



Pós-Graduação em Ciência da Computação

**“AVALIAÇÃO DO IMPACTO DE UMA
POLÍTICA DE MANUTENÇÃO NA
PERFORMABILIDADE DE SISTEMAS DE
TRANSFERÊNCIA ELETRÔNICA DE FUNDOS”**

Por

Erica Teixeira Gomes de Sousa

Dissertação de Mestrado



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/~posgraduacao

RECIFE, AGOSTO/2009



Universidade Federal de Pernambuco

CENTRO DE INFORMÁTICA

PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Erica Teixeira Gomes de Sousa

**“Avaliação do Impacto de uma Política
de Manutenção na Performabilidade
de Sistemas de Transferência
Eletrônica de Fundos”**

Este trabalho foi apresentado à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

ORIENTADOR: Prof. Dr. Paulo Romero Martins Maciel

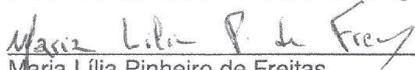
RECIFE, AGOSTO/2009



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Ata de Defesa de Dissertação de Mestrado do
Centro de Informática da Universidade Federal
de Pernambuco, 28 de agosto de 2009.

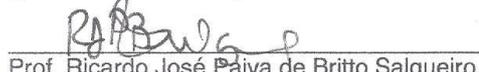
Ao vigésimo oitavo dia do mês de agosto do ano dois mil e nove, às dez horas, no Centro de Informática da Universidade Federal de Pernambuco, teve início a **octingentésima terceira** defesa de dissertação de mestrado em Ciência da Computação, intitulada "**Avaliação do Impacto da Política de Manutenção Preventiva na Performabilidade de Sistemas de Transferência Eletrônica de Fundos**" da candidata **Erica Teixeira Gomes de Sousa**, a qual já havia preenchido anteriormente as demais condições exigidas para a obtenção do grau de mestre. A Banca Examinadora, composta pelos professores Paulo Roberto Freire Cunha, pertencente ao Centro de Informática desta Universidade, Ricardo José Paiva de Britto Salgueiro, pertencente ao Departamento de Ciência da Computação e Estatística da Universidade Federal de Sergipe e Paulo Romero Martins Maciel, pertencente ao Centro de Informática desta Universidade, sendo o primeiro presidente da Banca Examinadora e o último orientador do trabalho de dissertação, resolveu: **Aprovar e dar o prazo de trinta dias para a entrega da versão final do trabalho**. E para constar lavrei a presente ata que vai por mim assinada e pela Banca Examinadora. Recife, 28 de agosto de 2009.



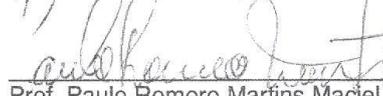
Maria Lília Pinheiro de Freitas
(secretária)



Prof. Paulo Roberto Freire Cunha
(primeiro examinador)



Prof. Ricardo José Paiva de Britto Salgueiro
(segundo examinador)



Prof. Paulo Romero Martins Maciel
(terceiro examinador)


 Maria do Socorro Chaves de Oliveira
Assistente em Administração
Pós-Graduação em
Ciência da Computação / UFPE

Confere Com o Original
Recife, 06, 10, 2009

A Deus.

À Minha Família.

Aos Meus Amigos.

Ao Prof. Dr. Paulo Romero Martins Maciel, orientador.

AGRADECIMENTOS

Gostaria de agradecer a todos que contribuíram para o desenvolvimento deste trabalho. Ao professor Paulo Maciel, pela orientação, apoio e enorme paciência, elementos essenciais para o desenvolvimento deste trabalho. Também gostaria de agradecer-lhe por todas as oportunidades de crescimento acadêmico e pessoal. Aos professores Paulo Cunha e Ricardo Salgueiro, por terem aceitado o convite para compor a banca de defesa do mestrado. A todos do grupo MoDCS (*Modeling of Distributed and Concurrent Systems*) em especial a Ângelo Ribeiro, Carlos Julian, Fábio Chicout e Gabriel Alves pela contribuição para a realização deste trabalho. Agradeço também à Cremilda Matos e Aig Issy por toda contribuição com esta pesquisa. Agradeço ao Laboratório de Performance Itautec-CIn e ao Laboratório HP-CIn pelo suporte financeiro e estrutura física cedida durante o desenvolvimento deste trabalho. Aos meus amigos do “Boa Nova”, principalmente Edinaldo, Elizabete, Eunice e Lizete (avós do coração), Jandete, João Cleber e Ubirajara pelo carinho. Aos meus amigos Ana Marta, Andrilene Maciel, Bil, Carlos Romerio (*In Memoriam*), Evandro Ventura, Flávia Valéria, Marcelo Marinho, Nívia Quental e Suilan Pim pelo carinho e apoio nos momentos mais difíceis desta jornada. Gostaria de agradecer à minha família, em especial à minha mãe, Elisabete, à minha avó, Elita, ao meu irmão Felipe e à meu padrao, Brivaldo, pelo incentivo, carinho e amor. Agradeço, principalmente, a Deus, que colocou todas essas pessoas em meu caminho.

*Se você pensar que pode ou que não pode, de qualquer forma, você
estará certo.*

—HENRY FORD

RESUMO

Os sistemas de transferência eletrônica de fundos (TEF) promovem benefícios financeiros para as instituições que utilizam esse serviço e conveniências para seus clientes. Os sistemas TEF realizam operações de crédito e débito em contas bancárias através de terminais de pontos de vendas (PDVs) em diferentes empresas. Assim, a avaliação de desempenho de sistemas TEF tem fundamental importância para as empresas que promovem o serviço de transferência eletrônica de fundos, uma vez que os recursos computacionais devem ser utilizados eficientemente, objetivando atingir os requisitos definidos nos acordos de nível de serviço (SLA). A ocorrência de eventos de falhas podem causar uma severa degradação no desempenho de sistemas. Assim, os resultados da modelagem de sistemas sem considerar os efeitos da dependabilidade sobre o desempenho são incompletos ou imprecisos. Este trabalho apresenta uma metodologia para avaliação da performabilidade composta de um método para avaliação de desempenho, um método para avaliação da dependabilidade e uma técnica de decomposição e composição. Este trabalho também apresenta dois modelos estocásticos expolinomiais para avaliação de performabilidade os quais são o modelo de desempenho de sistemas TEF e o modelo de dependabilidade de sistemas TEF. O modelo de desempenho de sistemas TEF proporciona a avaliação do desempenho dos recursos de processamento e armazenamento do servidor TEF, considerando variações na carga de trabalho (transações comerciais de crédito e débito). Já o modelo de dependabilidade de sistemas TEF promove a avaliação dos efeitos da ocorrência de eventos de falhas e de uma política de manutenção na disponibilidade e confiabilidade da infraestrutura de sistemas TEF. Os modelos de desempenho e de dependabilidade são combinados para a avaliação do impacto da dependabilidade no desempenho de sistemas TEF. A política de manutenção é empregada para prevenir a ocorrência de eventos de falhas no sistema TEF e permitir a recuperação de falhas para assegurar as expectativas de desempenho dos clientes através dos ajustes dos intervalos entre manutenções preventivas, dos intervalos entre substituições dos equipamentos, do número máximo de manutenções preventivas antes da substituição dos equipamentos, do tempo médio para reparo, do dimensionamento da equipe de manutenção e dos custos relacionados às manutenções. Estudos de caso são apresentados para análise do impacto das variações da carga de trabalho no desempenho de sistemas TEF, para análise do impacto de uma política de manutenção na performabilidade de sistemas TEF e para análise dos efeitos das variações de temperatura na performabilidade de sistemas TEF.

Palavras-chave: Avaliação de Performabilidade, Política de Manutenção, Avaliação de Desempenho, Avaliação de Dependabilidade, Modelos Estocásticos Expolinomiais, Redes de Petri Estocásticas, Sistema de Transferência Eletrônica de Fundos

ABSTRACT

The transfer electronic funds systems (EFT) promote financial benefits to institutions that utilize this kind of service and conveniences for its clients. EFT systems realize credit and debit operations in bank accounts through points of sales terminals (POSs) in different companies. This way, analyzing the performance of the EFT systems has a fundamental importance to companies that promote the service of transferring electronic funds, where the computational resources should be efficiently used with the objective of meeting the requirements defined in the service level agreements (SLA). The occurrence of fault events can lead to a severe degradation in the efficiency of systems. This way, the modeling results of systems are incomplete or imprecise without considering the dependability effects on performance. This work presents a methodology for performability evaluation composed of a method for evaluating performance, one for evaluating dependability and a decomposition and composition technique. This work also presents two expolynomial stochastic models for performability evaluation, which are the performance and dependability model of EFT systems. The EFT system performance model provides the performance evaluation of processing and storage resources of the EFT server taking the workload (credit and debit commercial transactions) variations into consideration. The EFT system dependability model promotes the evaluation of the fault events effects and maintenance policy in the availability and reliability of the EFT systems infrastructure. The performance and dependability models are combined for evaluation the impact of dependability in the performance of EFT systems. The maintenance policy is employed for preventing fault events in EFT systems and permits its recuperation so as to ensure the performance expectations of clients through adjustment of the intervals between preventive maintenance, intervals between replacement of equipments, maximum number of preventive maintenance before replacement of equipments, mean time to repair, maintenance team sizing and related costs. Case studies are presented in order to evaluate the impacts of workload variation on the performance of EFT systems, the impacts of maintenance policy on the performability of EFT systems and also to evaluate the effects of temperature variation on performability of EFT systems.

Keywords: Performability Evaluation, Maintenance Policy, Performance Evaluation, Dependability Evaluation, Expolynomial Stochastic Models, Stochastic Petri Nets, Electronic Funds Transfer System

SUMÁRIO

Lista de Figuras	xi
Lista de Tabelas	xiv
Lista de Abreviaturas	xvi
Capítulo 1—Introdução	2
1.1 Contexto	2
1.2 Motivação	3
1.3 Trabalhos Relacionados	4
1.4 Objetivos	6
1.5 Estrutura da Dissertação	6
Capítulo 2—Fundamentação Teórica	9
2.1 Sistema de Transferência Eletrônica de Fundos	9
2.2 Avaliação de Desempenho	12
2.3 Avaliação de Dependabilidade	15
2.3.1 Modelo de Arrhenius	17
2.4 Técnicas para Avaliação de Desempenho e Avaliação de Dependabilidade	18
2.4.1 Cadeias de Markov	18
2.4.1.1 Métodos de Solução para Probabilidade em Estado Estacionário	20
2.4.1.2 Métodos de Solução para Probabilidades em Estado Transiente	21
2.4.2 Diagrama de Blocos de Confiabilidade	21
2.4.3 Redes de Petri Estocásticas	23

2.4.4	Teoria das Filas	24
2.5	Avaliação de Performabilidade	26
2.6	Políticas de Manutenção	27
2.7	Considerações Finais	29
Capítulo 3—Redes de Petri		31
3.1	Introdução	31
3.1.1	Propriedades das Redes de Petri	33
3.1.1.1	Propriedades Comportamentais	33
3.1.1.2	Propriedades Estruturais	34
3.2	Rede de Petri Estocástica	36
3.3	Moment Matching	40
3.4	Considerações Finais	44
Capítulo 4—Metodologia para Avaliação de Performabilidade		46
4.1	Visão Geral	46
4.2	Método para Avaliação de Desempenho	46
4.3	Método para Avaliação de Dependabilidade	49
4.4	Estratégia de Decomposição e Composição	51
4.5	Considerações Finais	52
Capítulo 5—Modelos Propostos		54
5.1	Sistema TEF	54
5.2	Modelo de Desempenho de Sistemas TEF	55
5.3	Refinamento do Modelo de Desempenho de Sistemas TEF	57
5.4	Modelo de Dependabilidade de Sistemas TEF	64
5.4.1	Refinamento do Modelo de Dependabilidade de Sistemas TEF Con- siderando Variações de Temperatura	66
5.4.2	Refinamento do Modelo de Dependabilidade de Sistemas TEF Con- siderando uma Política de Manutenção	67
5.5	Avaliação de Performabilidade	71
5.6	Considerações Finais	71

Capítulo 6—Estudo de Caso	74
6.1 Estudo de Caso de Desempenho	74
6.2 Estudos de Caso de Dependabilidade	76
6.2.1 Estudos de Caso de Dependabilidade com Temperatura	79
6.3 Estudos de Caso de Performabilidade	84
6.3.1 Estudos de Caso de Performabilidade com Temperatura	86
6.4 Considerações Finais	87
Capítulo 7—Conclusões e Trabalhos Futuros	95
7.1 Contribuições	96
7.2 Trabalhos Futuros	97
Referências	102

LISTA DE FIGURAS

1.1	Número de Cartões	2
1.2	Número de Transações Comerciais	3
1.3	Ambiente TEF	3
2.1	Ponto de Venda	10
2.2	Sistema de Transferência Eletrônica de Fundos	10
2.3	Comunicação entre o Aplicativo de Automação Comercial e o Gerenciador Padrão	11
2.4	Curva da Banheira	16
2.5	Diagrama de Transição	19
2.6	Diagramas de Bloco de Confiabilidade	22
2.7	Diagrama de Blocos em Série e Paralelo 1	23
2.8	Diagrama de Blocos em Série e Paralelo 2	23
2.9	Sistema de Filas	24
3.1	Elementos de rede de Petri	32
3.2	Exemplo de rede de Petri	32
3.3	Períodos do Dia	32
3.4	Técnicas de Redução	36
3.5	Geração de Gráfico de Alcançabilidade	39
3.6	<i>Throughput Subnets</i>	40
3.7	Distribuição Empírica	42
3.8	Distribuição Erlang	43
3.9	Distribuição Hipoexponencial	43
3.10	Distribuição Hiperexponencial	44
4.1	Método para Avaliação de Desempenho	47

4.2	Fluxo de Medição de Desempenho	48
4.3	Método para Avaliação de Dependabilidade	50
4.4	Metodologia para Avaliação de Performabilidade	51
5.1	Sistema TEF	54
5.2	Modelo de Desempenho de Sistemas TEF	55
5.3	Sub-rede Cliente	56
5.4	Sub-rede <i>Buffer</i>	56
5.5	Sub-redes Processamento e Armazenamento de Transações	57
5.6	Fluxo da Transação TEF	58
5.7	Ambiente onde foram Realizados os Experimentos	59
5.8	Sub-rede Cliente Refinada	61
5.9	Sub-rede Buffer Refinada	61
5.10	Sub-redes de Processamento e Armazenamento de Transações Refinadas	62
5.11	Validação do Modelo de Desempenho de Sistemas TEF	63
5.12	Modelo de Dependabilidade de Sistemas TEF	64
5.13	Sub-rede Processamento de Transações	65
5.14	Sub-rede Armazenamento de Transações	65
5.15	Lugar Equipe	66
5.16	Sub-redes de Processamento e Armazenamento de Transações Refinadas	68
5.17	Modelo de Manutenção	70
5.18	Sub-rede Manutenção	71
6.1	Utilização do Processador e do Disco	75
6.2	Estudo de Caso 1	77
6.3	Estudo de Caso 2	77
6.4	Estudo de Caso 3	78
6.5	Confiabilidade X MTBM	78
6.6	Número de Noves X Custo de Manutenção	80
6.7	Estudo de Caso 4	82
6.8	Estudo de Caso 5	82
6.9	Estudo de Caso 6	82

6.10 Estudo de Caso 7	83
6.11 Número de Noves	83
6.12 Nível de Degradação	89
6.13 Nível de Degradação	90
6.14 Nível de Degradação	93

LISTA DE TABELAS

5.1	Média e Desvio Padrão	60
5.2	Parâmetros das Distribuições Hipoexponenciais	61
5.3	Métricas de Desempenho	64
5.4	MTTFs do Processador	66
5.5	MTTFs do Disco	66
5.6	Média e Desvio Padrão	67
5.7	Parâmetros das Distribuições Hipoexponenciais	67
5.8	Métricas de Dependabilidade	71
6.1	Utilização dos Recursos	75
6.2	Estudo de Caso 1	76
6.3	Estudo de Caso 2	76
6.4	Estudo de Caso 3	76
6.5	Disponibilidade	79
6.6	Confiabilidade	79
6.7	Custo Anual da Manutenção	79
6.8	Estudo de Caso 4	81
6.9	Estudo de Caso 5	81
6.10	Estudo de Caso 6	81
6.11	Estudo de Caso 7	81
6.12	Disponibilidade	83
6.13	Estudo de Caso 4	84
6.14	Estudo de Caso 8	84
6.15	Utilização do Processador para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Disponibilidade	85

6.16	Utilização do Disco para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Disponibilidade	86
6.17	Utilização do Processador para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Confiabilidade	87
6.18	Utilização do Disco para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Confiabilidade	88
6.19	Utilização do Processador para os Estudos de Caso 4, 5, 6 e 7 Considerando a Métrica de Disponibilidade	91
6.20	Utilização do Disco para os Estudos de Caso 4, 5, 6 e 7 Considerando a Métrica de Disponibilidade	92

LISTA DE ABREVIATURAS

ABECS - Associação Brasileira das Empresas de Cartões de Crédito e Serviços.

