


Metodologia proposta



1. Criação de **modelo de alto nível ou principal**
2. Criação de **sub-modelos** para componentes com comportamento **detalhado** mais conhecido (hosts, VMs, aplicações)
3. Identificação das **métricas** de interesse
4. **Resolução** do modelo hierárquico
5. Caso métricas **não** sejam **satisfatórias**:
6. Análise de **sensibilidade** do modelo **principal** e de **sub-modelos**
7. **Composição** dos índices de sensibilidade.
8. Construção e **análise** do **ranking geral**
9. Efetua **ações prioritárias** e reavalia métricas





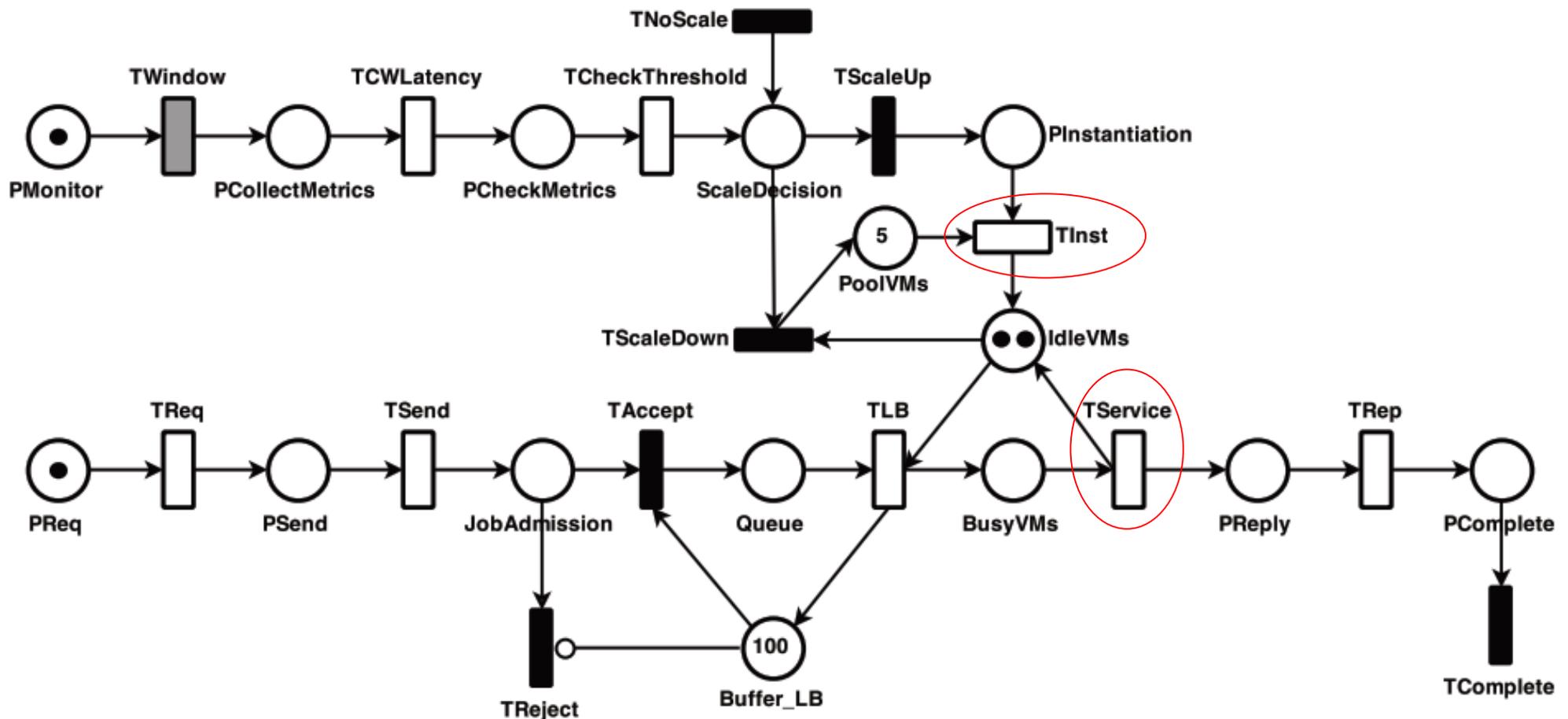
Web services em nuvem privada com autoscaling



- Composição de web services para recomendação de eventos musicais
- Criação e encerramento automático de VMs de acordo com a carga
- 3 modelos: 1 SPN + 2 CTMC:
 - Carga / autoscaling
 - Instanciação das VMs
 - Execução do web service (tempo de resp.)