ACH - Câmara de Compensação Bancária Automatizada.

CTMC - *Continuous-Time Markov Chains.*

DTMC - *Discrete-Time Markov Chains.*

EFT - *Electronic Funds Systems.*

ESMP - *EFT System Maintenance Policy.*

FIFO - *First In, First Out.*

IBD - Interface de Banco de Dados.

INA - *Integrated Net Analyzer.*

ISO - *Standard for Financial Transaction Card Originated Messages - Interchange Message Specifications.*

LIFO - *Last In, First Out.*

LT - *Local Teams.*

MDT - *Mean Downtime.*

MPT - *Maintenance Planning Team.*

MTBF - *Mean Time Between Failures.*

MTBM - *Mean Time Between Maintenance.*

MTBR - *Mean Time Between Replacement.*

MTFPR - *Mean Time to Failure Perception and Mean Time for Repairing.*

MTFPRA - *Mean Time to Failure Perception, Mean Time for Repairing and Mean Time for team Arrival to the site.*

MTTF - *Mean Time to Failure.*

MTTR - *Mean Time to Repair.*

ODBC - *Open Data Base Connectivity.*

PDV - *Pontos de Vendas.*

POS - *Points of Sales.*

PN - *Petri Nets.*

SCOPE - *Solução Completa para Pagamento Completo.*

SLA - *Service Level Agreement.*

SPN - *Stochastic Petri Nets.*

ST - *Specialist Teams.*

TEF - *Transferência Eletrônica de Fundos.*

TimeNET - *Timed Net Evaluation.*

Introdução

INTRODUÇÃO

Este capítulo apresenta uma breve introdução aos sistemas de transferência eletrônica de fundos, destacando-se aspectos de desempenho e dependabilidade a serem avaliados nesse contexto. Em seguida, são apresentadas as motivações, os trabalhos relacionados, os principais objetivos e contribuições, além da estrutura da dissertação.

1.1 CONTEXTO

A Associação Brasileira das Empresas de Cartões de Crédito e Serviços (ABECS) divulgou que o mercado brasileiro de cartões vem demonstrando um expressivo crescimento nos últimos anos, conforme mostra a Figura 1.1, no período de 2000 a 2008 [ABE09]. Esse crescimento é devido à praticidade e à segurança que as movimentações de sistemas TEF (Transferência Eletrônica de Fundos) oferecem. Essas características impulsionam a aplicação dos sistemas TEF nas mais variadas áreas do comércio.

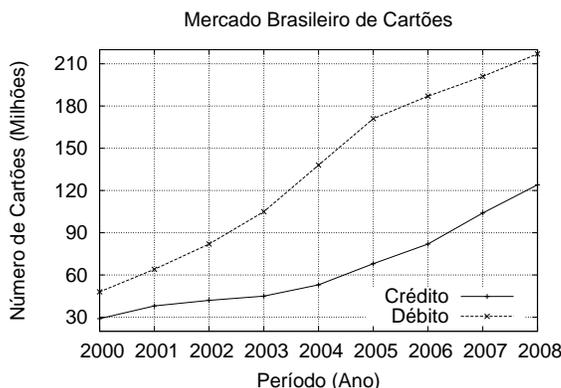


Figura 1.1: Número de Cartões

Em virtude disso, as movimentações de transações eletrônicas de fundos cresceram proporcionalmente. A Figura 1.2 mostra o crescimento do número de transações comerciais no período de 2000 a 2008 [ABE09].

O sistema de transferência eletrônica de fundos permite que clientes efetuem pagamentos a estabelecimentos comerciais por meio de troca de mensagens eletrônicas entre os pontos de vendas (PDVs) distribuídos em diversas empresas ou filiais e empresas autorizadas.

O ambiente TEF típico é composto de aplicativos de automação comercial, de um gerenciador padrão e de módulos TEF de empresas autorizadas. Os aplicativos de au-

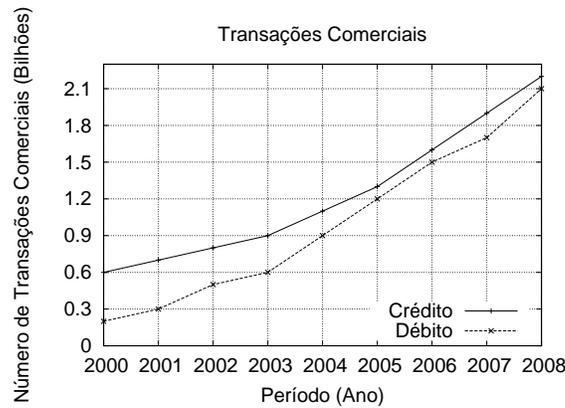


Figura 1.2: Número de Transações Comerciais

tomação comercial configurados em pontos de vendas são responsáveis pela impressão do cupom fiscal e do comprovante da transação TEF. Esses aplicativos proporcionam ao cliente a aquisição de um produto ou serviço pertencentes a um estabelecimento comercial. Cartões magnéticos e senhas, fornecidos pelas empresas autorizadas, permitem o acesso aos serviços e à identificação. O gerenciador padrão, ambiente de estudo, é o módulo responsável pelo direcionamento das transações comerciais para os Módulos TEF das empresas autorizadas e efetua o tratamento das atividades, permitindo a interação entre os aplicativos de automação comercial e as empresas administradoras de cartões de crédito e de débito. Finalmente, os módulos TEF das empresas autorizadas são responsáveis pela autorização de cada tipo de transferência eletrônica de fundos das transações comerciais efetuadas entre os clientes e os estabelecimentos comerciais [Mun09]. A Figura 1.3 descreve um ambiente TEF típico.

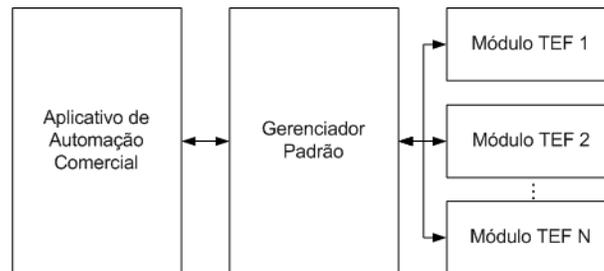


Figura 1.3: Ambiente TEF

1.2 MOTIVAÇÃO

O crescimento da capacidade computacional e a integração dos serviços de pagamento, assim como os avanços das novas tecnologias, têm proporcionado a crescente complexidade das transações eletrônicas. As empresas que promovem o serviço de transferência eletrônica de fundos (TEF) devem fornecer um serviço que atenda às expectativas de desempenho dos clientes.

A expansão do mercado de transferência eletrônica de fundos exige confiabilidade, alta disponibilidade, escalabilidade, segurança e preços acessíveis por parte das empresas que promovem esse serviço. A avaliação de desempenho e dependabilidade também são atividades essenciais para promover a melhoria da qualidade do serviço prestado e para o planejamento da infraestrutura do sistema TEF. Essa avaliação tem o objetivo de melhorar o desempenho de todo o sistema e reduzir os custos do serviço prestado. Assim, o emprego de técnicas de modelagem pode representar aspectos de desempenho e dependabilidade de sistemas com diferentes níveis de abstração, e cada nível é mais adequado para a resolução de um problema específico [Jai91, MA05].

Em sistemas TEF, eventos de falhas e atividades de reparo de um componente específico podem não necessariamente resultar na interrupção do serviço prestado pelo sistema, mas afetam o desempenho dele. A prestação ininterrupta do serviço promovida pelo sistema TEF pode resultar em uma degradação contínua no desempenho desse sistema. Nesse contexto, atividades de manutenção preventiva desempenham um importante papel para manter os níveis de desempenho e disponibilidade necessários para garantir a qualidade do serviço prestado. Assim, estratégias de manutenção têm um impacto fundamental sobre a disponibilidade e a confiabilidade de sistemas, bem como na prevenção de eventos de falhas e no custo do serviço oferecido. Dessa forma, a avaliação do desempenho desses sistemas, considerando defeitos, falhas [LAK92] e estratégias de manutenção é um assunto de maior importância para alcançar os acordos de nível de serviço contratados (SLA) [BM03, HMS02].

A modelagem integrada de aspectos de desempenho e dependabilidade é conhecida como modelagem de performabilidade. Essa modelagem permite a avaliação do desempenho considerando os níveis de degradação de serviços provocados pela ocorrência de eventos de falhas em um determinado período de tempo [HMRT01]. Aspectos de desempenho e dependabilidade são frequentemente modelados separadamente, com base no pressuposto de que falhas em componentes individuais ou sub-sistemas não afetam necessariamente o desempenho do sistema. Esse pressuposto não é válido para muitos sistemas que tenham mecanismos de recuperação ou tolerância a falhas. Esses mecanismos têm o objetivo de fornecer continuamente o serviço especificado mesmo que os níveis de desempenho sejam reduzidos. Ignorar os efeitos de eventos de falhas na degradação do sistema pode resultar em uma avaliação de desempenho incompleta ou inexata. Um modelo de performabilidade pode ser representado através de um modelo de dependabilidade, um modelo de desempenho e um método para combinar os resultados desses modelos [Rei90].

1.3 TRABALHOS RELACIONADOS

A combinação da avaliação de dependabilidade e desempenho é conhecida como avaliação de performabilidade, a qual considera a degradação do desempenho de sistemas provocada por aspectos de dependabilidade.

A maioria dos trabalhos que visa a avaliação de desempenho não trata da influência de aspectos de dependabilidade na degradação do desempenho de sistemas. O trabalho

de Araújo et al [ASM⁺09] propõe um modelo estocástico para avaliação de desempenho e planejamento de capacidade de sistemas TEF considerando um tráfego em rajada. No trabalho de Arteiro et al [ASRM07], apresenta-se um modelo de desempenho para sistemas de Middleware Orientados a Mensagem (MOMs) desenvolvido com redes de Petri estocásticas. Esse modelo permite a realização de experimentos de simulação visando à identificação da capacidade de entrega de mensagens e do ponto de saturação do sistema.

O trabalho de Silva et al [dSLJ⁺06] propõe uma metodologia baseada em redes de Petri estocásticas para modelar, avaliar e apoiar a escolha das composições de *WEB Service* considerando aspectos de desempenho. Souza et al [SARM06] apresentam um tratamento para modelagem de desempenho e avaliação de servidores de aplicação utilizando redes de Petri. Pacifici et al [PSST08] investigam o tráfego na WEB e a demanda por meio da análise de métricas de desempenho como vazão e utilização da CPU. Analisam a natureza do tráfego em relação aos diversos tipos de requisições e a correlação entre os fluxos de tráfego. A identificação da natureza do tráfego na WEB proporcionou a concepção e a implementação de um método on-line para estimação dinâmica da demanda da CPU.

Os trabalhos que objetivam a avaliação de desempenho considerando aspectos de dependabilidade não analisam o impacto de políticas de manutenção ou variações de temperatura. Shi [Shi04] desenvolve medidas de performabilidade e algoritmos para avaliação da performabilidade de redes sem fio nas quais uma política de controle de prioridade preemptiva é utilizada para garantir a qualidade do serviço de comunicação. Sesmun e Turner [ST00] propõem uma técnica que utiliza a performabilidade na concepção de redes de comunicação, cujo objetivo é desenvolver uma metodologia para concepção de redes de comunicação tolerantes a falhas. Os resultados mostram que a utilização da performabilidade para apoiar a decisão na topologia gera uma rede de comunicação mais confiável.

Nagaraja et al [NGB⁺05] propõem uma metodologia para avaliação de performabilidade (desempenho e disponibilidade) de serviços de internet baseados em *cluster*. Lanus et al [LYT⁺03] propõem a análise da performabilidade de sistemas complexos semelhantes a sistemas de telecomunicações por meio de sub-modelos. Essa proposta particiona o modelo gerado com base no sistema de telecomunicação em uma hierarquia de sub-modelos, evitando a explosão de espaço de estados causada pela complexidade e tamanho do modelo gerado.

Rabah [RK03] et al propõem modelos para avaliação de medidas de performabilidade em sistemas multiprocessadores. GosevaPopstojanova e Trivedi [GPT00] debatem a respeito de modelos analíticos de sistemas de computadores complexos. Conceitos básicos de modelagem de desempenho, dependabilidade e performabilidade são introduzidos através de um sistema multiprocessador.

Gupta et al [GLR⁺03] apresentam uma comparação dos pontos fortes e limitações de diferentes arquiteturas de servidores e utiliza modelos estocásticos para quantificação dos custos e dos benefícios de cada arquitetura, considerando aspectos de desempenho e dependabilidade. O trabalho de Mahevas et al [MRI01] propõe um novo método para análise da limitação de modelos concebidos para obtenção de medidas de desempenho,

dependabilidade e performabilidade de sistemas complexos. Essa abordagem permite a análise da limitação de modelos com grande espaço de estados na obtenção de medidas de desempenho, dependabilidade e performabilidade.

1.4 OBJETIVOS

Este trabalho propõe uma política de manutenção para sistemas de transferência eletrônica de fundos e avalia o impacto desta política na degradação do desempenho destes sistemas por meio de modelos estocásticos expolinomiais. Além disso, avalia o efeito de variações de temperatura na degradação de desempenho de sistemas TEF.

A avaliação da degradação do desempenho é realizada através de uma metodologia proposta. Essa metodologia produz um modelo de desempenho e um modelo de dependabilidade que são refinados por meio de distribuições expolinomiais [DAJ95], cuja utilização permite a avaliação de desempenho e dependabilidade com base em cadeias de Markov [BGdMT06, Mur89] bem como através de simulação estocástica.

A metodologia para a avaliação de performabilidade é baseada em um tratamento de modelagem hierárquica, que combina os resultados dos modelos de desempenho e dependabilidade. Esses modelos permitem o planejamento da capacidade de sistemas TEF, proporcionando o cumprimento dos acordos de níveis de serviços contratados (SLA). Esses modelos têm o objetivo de representar os recursos de armazenamento (disco físico) e o recurso de processamento [Bal01]. Os modelos de desempenho e dependabilidade devem ser capazes de fornecer métricas de desempenho tais como vazão e nível de utilização para a especificação da capacidade da infraestrutura de sistemas TEF e métricas de dependabilidade tais como disponibilidade e confiabilidade.

A metodologia adotada leva em consideração os efeitos da ocorrência de eventos de falhas e da política de manutenção na degradação do desempenho de sistemas TEF. A política de manutenção minimiza a ocorrência de defeitos e falhas através de ajustes nos intervalos entre manutenções, no número máximo de reparos antes da substituição do equipamento, no tempo médio de reparo, no tamanho das equipes de manutenção e nos custos relacionados.

1.5 ESTRUTURA DA DISSERTAÇÃO

O Capítulo 2 está dividido em quatro partes. A primeira parte apresenta os sistemas de transferência eletrônica de fundos, destacando as suas características e o seu funcionamento. A segunda apresenta conceitos sobre a avaliação de desempenho e a avaliação de dependabilidade. A terceira apresenta as principais técnicas para avaliação de desempenho e dependabilidade. A quarta apresenta conceitos sobre a avaliação de performabilidade e políticas de manutenção. O Capítulo 3 apresenta noções gerais sobre o formalismo matemático Redes de Petri (RP), a extensão Redes de Petri estocásticas (SPN) e a técnica de aproximação de fases. Essas técnicas são fundamentais para o desenvolvimento deste trabalho. O Capítulo 4 apresenta a metodologia proposta para avaliação de performabi-

lidade e suas atividades. O Capítulo 5 apresenta os modelos concebidos para avaliação de desempenho e dependabilidade de sistemas TEF e descreve seus componentes. Nele também é apresentada a validação do modelo de desempenho através da aplicação da metodologia de performabilidade de sistemas TEF. O Capítulo 6 apresenta estudos de caso baseados nos modelos propostos e adota a metodologia proposta para avaliação de performabilidade de sistemas TEF. O Capítulo 7 apresenta as conclusões obtidas durante o desenvolvimento desta dissertação, assim como as principais contribuições deste trabalho. Finalmente, são apresentados trabalhos futuros que darão continuidade ao estudo desenvolvido nesta dissertação.