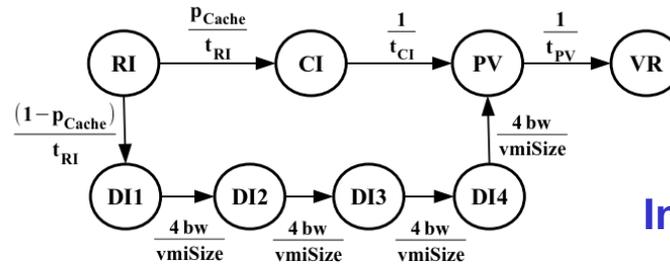
Web services em nuvem privada com autoscaling



Web services em nuvem privada com autoscaling

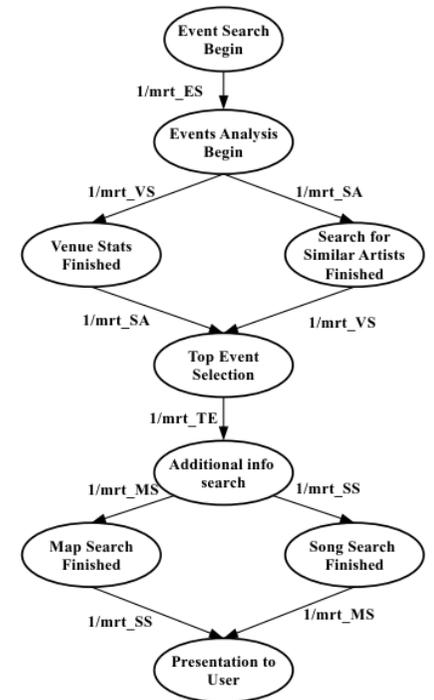
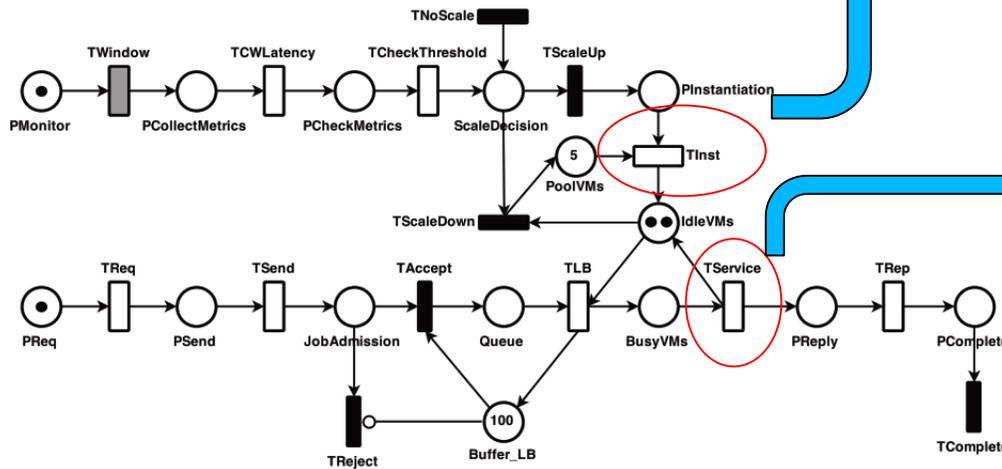