Fundamentação Teórica

FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma introdução sobre os sistemas de transferência eletrônica de fundos. Em seguida, apresenta os conceitos básicos sobre desempenho e dependabilidade e as principais técnicas para avaliação. Posteriormente, traz os conceitos básicos de avaliação de performabilidade e, por último, conceitos sobre políticas de manutenção.

2.1 SISTEMA DE TRANSFERÊNCIA ELETRÔNICA DE FUNDOS

Os brasileiros estão cada vez mais preferindo os meios eletrônicos para pagar as despesas. Os cartões de crédito e débito ultrapassaram os cheques como meio de pagamento no Brasil. Em 2008, ocorreram 2,1 bilhões de transações comerciais com cartões de débito e 2,2 bilhões de transações comerciais com cartões de crédito, contra 1,94 bilhão de transações com cheques [ABE09]. De 2005 a 2008, a utilização dos cartões de débito aumentou 84%, enquanto o uso de cheques caiu 23%. Em 2008, os cartões eletrônicos responderam por 53,8% dos pagamentos no varejo [BCB09].

Um pagamento eletrônico é qualquer tipo de pagamento que não utiliza dinheiro em espécie ou cheque de papel. Os métodos de pagamentos eletrônicos incluem cartões de crédito, cartões de débito e a rede de ACH (Câmara de Compensação Bancária Automatizada).

O sistema de transferência eletrônica de fundos (TEF) é um tipo importante de sistema de pagamento eletrônico dos setores bancário e varejista, que emprega diversas tecnologias para captar e processar o dinheiro e as transferências de crédito entre os bancos e as empresas com seus clientes. Bancos, empresas de cartões de crédito e outras empresas podem manter serviços de pagamento por telefone e serviços de pagamento de contas pela Internet, que possibilitam aos clientes a utilização de seus telefones ou computadores em rede para pagar contas eletronicamente. Além disso, os terminais de pontos de vendas em lojas de varejo estão interconectados aos sistemas de transferência eletrônica e aos bancos. Isso possibilita a utilização de cartões de crédito ou de débito para o pagamento instantâneo de bens em lojas de varejo.

Os componentes de um sistema TEF são: o aplicativo da automação comercial, o gerenciador padrão e os módulos TEF. O primeiro é responsável pela impressão do cupom fiscal e do comprovante da transação TEF. Através dele o cliente adquire um produto ou serviço de um estabelecimento comercial. É necessário possuir um cartão magnético e uma senha, fornecidos pela empresa autorizadora, para acessar o serviço e se identificar junto à empresa autorizadora. A Figura 2.1 mostra um ponto de venda que é uma solução típica de aplicativos de automação comercial de sistemas TEF.



Figura 2.1: Ponto de Venda

O gerenciador padrão é responsável pelo direcionamento da transação comercial para os respectivos módulos TEF e efetua o tratamento das atividades, permitindo que o aplicativo de automação comercial interaja com as empresas de cartões de crédito e de débito.

Os módulos TEF autorizam a transferência eletrônica de fundos por meio das transações efetuadas entre o cliente e o estabelecimento comercial. A Figura 2.2 descreve o sistema de transferência eletrônica de fundos.

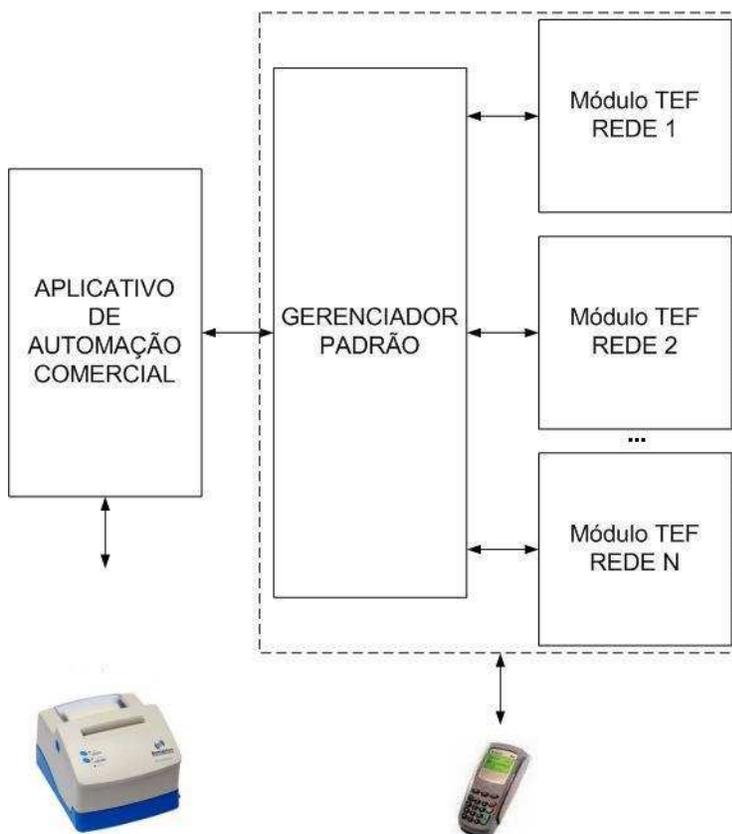


Figura 2.2: Sistema de Transferência Eletrônica de Fundos

Os sistemas TEF podem ser classificados como: dedicado, discado e IP [Mun09]. No sistema TEF dedicado, o ponto de venda fica conectado de modo permanente com a

empresa autorizadora através de uma linha privada. Esse sistema necessita de um modem da operadora responsável pelo link e um roteador. A vantagem nesta modalidade é que as transações são tratadas mais rapidamente, e também, pode existir apenas um servidor para tratar as diversas transações simultaneamente.

O sistema TEF discado caracteriza-se por utilizar uma linha telefônica discada como meio de comunicação para a realização das transações. A comunicação com as empresas autorizadoras faz-se a cada transação mediante acesso discado. Terminada a consulta e a aprovação do crédito, a conexão telefônica é desfeita; este ciclo completo leva aproximadamente 15 segundos. Desse modo, o processo torna-se mais demorado que no sistema TEF dedicado.

O sistema TEF IP tem as mesmas características do sistema TEF dedicado. A diferença é que sua conexão é estabelecida através de protocolos IP.

No processo de pagamento eletrônico através do sistema TEF, o aplicativo da automação comercial aciona o gerenciador padrão por meio da troca de pequenos arquivos contendo os parâmetros da transação, este, por sua vez, aciona o módulo TEF. A Figura 2.3 descreve um exemplo de comunicação entre o aplicativo de automação comercial e o gerenciador padrão [lin09].

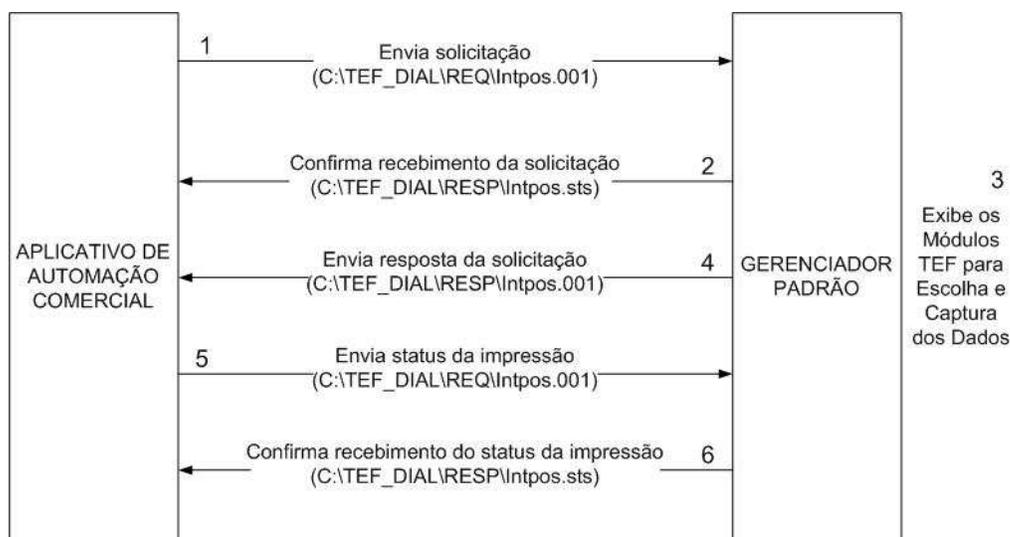


Figura 2.3: Comunicação entre o Aplicativo de Automação Comercial e o Gerenciador Padrão

A comunicação entre o aplicativo da automação comercial e o gerenciador padrão é realizada através de arquivos no formato “texto” com mensagens próprias. Nesse exemplo, os diretórios utilizados para a troca destes arquivos são: `C:\TEF_DIAL\REQ` e `C:\TEF_DIAL\RESP`. A comunicação entre o aplicativo de automação comercial e o gerenciador padrão está descrita abaixo.

1- Inicialmente, o aplicativo da automação comercial criará o arquivo *INTPOS.001* solicitando a realização da transação TEF e enviará para o diretório `C:\TEF_DIAL\REQ`. Este diretório é usado pelo aplicativo da automação comercial para envio de dados ao

gerenciador padrão.

Os dados de resposta do gerenciador padrão, após o processamento da transação pelo módulo TEF (American Express, Redecard ou Visanet), serão enviados para o diretório C:\TEF_DIAL\RESP. Este diretório é usado pelo aplicativo da automação comercial para receber as respostas do gerenciador padrão.

2- O aplicativo da automação comercial, após ter enviado o *INTPOS.001* para o gerenciador padrão C:\TEF_DIAL\REQ, deverá aguardar por alguns segundos o recebimento do arquivo *INTPOS.STS* enviado pelo gerenciador padrão C:\TEF_DIAL\RESP. Esse processo significa que o gerenciador padrão recebeu o *INTPOS.001* com a solicitação da transação TEF, enviado pelo aplicativo da automação comercial. Caso o gerenciador padrão não disponibilize este arquivo no tempo previsto, pode-se informar ao operador que houve algum problema.

3- Após o envio do *INTPOS.STS*, o gerenciador padrão irá exibir a tela com os módulos TEF disponíveis para a escolha.

4- Então, o aplicativo da automação comercial deverá aguardar o arquivo *INTPOS.001* com o resultado da transação. Este arquivo será gerado no diretório C:\TEF_DIAL\RESP.

5- Após o aplicativo da automação comercial realizar a impressão do comprovante TEF, deverá ser enviado um arquivo *INTPOS.001* ao gerenciador padrão C:\TEF_DIAL\REQ, confirmando ou não essa transação.

6- Após esse envio, o gerenciador padrão responderá um *INTPOS.STS* C:\TEF_DIAL\RESP, confirmando esse procedimento.

2.2 AVALIAÇÃO DE DESEMPENHO

A avaliação de desempenho de sistemas computacionais consiste de um conjunto de técnicas classificadas como as baseadas em medição e as baseadas em modelagem. As técnicas baseadas em modelagem podem ser classificadas como técnicas analíticas e técnicas baseadas em simulação [Lil00].

A medição de desempenho envolve essencialmente a monitoração do sistema enquanto está sob a ação de uma carga de trabalho. Para adquirir resultados representativos, a carga de trabalho deve ser cuidadosamente selecionada. Essa carga é utilizada nos estudos de desempenho, podendo ser real ou sintética. Embora a carga de trabalho real seja uma boa escolha por representar, de forma fiel, o sistema, ocasionalmente esta opção não é a desejável. Isso acontece quando o tamanho da carga não é considerável, e também quando esses dados receberam muitas perturbações ou, até mesmo, por questões de acessibilidade destes. Devido a esses motivos, outros tipos de cargas de trabalho também têm sido usados [Lil00]: Kernels, Programas sintéticos e Benchmarks.

A escolha da carga de trabalho é tão importante quanto a definição de qual estratégia de medição deve ser seguida. As diferentes estratégias de medição têm em sua base o conceito de evento, que é uma mudança no estado do sistema. A definição precisa de

um evento depende da métrica que está sendo medida. Por exemplo, um evento pode ser definido como um acesso ao disco, uma referência de memória, uma operação de comunicação de uma rede ou uma mudança interna de um processador. Esses diferentes tipos de métricas podem ser classificados em categorias baseadas no tipo de evento que compreende a métrica [Lil00]:

- Métricas baseadas em contagem de evento representam o registro da quantidade de vezes que um determinado evento ocorre. Exemplo: quantidade de requisições de leitura/escrita de um disco;
- Métricas baseadas em evento secundário representam o registro da quantidade de vezes que um evento ocorre devido à ocorrência de outro evento. Exemplo: para determinar a quantidade média de requisições enfileiradas em um *buffer*, será necessário registrar as requisições à medida que estas são adicionadas ou removidas do *buffer*. Assim, os eventos a serem monitorados serão as operações de enfileiramento e desenfileiramento, e a métrica será a quantidade média de requisições na fila;
- *Profiles* caracterizam o comportamento de um programa ou uma aplicação de um sistema. Usualmente é capaz de identificar detalhadamente em quais operações o programa ou sistema está consumindo mais tempo.

Diferentes estratégias de medição registram diferentes tipos de eventos:

- **Dirigida a evento:** é a estratégia que registra a frequência de ocorrência de um determinado evento. Uma vantagem dela é que a perturbação (*overhead*) gerada na medição ocorre apenas durante o registro do evento. Se o evento nunca ocorrer, ou apenas ocorrer raramente, a perturbação no sistema será relativamente pequena;
- **Tracing:** é uma estratégia similar à dirigida a evento, exceto que, em vez de simplesmente gravar os eventos que ocorrem, uma parte do estado do sistema é registrada para identificar o evento. Portanto é uma estratégia que requer mais armazenamento do que um simples contador de eventos;
- **Amostragem:** é uma estratégia que registra os dados do sistema em intervalos fixos de tempo, independentemente da ocorrência do evento. Como resultado, uma perturbação (*overhead*) pode ocorrer dependendo da frequência em que a medição é executada. Essa estratégia de medição produz um resumo estatístico do comportamento global do sistema. Eventos que ocorrem com pouca frequência são perdidos devido a esta aproximação estatística;
- **Indireta:** é uma estratégia que deve ser usada quando a métrica desejada não está acessível diretamente. Nesse caso, deve-se encontrar outra métrica que pode ser medida diretamente, a partir da qual se pode deduzir ou obter a métrica de desempenho desejada.

As ferramentas desenvolvidas para a avaliação de desempenho de sistemas de computadores modificam o comportamento do que está sendo medido. Quanto maior a quantidade de informações e resolução que a ferramenta de medição pode fornecer, maior será a perturbação introduzida por essa ferramenta. Essa perturbação introduzida pela ferramenta de medição torna os dados coletados por ela menos confiáveis. A ferramenta de medição dirigida a evento ocasiona uma perturbação (*overhead*) no sistema apenas quando os eventos ocorrem. A vantagem das ferramentas de medição dirigidas a evento é a pouca ou nenhuma perturbação ocasionada por eventos que ocorrem com pouca ou nenhuma frequência. A desvantagem dessa ferramenta de medição é a grande perturbação ocasionada por eventos que ocorrem frequentemente. A ferramenta de medição por amostragem ocasiona perturbações independente do número de vezes que o evento ocorre. A perturbação dependerá da frequência de amostragem determinada para coleta dos eventos [Lil00].

A qualidade das medições pode ser indicada através da precisão e exatidão da ferramenta de medição. A exatidão é a diferença absoluta entre o valor medido e o valor de referência. A precisão é a menor dispersão entre os valores obtidos através de múltiplas medições de uma determinada característica do sistema. Medições muito precisas são bem mais agrupadas em torno de um único valor medido. Medições imprecisas têm uma tendência a ter uma maior dispersão. A precisão dessas medidas é indicada pela dispersão delas em torno da média. A exatidão é a diferença entre o valor da média e o valor medido [Lil00].

A simulação é utilizada tanto em avaliação de desempenho, quanto na validação de modelos analíticos. Ao contrário das medições, as simulações baseiam-se em modelos abstratos do sistema, logo não exigem que o sistema esteja totalmente implantado para que sejam aplicadas. Assim, os modelos utilizados durante a simulação são elaborados através da abstração de características essenciais do sistema, sendo que a complexidade e o grau de abstração dele podem variar de um sistema para outro. Durante a simulação, controlam-se, com maior eficiência, os valores assumidos por parâmetros do sistema. Com isso, fica mais fácil obter informações relevantes para a avaliação de desempenho [Lil00].

A modelagem analítica utiliza um conjunto de equações e funções matemáticas para descrever o comportamento de um sistema. Os fatores que influenciam e interferem no comportamento do sistema são modelados e representados através dos parâmetros de equações matemáticas. Essas equações são chamadas de modelos analíticos. Apesar desses modelos considerarem parâmetros específicos de um sistema, podem ser facilmente adaptados para outros sistemas. Durante a construção dos modelos analíticos, deve-se levar em consideração a sua complexidade e praticidade. Os modelos analíticos permitem uma análise ampla e aprofundada em relação aos efeitos causados pelos parâmetros definidos nas equações sobre a aplicação. Além disso, também se podem estabelecer possíveis relacionamentos entre cada um dos parâmetros considerados. Essa modelagem, quando comparada às demais técnicas de avaliação de desempenho, apresenta menor custo de execução. Para validar os resultados alcançados através dos modelos elaborados, a modelagem analítica pode compará-los aos valores reais medidos em testes experimentais.

Esses valores poderão comprovar as predições realizadas através dos modelos analíticos [Lil00].

2.3 AVALIAÇÃO DE DEPENDABILIDADE

A avaliação de dependabilidade denota a capacidade que um sistema tem de oferecer um serviço de forma confiável. As medidas de dependabilidade são confiabilidade, disponibilidade, manutenibilidade, performabilidade, segurança, testabilidade, confidencialidade e integridade [LAK92]. Um resumo das principais medidas de dependabilidade é mostrado a seguir.

A confiabilidade de um sistema é a probabilidade de que este sistema execute a sua função, de modo satisfatório, sem a ocorrência de falhas, por um determinado período de tempo. A confiabilidade é representada pela Equação (2.1), onde T é uma variável aleatória que representa o tempo para ocorrência de falhas no sistema.

$$R(t) = P\{T \geq t\}, t \geq 0 \quad (2.1)$$

A probabilidade de ocorrência de falhas, ou inconfiabilidade, é representada pela Equação (2.2), onde T é uma variável aleatória que representa o tempo para ocorrência de falhas no sistema.

$$F(t) = 1 - R(t) = P\{T < t\} \quad (2.2)$$

A Equação (2.3) representa a confiabilidade, considerando a função de densidade $f(t)$ do tempo para ocorrência de falhas no sistema.

$$R(t) = P\{T \geq t\} = \int_t^{\infty} f(t)dt \quad (2.3)$$

O tempo médio para falha - *mean time to failure* - (MTTF) - é o tempo médio para a ocorrência de falhas no sistema. O MTTF é representado pela Equação (2.4).

$$MTTF = \int_0^{\infty} R(t)dt \quad (2.4)$$

A Figura 2.4 mostra a variação da taxa de falhas de componentes de hardware em função do tempo. A curva da banheira exhibe a taxa de falhas de componentes de hardware em três fases distintas [Ebe04]:

Durante a primeira fase, ocorre um curto período em que a taxa de falhas é bastante alta. Falhas ocorridas nesse período são decorrentes de defeitos de fabricação do equipamento. Com o intuito de encurtar esse período, fabricantes submetem os equipamentos a um processo chamado *burn-in*, onde eles são expostos a elevadas temperaturas de funcionamento.

Na segunda fase, as falhas ocorrem aleatoriamente. Valores de confiabilidade de equipamentos fornecidos por fabricantes aplicam-se a esse período.

Durante a fase final, a taxa de falhas cresce exponencialmente. O período de vida útil do equipamento normalmente não é uma constante. Ele depende do nível de estresse em que o equipamento é submetido durante esse período.

Em ambientes de alta disponibilidade, deve-se ter certeza de que a primeira fase tenha passado. Em alguns casos, é necessário deixar os equipamentos funcionando em um ambiente de testes durante um certo período. Ao mesmo tempo, deve-se tomar cuidado para que o equipamento seja substituído antes de entrar na fase final.

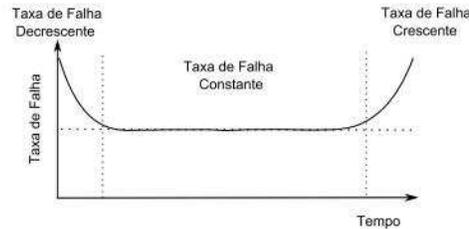


Figura 2.4: Curva da Banheira

A disponibilidade de um sistema é a probabilidade de que ele esteja operacional durante um determinado período de tempo, ou tenha sido restaurado após a ocorrência de um evento de falha. *uptime* é o período de tempo em que o sistema está operacional, *downtime* é o período de tempo em que o sistema não está operacional devido a ocorrência de um evento de falha ou atividade de reparo, e $uptime + downtime$ é o período de tempo de observação do sistema. A Equação (2.5) representa a disponibilidade de um sistema. A disponibilidade também pode ser expressa em termos de número de noves conforme a Equação (2.6). 100 representa o nível de disponibilidade máxima que o sistema pode atingir e A representa a disponibilidade do sistema.

$$A = \frac{uptime}{uptime + downtime} \quad (2.5)$$

$$N = 2 - \log(100 - A) \quad (2.6)$$

Mantenabilidade é a probabilidade de que um sistema seja reparado após a ocorrência de um evento de falha em um determinado período de tempo. A mantenabilidade é descrita pela Equação (2.7), onde T denota o tempo de reparo ou o tempo total de *downtime*. Essa equação representa a mantenabilidade, visto que o tempo de reparo T tem uma função de densidade $g(t)$.

$$V(t) = P\{T \leq t\} = \int_0^t g(t)dt \quad (2.7)$$

O tempo médio para reparo - *mean time to repair* (MTTR) - é o tempo médio em que o sistema está indisponível devido a atividades de manutenção. O MTTR é representado pela Equação (2.8).

$$MTTR = \int_0^{\infty} 1 - R(t)dt \quad (2.8)$$

O tempo médio de parada - *mean downtime* (MDT) - é o tempo médio em que o sistema está indisponível, qualquer que seja o motivo, como, por exemplo, atividades de manutenção corretivas, atividades de manutenção preventivas e indisponibilidade de recursos. O MDT é representado pela Equação (2.9)

$$MDT = 1 - A \quad (2.9)$$

O tempo médio entre falhas - *mean time between failures* (MTBF) - é o tempo médio entre as falhas do sistema, representado pela Equação (2.10).

$$MTBF = MTTR + MTTF \quad (2.10)$$

A performabilidade descreve a degradação do desempenho de sistemas provocada pela ocorrência de eventos de falhas; mesmo em decorrência deles, o sistema continuará funcionando, mas com degradações no nível de desempenho.

A dependabilidade de um sistema computacional pode ser afetada pela ocorrência de eventos de falhas, erros e defeitos, que são mecanismos destrutivos que tentam impedir o correto funcionamento de um sistema em função de uma sucessão de eventos indesejáveis. Estratégias de manutenção são estabelecidas para evitar a ocorrência de tais mecanismos, garantindo a qualidade do serviço oferecido. Dessa forma, estratégias de manutenção têm um impacto fundamental sobre a disponibilidade e a confiabilidade de sistemas [HMS02].

2.3.1 Modelo de Arrhenius

O modelo de Arrhenius relaciona o tempo de vida de um componente eletrônico com a temperatura a que ele é submetido [AR90]. A Equação (2.11) mostra a relação entre a temperatura e a o tempo médio para falhas (MTTF) do equipamento. Nessa equação, $MTTF_{temp_0}$ representa o tempo médio para falhas na temperatura de referência, $temp_0$ representa a temperatura de referência, $temp$ representa a temperatura analisada e $temp_r$ representa o aumento de temperatura.

$$MTTF(temp) = \frac{MTTF_{temp_0}}{2^{\frac{temp - temp_0}{temp_r}}} \quad (2.11)$$

2.4 TÉCNICAS PARA AVALIAÇÃO DE DESEMPENHO E AVALIAÇÃO DE DEPENDABILIDADE

Existem muitas técnicas usadas para avaliação de desempenho e dependabilidade. Esta seção apresenta os conceitos básicos de cadeias de Markov, diagrama de bloco de confiabilidade, redes de Petri estocásticas e teoria das filas [BGdMT06].

2.4.1 Cadeias de Markov

Um processo estocástico $X(t), t \in T$ é um conjunto de variáveis aleatórias definidas sobre o mesmo espaço de probabilidades, indexadas pelo parâmetro de tempo ($t \in T$) e assumindo valores no espaço de estados ($s_i \in S$) [CL08].

Os processos estocásticos podem ser classificados de acordo com o tipo de variável aleatória associada ao processo. Os processos estocásticos de tempo contínuo $X(t), t > 0$ são descritos por variáveis aleatórias de tempo contínuo, e seus espaços de estado podem ser discretos ou contínuos. Os processos estocásticos de tempo discreto $X(t), t = 0, 1, 2, \dots$ são descritos por variáveis aleatórias de tempo discreto e de forma similar seus espaços de estados podem ser discretos ou contínuos.

O processo estocástico é classificado como um processo de Markov se, para todo $t_0 < t_1 < \dots < t_n < t_{n+1}$ e para todo $X(t_0), X(t_1), X(t_2), \dots, X(t_n), X(t_{n+1})$, a distribuição condicional de $X(t_{n+1})$ depender somente do último valor anterior $X(t_n)$ e não dos valores anteriores $X(t_0), X(t_1), \dots, X(t_{n-1})$, isto é, para qualquer número real $X_0, X_1, X_2, \dots, X_n, X_{n+1}$, $P(X_{n+1} \leq s_{n+1} | X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = P(X_{n+1} \leq s_{n+1} | X_n = s_n)$ [BGdMT06].

Uma cadeia de Markov é descrita por uma sequência de variáveis aleatórias discretas, $X(t_n)$, em que t_n pode assumir um valor discreto ou contínuo, isto é, uma cadeia de Markov é um processo de Markov com um espaço de estados discretos.

A cadeia de Markov constitui um tipo particular de processo estocástico com estados discretos e com o parâmetro de tempo podendo assumir valores contínuos ou discretos. As cadeias de Markov de tempo contínuo são chamadas CTMC (*continuous-time Markov chains*) e as de tempo discreto são chamadas DTMC (*discrete-time Markov chains*). A propriedade markoviana (ausência de memória) destaca que eventos futuros não estão condicionados a eventos passados, ou seja, são dependentes apenas do estado presente [Jai91].

A cadeia de Markov possibilita a descrição do funcionamento de um sistema utilizando um conjunto de estados e transições entre esses estados. As transições entre os estados são modeladas por um processo estocástico de tempo contínuo ou discreto definidos por distribuições exponenciais ou geométricas.

O diagrama de transição é uma representação gráfica de uma cadeia de Markov. Nesse diagrama, são visualizados os estados (representados por círculos), as transições (representadas por arcos) e as probabilidades das transições. O diagrama de transição

representa o estado E_i e a probabilidade de transição p_{ij} , onde i e j são índices que identificam os vários estados possíveis (logo p_{ij} é a probabilidade de haver uma transição do estado E_i para o estado E_j). A Figura 2.5 mostra um diagrama de transição.

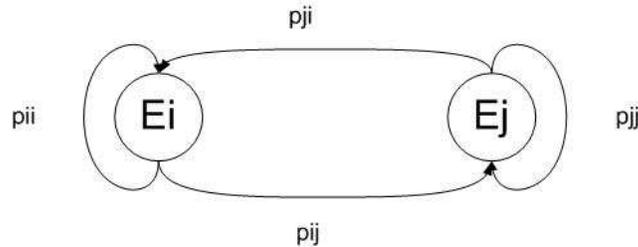


Figura 2.5: Diagrama de Transição

O vetor de probabilidade contém as probabilidades de transição de um estado para outros estados em um intervalo de tempo discreto. A Matriz (2.12) representa um vetor de probabilidade. Nessa matriz, p_{ii} indica a probabilidade de haver transição do estado E_i para o estado E_i , e p_{ij} indica a probabilidade de haver transição do estado E_i para o estado E_j . A soma dos elementos de um vetor de probabilidade sempre será igual a 1.

$$\pi = (p_{ii} \quad p_{ij}) \quad (2.12)$$

Para cada estado, deve haver um vetor de probabilidade. A matriz de transição de estados reúne os vetores de probabilidade de cada estado. Um modelo em cadeia de Markov é representado, matematicamente, por uma matriz de transição de estados (2.13). A probabilidade de cada estado em regime estacionário (solução de um modelo em cadeia de Markov) é a solução do sistema por meio da Equação linear (2.14). Nessa Equação, Q é a matriz de transição de estados e π (vetor de probabilidade) é o autovetor correspondente ao autovalor unitário da matriz de transição de estados.

$$Q = \begin{pmatrix} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{pmatrix} \quad (2.13)$$

$$\pi Q = 0 \quad (2.14)$$

Para os modelos em CTMC, a matriz de transição de estados é denominada de gerador infinitesimal, onde cada elemento não diagonal da linha i e coluna j da matriz representa a taxa de transição do estado i para o estado j do modelo. Os elementos diagonais representam o ajuste necessário para que a soma dos elementos de cada linha seja a zero. Para os modelos em DTMC, a matriz de transição de estados é denominada de matriz estocástica, onde cada elemento representa a probabilidade de transição entre os estados do modelo.

Em uma CTMC, as probabilidades de transição de estado são calculadas por meio da Equação (2.15).

$$p_{i,j}(s, t) = P\{X(t) = j | X(s) = i\} \quad (2.15)$$

O comportamento do processo markoviano é descrito pela Equação de probabilidade de transição de Champman-Kolmogoroff (2.16). Nessa equação, i, r e j são estados; s, u , e t , instantes de tempo [Her01].

$$p_{i,j}(s, t) = \sum_{\text{todor}} p_{i,r}(s, u) \cdot p_{r,j}(u, t), \quad s \leq u \leq t \quad (2.16)$$

Desenvolvendo-se essa equação (na notação matricial), obtém-se a Equação diferencial (2.17).

$$\frac{d\pi(t)}{dt} = \pi(t)Q \quad (2.17)$$

onde:

$\pi(t)$ é o vetor de probabilidade de estado i no tempo t .

$\pi(0)$ é o vetor de probabilidade de estado no tempo inicial.

Q é a matriz de transição de estados.

$\frac{d\pi(t)}{dt}$ é a primeira derivada do vetor de probabilidade com relação ao tempo.

A solução transiente, ou dependente do tempo, é importante quando o sistema a avaliar é dependente do tempo. Para modelos ergódicos [MLC96], considerando tempos de execução longos, pode-se mostrar que a probabilidade dos estados converge para valores constantes [Her01]. O comportamento transiente da cadeia de Markov nos fornece informações de desempenho e dependabilidade sobre os instantes iniciais do sistema. Assumindo-se que a probabilidade $\pi(t)$ é independente do tempo, isto é, $\pi_i = \lim_{t \rightarrow \infty} \pi_i(t)$ (homogeneidade), conseqüentemente, $\pi'(t) = 0$, resultando nas Equações (2.14) e (2.18).

$$\sum_{i=1}^N \pi_i = 1 \quad (2.18)$$

A Equação (2.18) é a condição de normalização, adicionada para assegurar que a solução obtida é um único vetor de probabilidade. A Equação (2.14) tem um conjunto de soluções infinitas. Normalizando as soluções, chega-se a um único vetor de probabilidades.

As cadeias de Markov têm importância fundamental para avaliações de desempenho e dependabilidade de sistemas computacionais [BGdMT06].

2.4.1.1 Métodos de Solução para Probabilidade em Estado Estacionário

Para computar as probabilidades em estado estacionário de uma CTMC finita, com N estados, faz-se necessário resolver o sistema de N equações lineares (Equações 2.14 e 2.18).

Assumindo-se que a cadeia de Markov é irredutível, e que π existe e é independente de $\pi(0)$, para se solucionar o sistema de equações, dois tipos de métodos de solução podem ser escolhidos: Métodos Diretos e Métodos Iterativos. Um método é chamado direto quando fornece a solução exata desejada, após um número finito de passos. Um método é chamado iterativo quando fornece uma sequência de soluções aproximadas que convergem para o valor exato.

A principal característica dos chamados métodos diretos é que eles focam na reescrita do sistema de equações de tal forma que se obtém expressões explícitas para o cálculo das probabilidades em estado estacionário. Os métodos diretos mais conhecidos são: a Decomposição LU [CL08] e a Eliminação Gaussiana [DAJ95]. Por razões computacionais e eficiência de memória, os métodos diretos não são utilizados quando o número de estados é maior que mil (1000). Nessas condições, utilizam-se métodos iterativos.