Colaboração com Eliomar e Artur

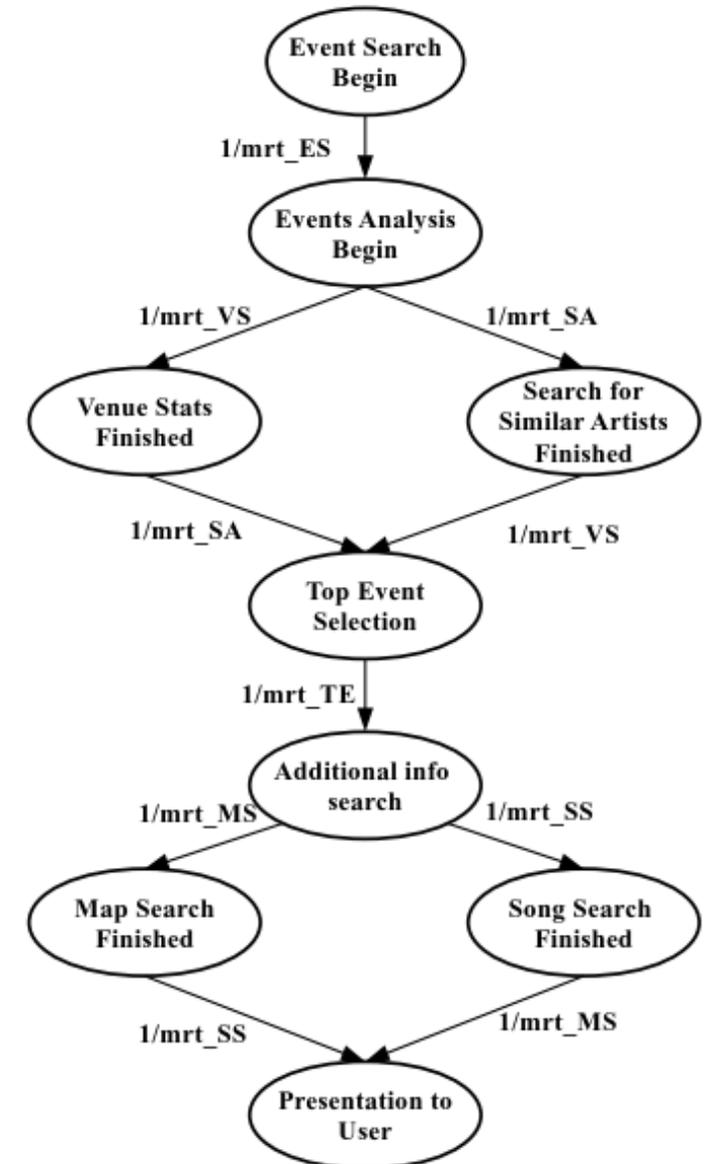
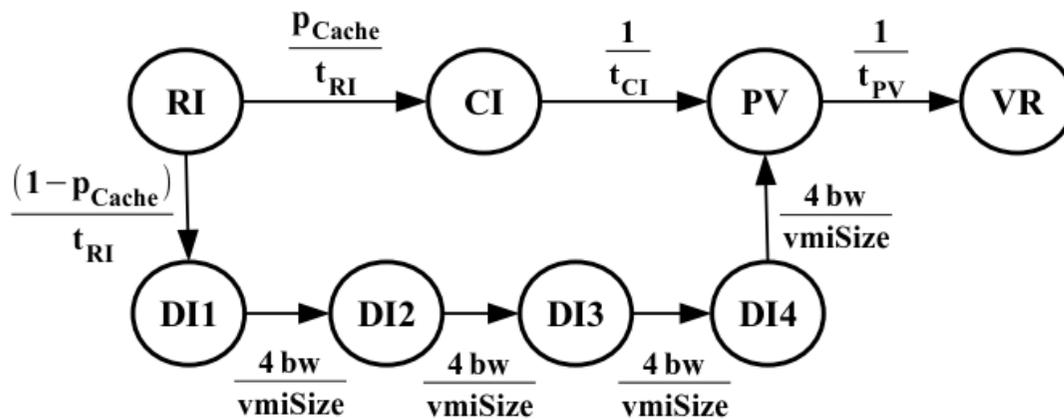