Os métodos iterativos são métodos numéricos para resolução de um sistema linear, que geram uma sequência de vetores $\{x^{(k)}\}$, a partir de uma aproximação inicial $x^{(0)}$. Sob certas condições, esta sequência converge para uma solução, caso ela exista. Os procedimentos iterativos não resultam em uma solução explícita do sistema de equações, não sendo possível estimar quantas interações serão necessárias para se obter a precisão desejada. Os métodos iterativos mais conhecidos são: Método da Potência, Método de Jacobi, Método de Gauss-Seidel e o Método SOR [CL08], [DAJ95].

2.4.1.2 Métodos de Solução para Probabilidades em Estado Transiente

As probabilidades de estado transiente de uma CTMC são especificadas por um sistema de equações lineares (Equação 2.19).

$$\pi'(t) = \pi(t)Q \quad (2.19)$$

As medidas são especificadas em termos de $\pi(t)$ e chamadas de medidas de instantes do tempo. Existem diversos métodos de resolução da Equação (2.19), para computação das probabilidades transientes.

A solução numérica de sistemas de equações diferenciais do tipo (2.19) tem sido, desde muito tempo, um tópico importante em matemática numérica. Muitos procedimentos numéricos têm sido desenvolvidos para esta finalidade. Entre eles, os mais utilizados são: Métodos de Runge-Kutta e Método de Uniformização [Gra91].

2.4.2 Diagrama de Blocos de Confiabilidade

O diagrama de blocos de confiabilidade é uma das técnicas mais usadas para a análise de confiabilidade de sistemas. Uma grande vantagem do diagrama de bloco de confiabilidade

é a facilidade de analisar a confiabilidade de sistemas [XDP04].

Em um modelo de diagrama de blocos de confiabilidade, os componentes são representados com blocos combinados com outros blocos (ou seja, componentes) em série, paralelo ou combinações dessas estruturas. Um diagrama que tem componentes conectados em série exige que cada componente esteja funcionando para que o sistema seja operacional. Um diagrama que tem componentes conectados em paralelo exige que apenas um componente esteja funcionando para que o sistema seja operacional [THGF96]. Assim, o sistema é descrito como um conjunto de blocos funcionais interconectados para representar o efeito da confiabilidade de cada bloco na confiabilidade do sistema [Smi05]. A Figura 2.6(a) mostra a conexão dos blocos em série, e a Figura 2.6(b) mostra a conexão dos blocos em paralelo.

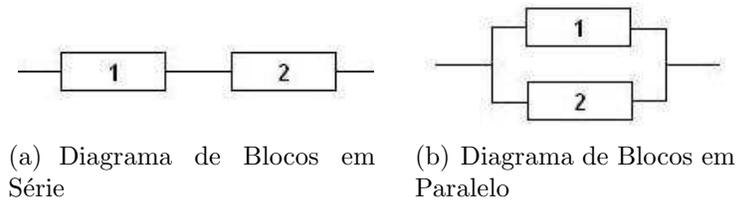


Figura 2.6: Diagramas de Bloco de Confiabilidade

A confiabilidade de dois blocos conectados em série é obtida através da Equação (2.20).

$$R_S = R_1 \times R_2 \quad (2.20)$$

onde:

R_1 descreve a confiabilidade do bloco 1.

R_2 descreve a confiabilidade do bloco 2.

A confiabilidade de dois blocos conectados em paralelo é obtida através da Equação (2.21).

$$R_P = 1 - \prod_{i=1}^2 (1 - R_i) \quad (2.21)$$

Blocos conectados em série e blocos conectados em paralelo podem ser combinados para formar um novo bloco, e a confiabilidade desse bloco é obtida através das Equações (2.20) e (2.21). A Figura 2.7 mostra a conexão de blocos em série e paralelo.

Os blocos 1 e 2 conectados em paralelo são combinados e formam o bloco $P1$ e sua confiabilidade é obtida por meio da Equação (2.22). Os blocos 4 e 5 são combinados e formam o bloco $P2$, cuja confiabilidade é obtida através da Equação (2.23). A Figura 2.8 mostra o resultado da combinação desses blocos.

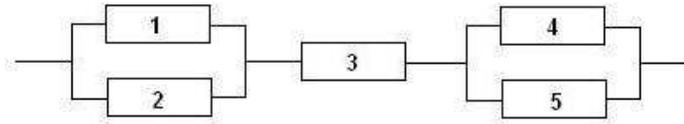


Figura 2.7: Diagrama de Blocos em Série e Paralelo 1

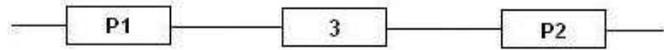


Figura 2.8: Diagrama de Blocos em Série e Paralelo 2

$$R_{P1} = 1 - \prod_{i=1}^2 (1 - R_i) \quad (2.22)$$

$$R_{P2} = 1 - \prod_{i=4}^5 (1 - R_i) \quad (2.23)$$

O diagrama de blocos de confiabilidade é utilizado, principalmente, em sistemas modulares que consistam de muitos módulos independentes, onde cada um pode ser facilmente representado por um bloco de confiabilidade.

2.4.3 Redes de Petri Estocásticas

Redes de Petri estocásticas (SPNs) é uma ferramenta gráfica cujo propósito é a abstração de um sistema real, um formalismo do fluxo de dados do sistema modelado em questão. O formalismo das Redes de Petri é utilizado em sistemas que possam apresentar atividades assíncronas, concorrentes e não-determinísticas, além de conflitos. A alocação de recursos, sistemas operacionais, redes de filas, entre outros, são exemplos de áreas de estudo em que se aplicam a modelagem de redes de Petri estocásticas.

As redes de Petri estocásticas permitem a modelagem e análise probabilística de sistemas. As transições em SPNs podem ser imediatas e temporizadas. As transições temporizadas possuem um atraso exponencialmente distribuído. A propriedade de ausência de memória da distribuição exponencial no atraso dos disparos, implica no fato das SPNs serem isomórficas às cadeias de Markov de tempo contínuo (continuous time Markov chain, CTMCs), provendo então medidas de desempenho e dependabilidade. Desta forma, as redes de Petri estocásticas têm importância fundamental para avaliações de desempenho e dependabilidade de sistemas computacionais [MLC96]. As redes de Petri estocásticas foram descritas no Capítulo 3.

2.4.4 Teoria das Filas

Um sistema de filas é composto de clientes que chegam para serem atendidos pelos servidores; uma fila de tamanho finito ou infinito, onde os clientes que chegam e não são atendidos imediatamente, esperam pelo atendimento; um ou mais servidores idênticos atendendo os clientes; e clientes saindo após o atendimento. Um servidor pode servir um cliente durante determinado período de tempo. Um servidor está em estado ocupado durante o tempo em que está servindo um cliente, e em estado ocioso em caso contrário. Se todos os servidores estiverem ocupados atendendo aos clientes, os clientes recém-chegados devem esperar na fila antes de serem atendidos. Quando acaba o atendimento do cliente, um dos clientes que estava esperando na fila será selecionado para o atendimento de acordo com a disciplina de filas [BGdMT06]. A Figura 2.9 representa um sistema de filas.

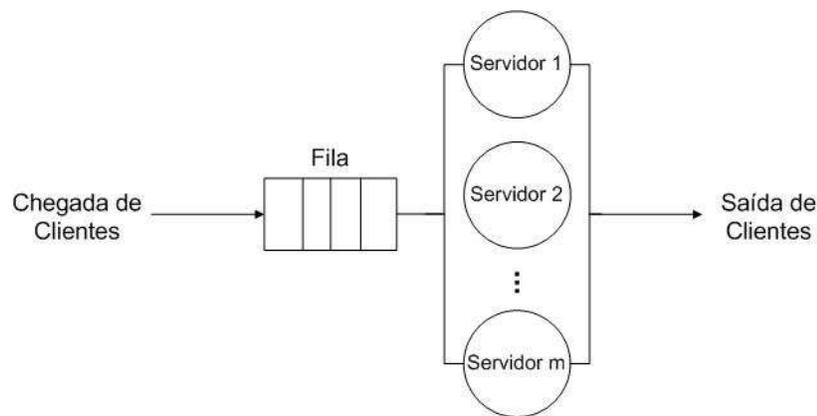


Figura 2.9: Sistema de Filas

As características básicas que descrevem um sistema de fila são: processo de chegada de clientes, processo de atendimento de clientes, a disciplina de fila, a capacidade do sistema, o número de servidores e o número de fases do serviço [Gro08].

O processo de chegada é geralmente um processo estocástico, sendo necessário conhecer a distribuição de probabilidade que descreve os tempos entre chegadas.

No processo de atendimento de clientes, o intervalo de tempo de atendimento de clientes é geralmente descrito por uma distribuição de probabilidade. Geralmente, os tempos de atendimento de clientes são considerados independentes do processo de chegada e do servidor que executa o serviço, e são identicamente distribuídos.

A disciplina de filas descreve a forma como os clientes são selecionados para serem atendidos quando estão esperando na fila. As disciplinas de filas consideram o atendimento em primeiro lugar de quem chegou primeiro na fila (FIFO - “*First In, First Out*”) ou o atendimento em primeiro lugar de quem chegou por último na fila (LIFO - “*Last In, First Out*”).

Em alguns sistemas existe uma limitação física no dimensionamento da fila. Quando

esta atinge um determinado tamanho, não é permitida a entrada de mais clientes até que haja espaço disponível, que será proveniente do atendimento de um cliente e, conseqüentemente, diminuição do tamanho da fila.

O número de servidores determina o número de clientes que podem ser atendidos em simultâneo. Um sistema de fila pode ter várias fases de serviço, cada cliente tem de passar por várias fases.

O número total de clientes no sistema no instante t ($N(t)$) é a soma do número de clientes na fila ($N_q(t)$) com o número de clientes em atendimento ($N_s(t)$). Então, seja $p_n(t) = P(N(t) = n)$, e $p_n = P(N = n)$ no estado estacionário. Considerando C servidores no estado estacionário, podem ser deduzidos o número médio de clientes no sistema (Equação 2.24) e o número de número de clientes esperando na fila (Equação 2.25).

$$L_s = E[N] = \sum_{n=0}^{\infty} np_n \quad (2.24)$$

$$L_q = E[N] = \sum_{n=c+1}^{\infty} (n - C)p_n \quad (2.25)$$

Uma relação que tem muita utilidade na teoria das filas foi desenvolvida por John D. C. Little. A fórmula de Little relaciona o comprimento médio da fila com o tempo médio de espera na fila.

O comprimento médio da fila L_q pode ser obtido por meio da fórmula de Little (Equação 2.26), onde λ representa a taxa de chegada de clientes ao sistema e W_q representa o tempo médio de espera na fila [Jai91].

$$L_q = \lambda W_q \quad (2.26)$$

O tempo médio de permanência no sistema W é a soma do tempo médio de espera na fila W_q com o tempo médio de atendimento W_s . O tempo médio de atendimento é o inverso da taxa de serviço μ . A Equação (2.27) representa o tempo médio de permanência no sistema.

$$W = W_q + W_s = W_q + \frac{1}{\mu} \quad (2.27)$$

O número médio de clientes no sistema é obtido por meio da Equação (2.28), onde W representa o tempo médio de permanência no sistema e λ representa a taxa de chegada dos clientes ao sistema.

$$L_s = \lambda W \quad (2.28)$$

Algumas medidas importantes de desempenho são utilização, vazão, comprimento médio da fila, tempo médio de resposta, tempo médio de espera na fila e tempo médio de serviço [BGdMT06].

A utilização ρ é a fração de tempo em que o servidor está ocupado atendendo o cliente. A Equação (2.29) representa a utilização de um servidor, e a Equação (2.30) representa a utilização de múltiplos servidores.

$$\rho = \frac{\lambda}{\mu} \quad (2.29)$$

$$\rho = \frac{\lambda}{m\mu} \quad (2.30)$$

A vazão representa o número médio de clientes atendidos pelo servidor por unidade de tempo. A Equação (2.31) descreve a vazão quando um servidor atende os clientes, e a Equação (2.32) representa a vazão quando vários servidores atendem os clientes.

$$\lambda = \rho\mu \quad (2.31)$$

$$\lambda = m\rho\mu \quad (2.32)$$

O tempo de médio de resposta representa o tempo médio de permanência no sistema W .

O tempo de médio de serviço representa o tempo médio de atendimento W_s .

São sistemas computacionais reais aqueles com diversos recursos, como processador, disco, memória e rede, e podem ser modelados por meio de redes de filas [BGdMT06].

2.5 AVALIAÇÃO DE PERFORMABILIDADE

A avaliação de performabilidade descreve o efeito de eventos de falhas e atividades de reparo na degradação do desempenho de sistemas. Para a avaliação de performabilidade é comum a utilização de técnicas de decomposição hierárquicas para combinação de um modelo de dependabilidade de alto nível e modelos de desempenho de baixo nível, um modelo de desempenho para cada estado do modelo de dependabilidade [PRS96] [STP96]. Este método hierárquico tem o objetivo de evitar os problemas *largeness* e *stiffness*.

Largeness é consequência do tamanho do espaço de estados do modelo e *stiffness* é consequência das diferentes ordens de magnitude entre os tempos das atividades de desempenho, os tempos dos eventos de falhas e os tempos das atividades de reparo. O problema *stiffness* pode causar sérios problemas durante a solução analítica do modelo,

mesmo que o modelo não tenha um grande espaço de estados devido aos diferentes tempos associados as transições temporizadas.

A integração da modelagem de aspectos de desempenho e dependabilidade de sistemas é conhecida como modelagem de performabilidade. A modelagem de performabilidade permite a avaliação de desempenho considerando a degradação dos níveis de serviço provocados pelos eventos de falhas durante um determinado período de tempo.

2.6 POLÍTICAS DE MANUTENÇÃO

As atividades de manutenção têm um papel primordial, uma vez que o constante funcionamento de máquinas e equipamentos, pelo processo produtivo, exerce grande influência sobre o seu estado, que se traduz, na maioria das vezes, por um processo de degradação. Nesse contexto, é de extrema importância a coordenação das atividades de manutenção às características particulares de cada sistema, bem como os objetivos que se desejam atingir, quais sejam: a minimização dos custos, dos tempos de parada, a maximização da confiabilidade ou da disponibilidade. A esse processo dá-se o nome de política de manutenção.

A manutenção preventiva é aquela planejada com o objetivo de prevenir falhas, ou seja, ações de manutenção são tomadas antes da ocorrência de eventos de falhas.

Os benefícios em longo prazo da manutenção preventiva são o aumento da confiabilidade do sistema; a diminuição no custo da substituição de equipamentos; a diminuição do tempo de parada; e o melhor gerenciamento do inventário de peças extras. Outro fator que influencia na decisão de se realizar uma manutenção preventiva é o custo, que deve ser menor que o de uma manutenção corretiva.

A manutenção corretiva é uma manutenção com o objetivo de corrigir falhas, ou seja, ações de manutenção são tomadas após a ocorrência de eventos de falhas. A manutenção corretiva pode ser planejada quando a falha não provoca a parada do sistema, ou não planejada quando a falha ocorre de maneira inesperada.

A manutenção preditiva é aquela na qual o sistema dispõe de meios para informar o *status* atual do sistema fornecendo, assim, uma estimativa de quando a manutenção deve ser realizada.

As ações de manutenção realizadas variam, além do tipo de manutenção, com a política de manutenção adotada.

Existem vários tipos de políticas de manutenção [WP06], entre eles a política de substituição por idade, a política de substituição em bloco, a de manutenção preventiva periódica, a política de limite de falhas, a política de manutenção preventiva sequencial, a política de custo de reparo limitado, a política de tempo de reparo limitado, a política de contagem do número de reparos e a política de tempo de referência.

De acordo com [Smi05], caso a taxa de falhas seja crescente e o custo de uma substituição não planejada (devido à falha) seja maior que o custo de uma substituição

planejada, é necessário saber o custo da substituição não planejada em um intervalo de tempo em particular e o custo de substituições planejadas durante esse intervalo.

O intervalo de substituição ótimo é aquele que minimiza a soma dos dois custos descritos. Esse intervalo pode ser encontrado de acordo com duas políticas: a substituição por idade e a substituição em bloco.