Instanciação de VMs

Representação global do sistema



Web services em nuvem privada com autoscaling



Web services em nuvem privada com autoscaling



Performance measures

Measure	Expression	Value
Utilization of VMs (%)	$E\{\#BusyVMs\} / (E\{\#IdleVMs\} + E\{\#BusyVMs\})$	38.3 %
Average number of busy VMs	$E\{\#BusyVMs\}$	1.716
Average number of idle VMs	$E\{\#IdleVMs\}$	2.773
LB queue size (#of requests)	$E\{\#Queue\}$	0.432
Mean response time - Rsp - (s)	$NRequests / (P\{\#PReply > 0\} \times (1 / TReply))$	9.029 s



Web services em nuvem privada com autoscaling



Sensitivity ranking for the main model

Parameter	S(Rsp)
TService	0.45763
TLB	0.13788
TRep	0.11303
TSend	0.11466
TReq	-0.05808
TWindow	0.00617
TCWLatency	0.00489
TInst	0.00256
TCheckThreshold	0.00176

$$S_{pCache}(Rsp) = S_{TInst}(Rsp) \times SS_{pCache}(TInst)$$

$$S_{mrtES}(Rsp) = S_{TService}(Rsp) \times SS_{mrtES}(TService)$$

Sensitivity ranking for the VM instantiation submodel

Parameter	SS(TInst)
pCache	-4.52843
vmiSize	0.52363
bw	-0.52363
t_PV	0.28465
t_CI	0.18421
t_RI	0.00752

Sensitivity ranking for the mashup sub-model

Parameter	SS(TService)
mrt_ES	0.33906
mrt_SA	0.32711
mrt_SS	0.26727
mrt_TS	0.03284
mrt_MS	0.02274
mrt_VS	0.01096

Web services em nuvem privada com autoscaling

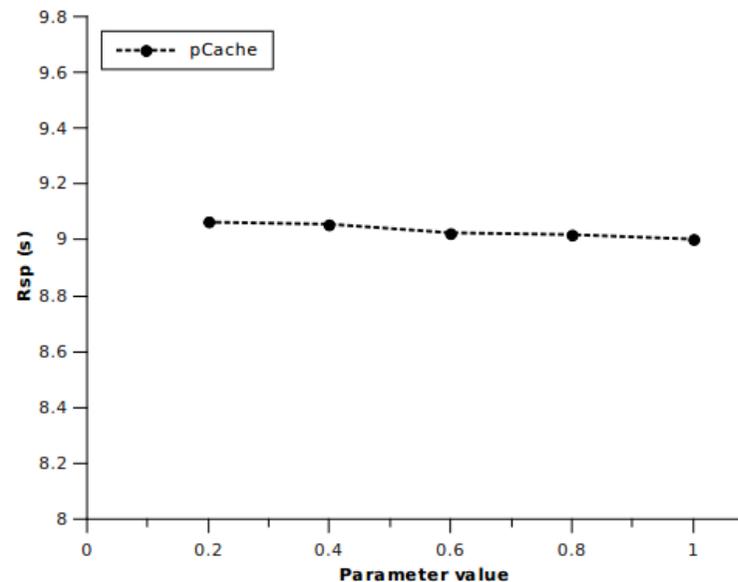
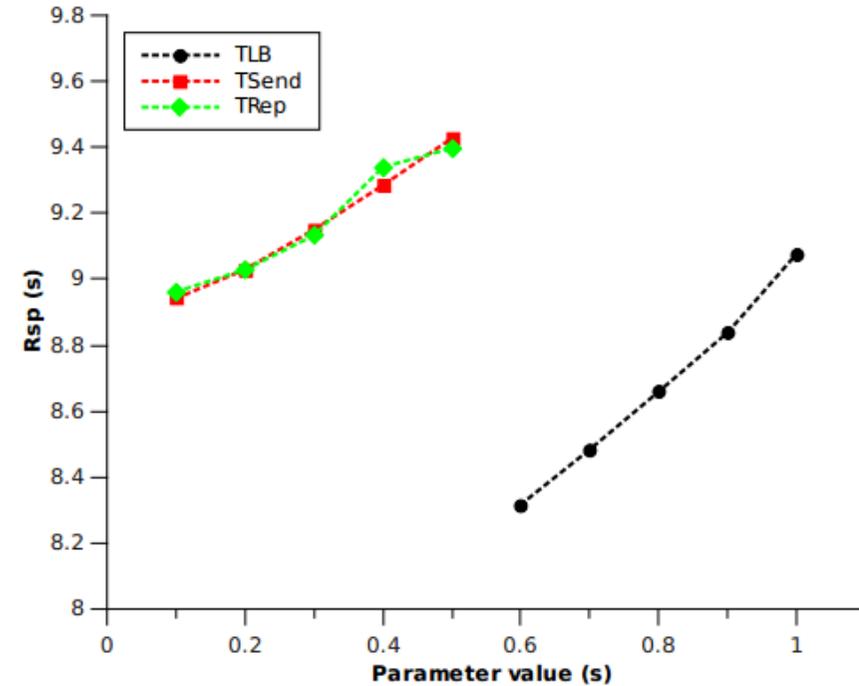
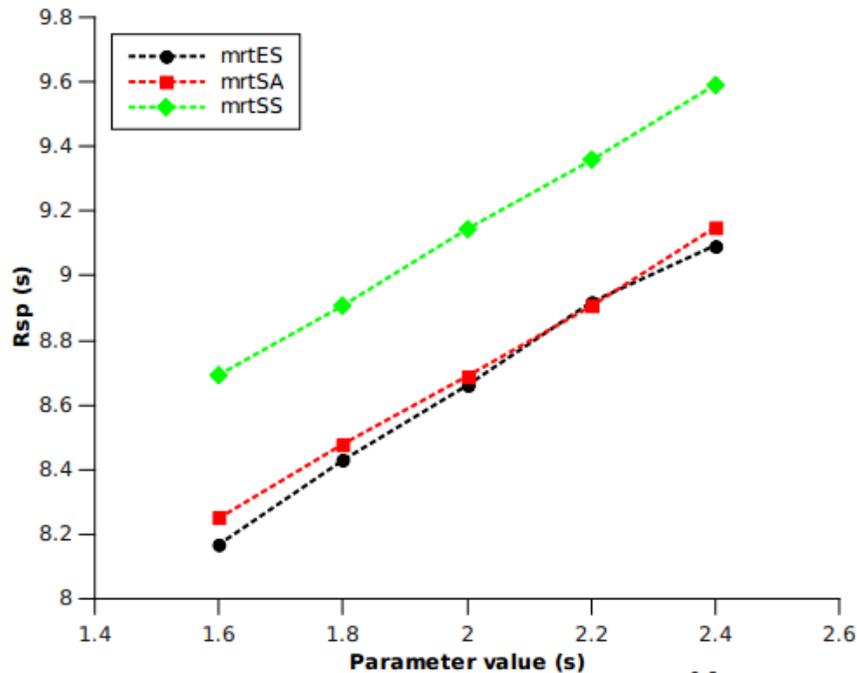


Unified sensitivity ranking for the general model and submodels

Parameter	S(Rsp)
mrt_ES	0.15517
mrt_SA	0.14969
TLB	0.13977
mrt_SS	0.12231
TSend	0.11466
TRep	0.11303
TReq	-0.05808
mrt_TS	0.01503
pCache	-0.01162
mrt_MS	0.01041
TWindow	0.00617
mrt_VS	0.00502
TCWLatency	0.00489
TCheckThreshold	0.00176
vmiSize	0.00134
bw	-0.00134
t_PV	0.00073
t_CI	0.00047
t_RI	0.00002

- O tempo de resposta dos seguintes web services:
 - **Event Search**
 - **Similar Artists**
 - **SongSearch**
- Tempo de execução do **Load Balancer**
- **Latência de rede** para enviar e receber
- O parâmetro mais importante do processo de instanciação (**pCache**) é apenas **intermediário** quando trata-se do tempo de resposta total da aplicação

Web services em nuvem privada com autoscaling



Passos em andamento



- Implementação das funções de A.S. para modelagem hierárquica em CTMC no Mercury
- Experimentos para validar os resultados do modelo e da A.S. do estudo de caso do mashup com autoscaling