A política de substituição por idade consiste em substituir um item, se este chegar a um tempo t (idade de substituição) ou se falhar antes. O uso dessa política só é eficaz se o custo da substituição antes da ocorrência da falha proporcionar alguma economia. O principal objetivo da política de manutenção por idade é a garantia do nível de confiabilidade a um custo mínimo, através de sucessivas substituições que se antecedem à falha. A política de substituição em bloco caracteriza-se pela manutenção preventiva ser aplicada a um grupo de itens em um tempo determinado, permitindo também a substituição imediata de itens que venham a falhar antes do tempo previsto para execução da substituição. O uso dessa política difere da política de substituição por idade, pois não requer que se mantenha um arquivo das extensões dos tempos de operação de cada unidade. O sistema requer, portanto, contínuas inspeções para se detectar qualquer falha que possa ocorrer entre os blocos de substituições.

Para o caso de uma substituição por idade, esse intervalo começa em $t = 0$ e termina quando ocorre uma falha ou em uma idade de substituição $t = T$, o que ocorrer primeiro.

A probabilidade de sobreviver até o tempo $t = T$ é $R(t)$, descrito na Equação (2.3) e a probabilidade de falhar antes do tempo $t = T$ é, conseqüentemente, $1 - R(T)$. Com isso, o custo por unidade de tempo é descrito pela Equação (2.33) onde C_U é o custo de uma substituição não planejada, C_P é o custo de uma substituição planejada e $\int_0^T R(t)dt$ é o período esperado de uso do equipamento.

$$C(T) = \frac{C_U[1 - R(t)] + C_P R(t)}{\int_0^T R(t)dt} \quad (2.33)$$

Para o caso da política de substituição em bloco, as substituições sempre ocorrem em um tempo $t = T$ independente da possibilidade da ocorrência de eventos de falhas antes do tempo $t = T$.

O custo por unidade de tempo é dado pela Equação (2.34).

$$C(T) = \frac{C_U}{MTBF} + \frac{C_P}{T} \quad (2.34)$$

A política de substituição por idade é ideal para equipamentos caros, enquanto a política de substituição em bloco é mais apropriada para equipamentos baratos e com grande quantidade de itens extras em estoque.

Na política de manutenção preventiva periódica, a manutenção é realizada periodicamente em intervalos de tempo fixos kT ($k = 1, 2, \dots$) independentemente do histórico de falhas do componente.

Na política de limite de falhas, a manutenção só é realizada quando a taxa de falhas atinge um nível pré-determinado. Essa política faz com que um componente trabalhe num nível de confiança aceitável ou acima dele.

Na política de manutenção preventiva sequencial, a manutenção de um componente é realizada em intervalos de tempo desiguais. Geralmente, esse intervalo de tempo vai se tornando menor com o tempo devido ao fato de que o componente vai precisar de manutenções mais frequentes com o passar do tempo.

Existem dois tipos de política de reparo limitado: a de custo de reparo limitado e a de tempo de reparo limitado. Na primeira, o custo de reparo é estimado e o reparo é realizado caso a estimativa esteja abaixo de um valor pré-determinado, caso contrário, o componente é substituído. Na política de tempo de reparo limitado, se o reparo não for concluído dentro de um intervalo de tempo pré-determinado, o componente é substituído por um novo.

Na política de contagem do número de reparos, o componente é substituído na k -ésima falha e as primeiras $(k - 1)$ falhas são removidas com reparos mínimos.

Na política de tempo de referência, se a k -ésima falha ocorrer antes de um tempo T de referência, um reparo mínimo é realizado e, na próxima falha, o item é substituído. Caso a k -ésima falha ocorra após o tempo T , o item é substituído imediatamente.

2.7 CONSIDERAÇÕES FINAIS

Este capítulo apresentou os conceitos básicos sobre os sistemas de transferência eletrônica de fundos. Em seguida, conceitos sobre avaliação de desempenho e avaliação de dependabilidade foram apresentados. Então, técnicas para avaliação de desempenho e avaliação de dependabilidade foram descritas. Posteriormente, os conceitos básicos de avaliação de performabilidade foram descritos. Finalmente, conceitos sobre políticas de manutenção são apresentados.

Redes de Petri

Este capítulo apresenta os principais conceitos sobre redes de Petri (*Petri Nets* - PNs), assim como características, propriedades e técnicas de análise. Em seguida, são apresentadas as redes de Petri estocásticas (*Stochastic Petri Nets* - SPNs), que são uma extensão à teoria inicial das redes de Petri. Finalmente, são introduzidos o *moment matching* e a técnica de aproximação de fases.

3.1 INTRODUÇÃO

O conceito de redes de Petri foi introduzido por Carl Adam Petri, no ano de 1962, com a apresentação da sua tese de doutorado “*Kommunikation mit Automaten*” (comunicação com autômatos) [Mur89] na faculdade de Matemática e Física da Universidade Darmstadt na Alemanha. Redes de Petri são ferramentas gráficas e matemáticas usadas para descrição formal de sistemas caracterizados pelas propriedades de concorrência, paralelismo, sincronização, distribuição, assincronismo e não-determinismo.

A aplicabilidade das Redes de Petri como ferramenta para estudo de sistemas é importante por permitir representação matemática, análise dos modelos e também por fornecer informações úteis sobre a estrutura e o comportamento dinâmico dos sistemas modelados. As aplicações das Redes de Petri podem se dar em muitas áreas (sistemas de manufatura, desenvolvimento de *software*, sistemas administrativos, entre outros).

As redes de Petri são formadas por lugares (Figura 3.1(a)), transições (Figura 3.1(b)), arcos (Figura 3.1(c)) e marcas (Figura 3.1(d)). Os lugares correspondem às variáveis de estado e às transições, às ações ou eventos realizados pelo sistema. A realização de uma ação está associada a algumas pré-condições, ou seja, existe uma relação entre os lugares e as transições que possibilita ou não a realização de uma determinada ação. Após a realização de uma determinada ação, alguns lugares terão suas informações alteradas, ou seja, a ação criará uma pós-condição. Os arcos representam o fluxo das marcas pela rede de Petri, e as marcas representam o estado em que o sistema se encontra em determinado momento. Graficamente, os lugares são representados por elipses ou círculos; as transições, por retângulos, os arcos, por setas e as marcas, por meio de pontos. A Figura 3.2 mostra um exemplo de rede de Petri.

Os dois elementos, lugar e transição, são interligados por meio de arcos dirigidos. Os arcos que interligam lugares às transições (Lugar \rightarrow Transição) correspondem à relação entre as condições verdadeiras (pré-condição), que possibilitam a execução das ações. Os arcos que interligam as transições aos lugares (Transição \rightarrow Lugar) representam a relação entre as ações e as condições que se tornam verdadeiras com a execução das ações

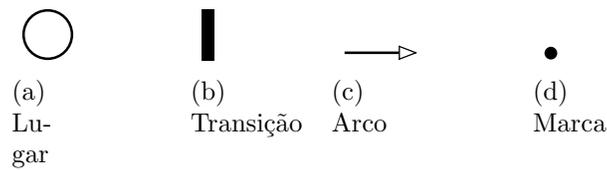


Figura 3.1: Elementos de rede de Petri

(pós-condição).

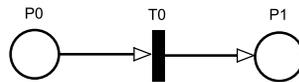


Figura 3.2: Exemplo de rede de Petri

A Figura 3.3 [MLC96] apresenta o ciclo repetitivo dos períodos do dia através de um modelo PN. Um dia pode ser dividido em três períodos distintos: manhã, tarde e noite. As transições entre os períodos do dia devem respeitar algumas condições. O período anterior à tarde é a manhã, o anterior à noite é a tarde e assim por diante. Dessa forma, podem ser estabelecidas as pré-condições e as pós-condições. Para modelar esse sistema através de uma rede de Petri são necessários três lugares, os quais representam os três períodos do dia, e três transições, as quais representam as três mudanças de período do dia. Esse modelo tem o seu estado inicial com uma marca (ponto) no lugar Manhã (Figura 3.3(a)). Com essa marcação, o único evento possível de ocorrer é o evento entardecer, representado pela transição entardecer. Após a execução desse evento, é depositada uma marca no lugar Tarde (Figura 3.3(b)). Com uma marca no lugar Tarde, o próximo evento que ocorrerá é anoitecer, representado pela transição anoitecer. A execução desse evento deposita uma marca no lugar Noite (Figura 3.3(c)). Uma marca no lugar Noite possibilita a ocorrência do evento amanhecer, este representado pela transição amanhecer. A execução desse evento reinicia o processo.

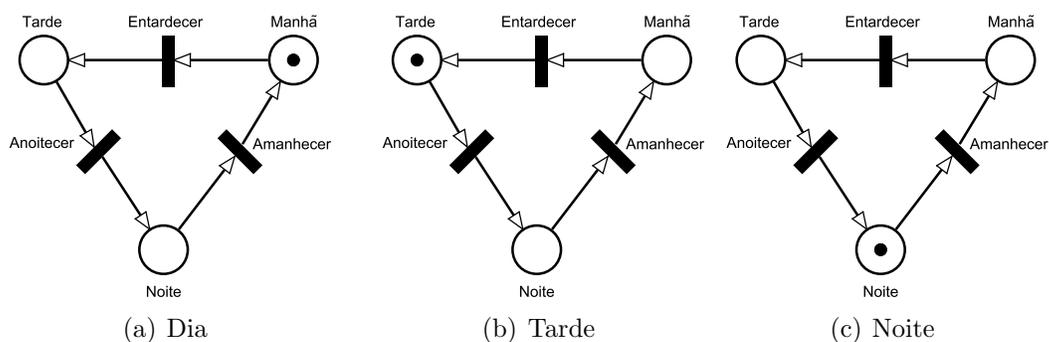


Figura 3.3: Períodos do Dia

A representação formal de um modelo PN é a quintupla $PN = \{P, T, F, W, \mu_0\}$, onde:

- P é o conjunto finito de lugares;

- T é o conjunto finito de transições, $P \cap T = \emptyset$;
- $F \subseteq (P \times T) \cup (T \times P)$ é o conjunto de arcos;
- $W : F \rightarrow IR^+ \cup \{0\}$ é a função de atribuição de peso aos arcos;
- $\mu_0 : P \rightarrow IN$ é a função de marcação inicial, onde $P \cap T = \emptyset$ e $P \cup T \neq \emptyset$.

3.1.1 Propriedades das Redes de Petri

O estudo das propriedades de redes de Petri permite a análise do sistema modelado. Os tipos de propriedades podem ser divididos em duas categorias: as propriedades dependentes de marcação inicial, conhecidas como propriedades comportamentais, e as propriedades não dependentes de marcação, conhecidas como propriedades estruturais [MLC96, Mur89].

3.1.1.1 Propriedades Comportamentais

As propriedades comportamentais são aquelas que dependem apenas da marcação inicial da rede de Petri. As propriedades abordadas são alcançabilidade, limitação, segurança, liveness e cobertura.

Alcançabilidade ou *reachability* indica a possibilidade de uma determinada marcação ser atingida pelo disparo de um número finito de transições a partir de uma marcação inicial. Dada uma rede de Petri marcada $RM = (R; M_0)$, o disparo de uma transição t_0 altera a marcação da rede. Uma marcação M' é acessível a partir de M_0 se existe uma sequência de transições que, disparadas, levam à marcação M' . Ou seja, se a marcação M_0 habilita a transição t_0 , disparando-se esta transição, atinge-se a marcação M_1 . A marcação M_1 habilita t_1 a qual, sendo disparada, atinge-se a marcação M_2 e assim por diante até a obtenção da marcação M' .

Seja $M_i[t_j > M_k$ e $M_k[t_h > M_1$, então $M_i[t_j t_h > M_1$, por recorrência, o disparo de uma sequência $s \in T^*$ é designado por $M[s > M']$. O conjunto de todas as possíveis marcações obtidas a partir da marcação M_0 na rede $RM = (R; M_0)$ é denotado por $CA(R; M_0) = \{M' \in IN^m \mid \exists s, M_0[s > M']\}$, onde m é a cardinalidade do conjunto de lugares da rede.

A análise da alcançabilidade de uma marcação consiste em determinarmos uma dada marcação $M' \in CA(R; M_0)$ da rede marcada RM . Em alguns casos, deseja-se observar apenas alguns lugares específicos da rede em estudo. Esse problema é denominado sub-marcação alcançável.

Seja um lugar $p_i \in P$, de uma rede de Petri marcada $RM = (R; M_0)$, esse lugar é k -limitado ($k \in IN$) ou simplesmente limitado se para toda marcação acessível $M \in CA(R; M_0)$, $M(p_i) \leq k$.

O limite k é o número máximo de marcas que um lugar pode acumular. Uma rede de

Petri marcada $RM = (R; M_0)$ é k -limitada se o número de marcas de cada lugar de RM não exceder k em qualquer marcação acessível de $RM(\max(M(p)) = k, \forall p \in P)$.

Segurança ou *safeness* é uma particularização da propriedade de limitação. O conceito de limitação define que um lugar p_i é k -limitado se o número de marcas que esse lugar pode acumular estiver limitado ao número k . Um lugar que é 1-limitado pode ser simplesmente chamado de seguro.

Seja $p_i \in P$ um lugar de uma rede de Petri marcada $RM = (R; M_0)$, p_i é seguro se para toda marcação $M' \in CA(R; M_0)$, $M(p_i) \leq 1$. Uma rede é segura se todos os lugares pertencentes a essa rede forem seguros, ou seja, todos os lugares dessa rede podem conter no máximo uma única marca.

Vivacidade ou *liveness* está definida em função das possibilidades de disparo das transições. Uma rede é considerada *live* se, independente das marcações que sejam alcançáveis a partir de M_0 , for sempre possível disparar qualquer transição da rede através de uma sequência de transições $L(M_0)$. A ausência de bloqueio (*deadlock*) em sistemas está fortemente ligada ao conceito de vivacidade. Pois, *deadlock* em uma rede de Petri é a impossibilidade do disparo de qualquer transição da rede. O fato de um sistema ser livre de *deadlock* não significa que seja *live*, entretanto um sistema *live* implica um sistema livre de *deadlocks*.

Uma rede $RM = (R; M_0)$ é viva (*live*) se para toda $M \in CA(R; M_0)$ for possível disparar qualquer transição de RM através do disparo de alguma sequência de transições.

O conceito de cobertura está associado ao conceito de alcançabilidade e *live*. Uma marcação M_i é coberta se existir uma marcação $M_j \neq M_i$, tal que $M_j \geq M_i$.

3.1.1.2 Propriedades Estruturais

As propriedades estruturais são aquelas que dependem apenas da estrutura da rede de Petri. Essas propriedades refletem características independentes de marcação. As propriedades analisadas neste trabalho são limitação estrutural e consistência.

Uma rede de Petri $R = (P, T, F, W, \mu_0)$ é classificada como estruturalmente limitada se for limitada para qualquer marcação inicial.

Ela será considerada consistente se, disparando uma sequência de transições habilitadas a partir de uma marcação M_0 , retornar a M_0 , porém todas as transições da rede são disparadas pelo menos uma vez.

Seja $RM = (R; M_0)$ uma rede marcada e s uma sequência de transições, RM é consistente se $M_0[s > M_0$ e toda transição T_i , disparar pelo menos uma vez em s .

Os métodos de análise das propriedades das redes de Petri são classificados como análise baseada na geração do espaço de estados, análise baseada na equação de estado, métodos baseados na estrutura da rede (análise de invariantes) e técnicas de redução. A validação dos modelos pode ser realizada através de simulação [MLC96, Mur89].

O método de análise baseada na geração do espaço de estados envolve essencialmente

a enumeração das marcações alcançáveis (marcações cobertas). Esse método é aplicável a todas as classes de redes, mas é limitado a redes pequenas devido à complexidade do problema de explosão de espaço de estados. Esse método baseia-se na construção de um grafo que representa todas as marcações que a rede de Petri pode alcançar. Cada nó corresponde a uma marcação, e cada arco corresponde ao disparo de um conjunto não vazio de transições. Se a rede de Petri for limitada, é possível construir este tipo de grafo e, nesse caso, ele denomina-se grafo de ocorrências. Caso a rede de Petri não seja limitada, o grafo de ocorrências é infinito. Nesse caso, ainda é possível construir um grafo que se denomina grafo de cobertura [MLC96, Mur89].

O método de análise baseada na equação de estado ou equação fundamental possibilita a verificação da acessibilidade das marcações, assim como o número de vezes que cada transição tem que ser disparada para atingir determinada marcação. Uma rede de Petri pode ser representada por duas matrizes, uma indicando os conjuntos de lugares que servem de entrada para cada uma das transições da rede e outra indicando os conjuntos de lugares que servem de saída para tais transições. A primeira delas é a matriz de entrada, também chamada de matriz de incidência reversa. A matriz de incidência A de uma rede de Petri é uma matriz $n \times m$ de inteiros, definida como $A = [a_{ij}]$ e $a_{ij} = a_{ij}^+ - a_{ij}^-$, onde $a_{ij}^+ = w(i, j)$ é o peso do arco da transição i para seu lugar de saída j e $a_{ij}^- = w(i, j)$ é o peso do arco do lugar de entrada j para a transição i [MLC96, Mur89].

O método baseado na estrutura da rede (análise de invariantes) verifica a existência de componentes repetitivos estacionários nos modelos, onde esses componentes correspondem a comportamentos cíclicos da rede. Os invariantes em uma rede de Petri representam os componentes conservativos e repetitivos da rede. Há conjuntos de lugares e de transições da rede, cujo comportamento não se altera durante o seu funcionamento. A identificação e a interpretação de cada um destes conjuntos são importantes, pois eles refletem certas propriedades da rede que podem ser de interesse para a análise do sistema modelado. Os componentes conservativos da rede são representados em seus invariantes de lugar, ou seja, são conjuntos de lugares da rede nos quais a soma das marcas é constante durante todo o seu funcionamento. Os componentes repetitivos são representados em seus invariantes de transição, isto é, são conjuntos de transições da rede que, ao serem disparadas em determinada sequência, retornam à marcação de partida [MLC96, Mur89].

As técnicas de redução são transformações aplicadas ao modelo de um sistema com o objetivo de simplificá-lo, preservando as propriedades do sistema analisado. Normalmente essas técnicas são utilizadas para facilitar a análise de sistemas complexos. Essas técnicas são baseadas nas transformações de redes originais em um modelo mais abstrato de tal maneira que as propriedades como *liveness*, *boundedness* e *safeness* são preservadas nos modelos obtidos por estas reduções. A transformação reversa (refinamento) pode ser usada para processos de síntese. As regras de transformação das redes podem ser a partir de aplicação das fusões, tanto de lugares, quanto de transições [Mur89]. A Figura 3.4 apresenta algumas das técnicas de redução [MLC96, Mur89].

A simulação é utilizada quando o sistema é relativamente complexo e sua análise através de outros métodos analíticos se mostra inviável.

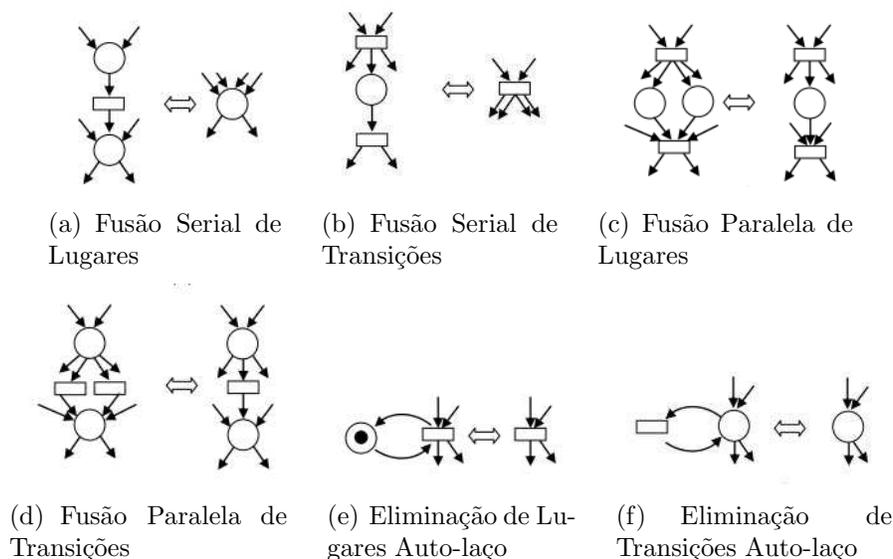


Figura 3.4: Técnicas de Redução

3.2 REDE DE PETRI ESTOCÁSTICA

Rede de Petri estocástica (SPN) [Ger00] é uma das extensões de rede de Petri (PN) [Mur89] utilizada para a modelagem de desempenho e dependabilidade. Uma rede de Petri estocástica adiciona tempo ao formalismo de redes de Petri, com a diferença de que os tempos associados às transições temporizadas são distribuídos exponencialmente, enquanto o tempo associado às transições imediatas é zero. As transições temporizadas modelam atividades através dos tempos associados, de modo que o período de habilitação da transição temporizada corresponde ao período de execução da atividade, e o disparo da transição temporizada corresponde ao término da atividade. Níveis diferentes de prioridade podem ser atribuídos às transições. A prioridade de disparo das transições imediatas é superior à das transições temporizadas. As prioridades podem solucionar situações de confusão [MBC⁺98]. As probabilidades de disparo associadas às transições imediatas podem solucionar situações de conflito [Bal01, MBC⁺98].

Uma SPN é definida pela 9-tupla $SPN = \{P, T, I, O, H, \Pi, G, M_0, Atts\}$, onde:

- $P = \{p_1, p_2, \dots, p_n\}$ é o conjunto de lugares;
- $T = \{t_1, t_2, \dots, t_m\}$ é o conjunto de transições imediatas e temporizadas, $P \cap T = \emptyset$;
- $I \in (\mathbb{N}^n \rightarrow \mathbb{N})^{n \times m}$ é a matriz que representa os arcos de entrada (que podem ser dependentes de marcações);
- $O \in (\mathbb{N}^n \rightarrow \mathbb{N})^{n \times m}$ é a matriz que representa os arcos de saída (que podem ser dependentes de marcações);
- $H \in (\mathbb{N}^n \rightarrow \mathbb{N})^{n \times m}$ é a matriz que representa os arcos inibidores (que podem ser dependentes de marcações);

- $\Pi \in \mathbb{N}^m$ é um vetor que associa o nível de prioridade a cada transição;
- $G \in (\mathbb{N}^n \rightarrow \{true, false\})^m$ é o vetor que associa uma condição de guarda relacionada a marcação do lugar à cada transição;
- $M_0 \in \mathbb{N}^n$ é o vetor que associa uma marcação inicial de cada lugar (estado inicial);
- $Atts = (Dist, Markdep, Policy, Concurrency, W)^m$ compreende o conjunto de atributos associados às transições, onde:
 - $Dist \in \mathbb{N}^m \rightarrow \mathcal{F}$ é uma possível função de distribuição de probabilidade associada ao tempo de uma transição (esta distribuição pode ser dependente de marcação) (o domínio de \mathcal{F} é $[0, \infty)$);
 - $Markdep \in \{constante, enabdep\}$, onde a distribuição de probabilidade associada ao tempo de uma transição pode ser independente (*constante*) ou dependente de marcação (*enabdep*— a distribuição depende da condição de habilitação atual);
 - $Policy \in \{prd, prs\}$ define a política de memória adotada pela transição (*prd*— *preemptive repeat different*, valor padrão, de significado idêntico à *race enabling policy*; *prs*— *preemptive resume*, corresponde ao *age memory policy*);
 - $Concurrency \in \{ss, is\}$ é o grau de concorrência das transições, onde *ss* representa a semântica *single server* e *is* representa a semântica *infinity server*.
 - $W : T \rightarrow IR^+ \cup \{0\}$ é a função peso, que representa o peso (w_t) de transições imediatas e a taxa λ_t de transições temporizadas, onde:

$$\pi(t) = \begin{cases} \geq 1, & \text{se } t \text{ é uma transição imediata;} \\ 0, & \text{caso contrário.} \end{cases}$$

Se t é uma transição temporizada, então λ_t será o valor do parâmetro da função densidade probabilidade exponencial;

Se t é uma transição imediata, então W_t será um peso, que é usado para o cálculo das probabilidades de disparo das transições imediatas em conflitos.

Os arcos inibidores são usados para prevenir transições de serem habilitadas quando certa condição é verdadeira.

Os modelos SPN possuem dois tipos de estados (marcações), os estados tangíveis (*tangible*) e os estados voláteis (*vanish*). Os estados voláteis são criados em decorrência da marcação dos lugares que são pré-condições de habilitação de uma transição imediata. O termo *vanish* é usado porque as marcações chegam a esses lugares e são instantaneamente consumidas. O tempo de permanência das marcações nesses lugares é zero. Os estados tangíveis são criados em decorrência da marcação dos lugares que são pré-condições de habilitação de uma transição temporizada [MBC⁺98].

As transições temporizadas podem ser caracterizadas por diferentes políticas de memória tais como *Resampling*, *Enabling memory* e *Age memory* [MBC⁺98].

Resampling: A cada disparo de toda e qualquer transição do modelo, todos os temporizadores existentes são reiniciados (*Restart*), e, sendo assim, não há memória. O temporizador de cada transição será reiniciado sempre que a transição tornar-se habilitada;

Enabling memory: A cada disparo de transição, os temporizadores das transições que estavam desabilitadas são reiniciados, enquanto que os temporizadores das transições que estavam habilitadas mantêm o valor atual (*Continue*). Assim que estas transições tornarem-se habilitadas novamente, seus temporizadores continuam do ponto onde foram parados. Uma variável (*enabling memory variable*) mede o tempo que a transição passou habilitada desde o último instante de tempo em que ela se tornou habilitada;

Age memory: Após cada disparo, os temporizadores de todas as transições mantêm seus valores atuais (*Continue*). Uma memória do passado é mantida por uma variável (*age memory variable*) associada a cada transição temporizada. Esta variável contabiliza o tempo gasto na atividade modelada pela transição, medindo o tempo cumulativo de habilitação, desde o instante do seu último disparo.

As transições temporizadas podem ser caracterizadas por diferentes semânticas de disparo conhecidas como *single server*, *multiple server* e *infinite server* [MBC⁺98].

Na semântica *single server*, as marcações são processadas serialmente. Após o primeiro disparo da transição temporizada, o temporizador é reiniciado como se a transição temporizada tivesse sido habilitada novamente. Esse tipo de semântica é utilizada nos modelos de disponibilidade, considerando-se que haja apenas uma única equipe de manutenção, quando vários componentes do sistema entram numa condição de falha;

Na semântica *multiple server*, as marcações são processadas com um grau máximo K de paralelismo. Caso o grau de habilitação seja maior do que K , não será criado nenhum novo temporizador para processar o tempo para o novo disparo até que o grau de habilitação tenha diminuído abaixo de K . Esse tipo de semântica é utilizado nos modelos de disponibilidade considerando-se que haja um número de equipes de manutenção menor do que o número de componentes na condição de falha. Os componentes em excesso ficarão em fila;

Na semântica *infinite server*, o valor de K é infinito, todas as marcações são processadas em paralelo, e as temporizações associadas são decrementadas a zero em paralelo. Esse tipo de semântica é utilizada nos modelos de disponibilidade, considerando-se que haja tantas equipes de manutenção quantos sejam os componentes em falha. Para cada componente existe uma equipe de manutenção exclusiva e independente. Nesse tipo de semântica, todas as marcações são processadas em paralelo.

Nos modelos SPN, as transições são disparadas obedecendo à semântica *interleaving* de ações [MBC⁺98]. Essa semântica define que as transições são disparadas uma a uma, mesmo que o estado compreenda transições imediatas não conflitantes. A análise de um modelo SPN requer a solução de um sistema de equações igual ao número de marcações tangíveis. O gerador infinitesimal Q da cadeia de Markov de tempo contínuo (CTMC) associado ao modelo SPN é derivado de uma redução de um gráfico de alcançabilidade,

rotulado com as taxas das transições temporizadas ou pesos das transições imediatas.

Modelos SPN permitem a geração de gráficos de alcançabilidade a partir dos quais cadeias de Markov de tempo contínuo (CTMC) são diretamente derivadas. A Figura 3.5 apresenta um exemplo de geração de gráfico de alcançabilidade a partir de um modelo SPN. No modelo SPN mostrado na Figura 3.5(a), existe um conflito entre duas transições imediatas ($T1$ e $T2$). A Figura 3.5(b) mostra o gráfico de alcançabilidade com a indicação de que o estado $P1$ é volátil. O disparo da transição temporizada $T0$ torna o lugar $P1$ marcado, habilitando as duas transições imediatas, $T1$ e $T2$, gerando o estado $P1$. Há uma mudança imediata (tempo zero) para o estado $P2$ ou $P3$, através do disparo da transição imediata $T1$ ou $T2$, com probabilidades $\frac{\alpha}{\alpha+\beta}$ e $\frac{\beta}{\alpha+\beta}$, respectivamente. A Figura 3.5(c) mostra o gráfico de alcançabilidade tangível após a eliminação do estado volátil $P1$.

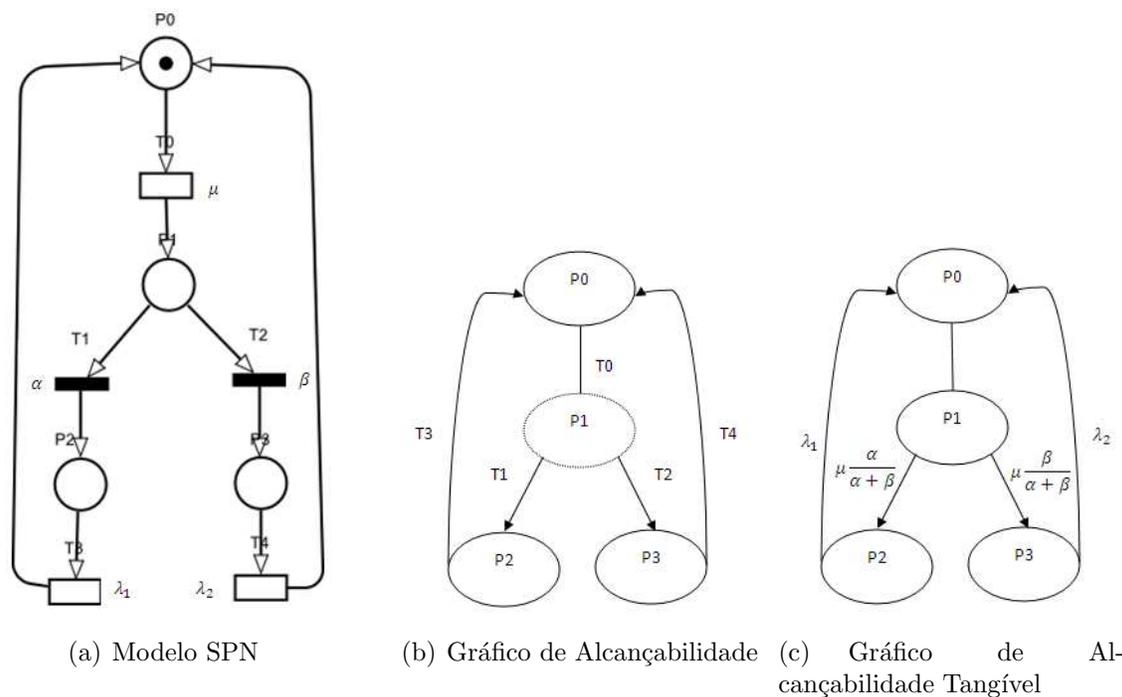


Figura 3.5: Geração de Gráfico de Alcançabilidade

A taxa na qual o sistema se move do estado $P0$ para $P2$ ou $P3$ é obtida pelo produto da taxa λ da transição do estado $P0$ para o estado volátil $P1$, com a probabilidade de ir do estado $P1$ para o estado $P2$ ou $P3$.

Redes de Petri estocásticas marcadas, com um número finito de lugares e transições, são isomórficas as cadeias de Markov [Mur89]. O isomorfismo de um modelo SPN com uma cadeia de Markov é obtido a partir do gráfico de alcançabilidade reduzido, que é dado através da eliminação dos estados voláteis e rótulo dos arcos com as taxas das transições temporizadas e pesos das transições imediatas. As medições de desempenho e dependabilidade são obtidas através de simulações e análises em estado estacionário e

transiente baseadas na cadeia de Markov embutida no modelo SPN [BGdMT06].

Os modelos SPN são usados para análise de desempenho e dependabilidade de sistemas, visto que permitem a descrição das atividades de sistemas através de gráficos de alcançabilidade. Esses gráficos podem ser convertidos em modelos Markovianos, que são utilizados para avaliação quantitativa do sistema analisado.

3.3 MOMENT MATCHING

Modelos SPN consideram somente transições imediatas e transições temporizadas com tempos de disparo distribuídos exponencialmente. Essas transições modelam ações, atividades e eventos.

Uma variedade de atividades podem ser modeladas através do uso dos construtores *throughput subnets* e *s-transitions*. Esses construtores são utilizados para representar distribuições expolinomiais, tais quais as distribuições Erlang, Hipoexponencial e Hiperexponencial [DAJ95].

Combinações de lugares, transições exponenciais e transições imediatas podem ser usadas entre dois lugares para representar diferentes tipos de distribuições. As Figuras 3.6(a), 3.6(b) e 3.6(c) representam três *throughput subnets*.

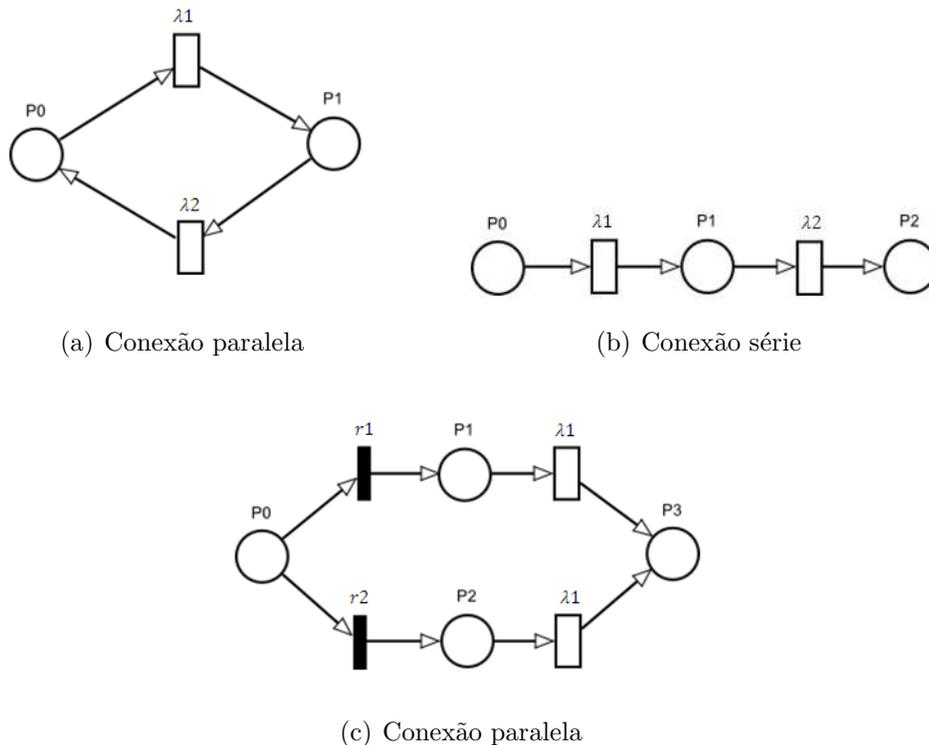


Figura 3.6: *Throughput Subnets*

A Figura 3.6(a) descreve uma *throughput subnet* formada por duas transições expo-

nenciais em paralelo com taxas λ_1 e λ_2 , respectivamente.

Uma marcação no lugar $P0$ aparecerá no lugar $P1$ após o disparo de uma das transições exponenciais que estão em paralelo as quais têm tempos associados τ_1 e τ_2 , respectivamente, (Equação (3.1)). A função de densidade para esses tempos é dada pela Equação (3.2).

$$\tau = \min(\tau_1 + \tau_2) \quad (3.1)$$

$$f_\tau(t) = (\lambda_1 + \lambda_2) \exp^{-(\lambda_1 + \lambda_2)t}, t \geq 0 \quad (3.2)$$

Essas transições exponenciais em paralelo são equivalentes a uma transição exponencial com taxa $\lambda_1 + \lambda_2$.

A Figura 3.6(b) descreve uma *throughput subnet* formada por duas transições exponenciais em série com os parâmetros λ_1 e λ_2 , respectivamente. Uma marcação no lugar $P0$ aparecerá no lugar $P2$ após o disparo das transições exponenciais, as quais têm um tempo associado $\tau = \tau_1 + \tau_2$, cuja função de densidade é dada pela Equação (3.3).

$$f_\tau(t) = (f_{\tau_1} * f_{\tau_2})(t) = \frac{\lambda_1 \lambda_2 (\exp^{-\lambda_1 t} - \exp^{-\lambda_2 t})}{\lambda_2 - \lambda_1}, t \geq 0 \quad (3.3)$$

* é o operador de convolução. Para o caso onde $\lambda_1, \lambda_2 = \dots = \lambda_n$, a função densidade é dada pela Equação (3.4).

$$f_\tau(t) = \frac{\lambda^n t^{n-1} \exp^{-\lambda t}}{(n-1)!}, t > 0 \quad (3.4)$$

Essa expressão representa uma distribuição do tipo Erlang de ordem N . Uma distribuição do tipo Erlang é especificada por dois parâmetros $\lambda > 0$ e $n > 0$.

A Figura 3.6(c) descreve uma *throughput subnet* formada por duas subredes paralelas, cada uma contendo uma transição imediata e uma transição exponencial. Uma marcação no lugar $P0$ aparecerá no lugar $P3$ após o disparo das transições imediatas e exponenciais em cada sub-rede. A probabilidade de cada sub-rede é determinada pelos pesos r_1 e r_2 das transições imediatas. A função de densidade dos tempos associados as transições exponenciais é dada pela Equação (3.5), que é uma distribuição hiperexponencial.

$$f_\tau(t) = r_1 f_{\tau_1}(t) + r_2 f_{\tau_2}(t) = r_1 \lambda_1 \exp^{-\lambda_1 t} + r_2 \lambda_2 \exp^{-\lambda_2 t}, t > 0 \quad (3.5)$$

Essa *throughput subnet* implementa uma função de densidade com tempo hiperexponencial, cuja distribuição hiperexponencial é descrita pela Equação (3.6).

n , a ordem

$r_j, j = 1 \dots n$,

$$\lambda_j, j = 1 \dots n.$$

$$\sum r_j = 1 \quad (3.6)$$

A técnica de aproximação de fases pode ser aplicada para modelar ações, atividades e eventos não-exponenciais através do *moment matching*. O método apresentado calcula o primeiro momento em torno da origem (média) e o segundo momento central (variância) e estima os momentos respectivos da *s-transition* [DAJ95].

Dados de desempenho ou dependabilidade medidos ou obtidos de um sistemas (distribuição empírica) com média μ_D e desvio-padrão σ_D podem ter seu comportamento estocástico aproximados através da técnica de aproximação de fases. O inverso do coeficiente de variação dos dados medidos (Equação (3.7)) permite a seleção da distribuição expolinomial que melhor se adapta à distribuição empírica.

$$\frac{1}{CV} = \frac{\mu_D}{\sigma_D} \quad (3.7)$$

A rede de Petri descrita na Figura 3.7 representa uma atividade com distribuição de probabilidade genérica.

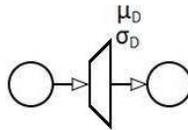


Figura 3.7: Distribuição Empírica

Dependendo do valor de inverso do coeficiente de variação dos dados medidos (Equação (3.7)), a respectiva atividade tem uma dessas distribuições atribuídas: Erlang, Hipoexponencial ou Hiperexponencial.

Quando o inverso do coeficiente de variação é um número inteiro e diferente de um, os dados devem ser caracterizados através da distribuição Erlang, que é representada por uma sequência de transições exponenciais, cujo tamanho é calculado através da Equação (3.8). A taxa de cada transição exponencial é calculada através da Equação (3.9). Os modelos de Redes de Petri descritos na Figura 3.8 representam uma distribuição Erlang.

$$\gamma = \left(\frac{\mu}{\sigma}\right)^2 \quad (3.8)$$

$$\lambda = \frac{\gamma}{\mu} \quad (3.9)$$

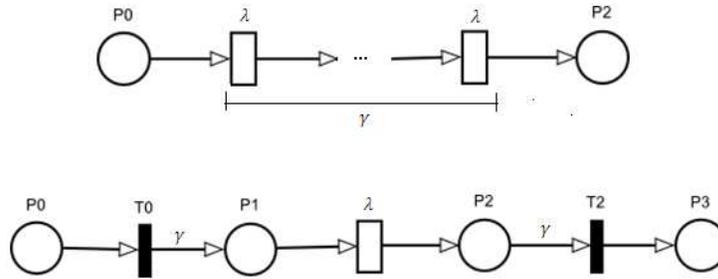


Figura 3.8: Distribuição Erlang

Quando o inverso do coeficiente de variação é um número maior que um (mas não é um número inteiro), os dados são representados através da distribuição hipoexponencial, a qual é representada por uma sequência de transições exponenciais, cujo tamanho é calculado através da Equação (3.10). As taxas das transições exponenciais são calculadas através das Equações (3.11) e (3.12), e os tempos médios atribuídos às transições exponenciais são calculados através das Equações (3.13) e (3.14). Os modelos de Redes de Petri apresentados na Figura 3.9 descrevem uma distribuição hipoexponencial.

$$\left(\frac{\mu}{\sigma}\right)^2 - 1 \leq \gamma < \left(\frac{\mu}{\sigma}\right)^2 \tag{3.10}$$

$$\lambda_1 = \frac{1}{\mu_1} \tag{3.11}$$

$$\lambda_2 = \frac{1}{\mu_2} \tag{3.12}$$

$$\mu_1 = \mu \mp \frac{\sqrt{\gamma(\gamma + 1)\sigma^2 - \gamma\mu^2}}{\gamma + 1} \tag{3.13}$$

$$\mu_2 = \gamma\mu \pm \frac{\sqrt{\gamma(\gamma + 1)\sigma^2 - \gamma\mu^2}}{\gamma + 1} \tag{3.14}$$

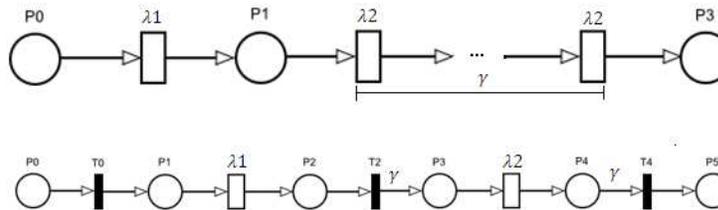


Figura 3.9: Distribuição Hipoexponencial

Quando o inverso do coeficiente de variação é um número menor que um, os dados devem ser caracterizados através de uma distribuição hiperexponencial. A taxa da transição exponencial deve ser calculada através da Equação (3.15), e os pesos das transições imediatas são calculados através das Equações (3.16) e (3.17). O modelo de Redes de Petri que representa essa distribuição hiperexponencial é descrito na Figura 3.10.

$$\lambda_h = \frac{2\mu}{\mu^2\sigma^2} \quad (3.15)$$

$$r_1 = \frac{2\mu^2}{\mu^2\sigma^2} \quad (3.16)$$

$$r_2 = 1 - r_1 \quad (3.17)$$

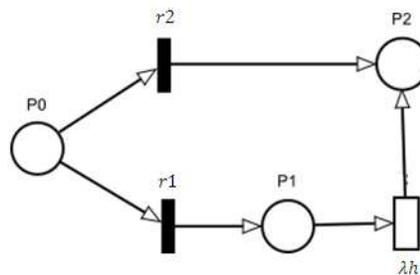


Figura 3.10: Distribuição Hiperexponencial

3.4 CONSIDERAÇÕES FINAIS

Este capítulo apresentou uma introdução sobre redes de Petri, assim como definições, conceitos básicos e propriedades, as quais podem ser divididas em duas categorias: propriedades comportamentais e propriedades estruturais. Em seguida, foram apresentadas as redes de Petri estocásticas (SPNs), que são de particular interesse deste trabalho. As SPNs possuem transições com tempos exponencialmente distribuídos e transições imediatas. Finalmente, foram apresentados o *moment matching* e a técnica de aproximação de fases, que são técnicas para aproximar o comportamento estocástico de medidas não-exponenciais.

Metodologia

METODOLOGIA PARA AVALIAÇÃO DE PERFORMABILIDADE

Este capítulo apresenta uma metodologia para avaliação de performabilidade de sistemas de transferência eletrônica de fundos (TEF). Essa metodologia considera um método para avaliação de desempenho de sistemas TEF, um método para avaliação de dependabilidade de sistemas TEF e uma estratégia para decomposição e composição, cujo objetivo é reduzir a complexidade do processo de avaliação.

4.1 VISÃO GERAL

Sistemas TEF realizam operações de crédito e débito em contas bancárias através de terminais de pontos de venda ou através de caixas eletrônicos. O desenvolvimento de técnicas, estratégias e modelos que proporcionem meios para avaliação de desempenho e dependabilidade desses sistemas são de fundamental importância para empresas que fornecem essa classe de serviços, uma vez que os recursos computacionais devem ser usados de forma a atender os níveis de serviços estabelecidos e que se faça um uso eficiente dos recursos financeiros aplicados na infraestrutura computacional. Em sistemas TEF, eventos de falhas e atividades de reparo podem causar degradação no desempenho do sistema, portanto a análise de desempenho sem considerar os efeitos da dependabilidade pode ser imprecisa. A metodologia adotada para avaliação de performabilidade de sistemas TEF combina os resultados da avaliação de desempenho e os resultados da avaliação de dependabilidade através de uma estratégia de decomposição e composição.

4.2 MÉTODO PARA AVALIAÇÃO DE DESEMPENHO

A estratégia adotada para avaliação de desempenho de sistemas TEF é composta de dez etapas: a compreensão do problema e entendimento do sistema, medição, geração do modelo abstrato, análise de propriedades e validação do modelo abstrato, geração do modelo refinado, análise de propriedades e validação do modelo refinado, mapeamento das métricas de desempenho, validação quantitativa, escolha do método de avaliação e avaliação de cenários. A Figura 4.1 mostra o diagrama de atividades do método para avaliação de desempenho de sistemas TEF.

A primeira etapa do método para avaliação de desempenho corresponde ao estudo e compreensão do problema a ser analisado, a identificação dos componentes do sistema TEF. Nesta etapa, deve-se analisar o impacto das transações comerciais de crédito e débito no desempenho da infraestrutura do sistema TEF, assim como deve-se definir os

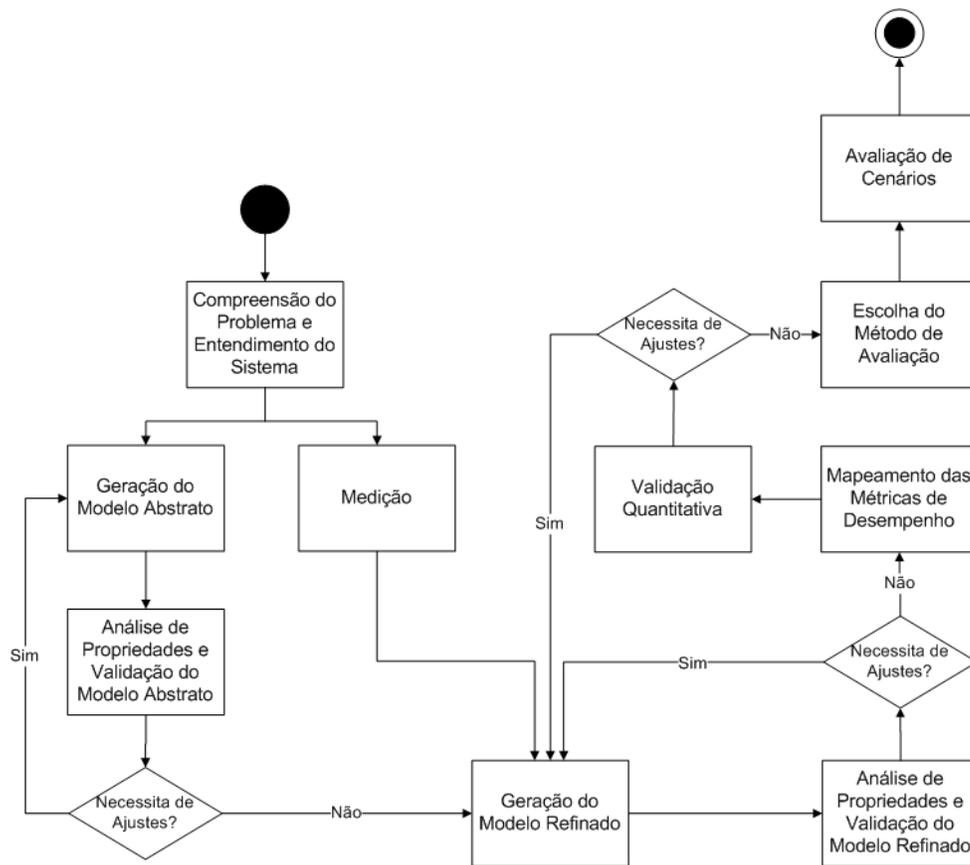


Figura 4.1: Método para Avaliação de Desempenho

critérios de desempenho que serão considerados no processo de avaliação.

A etapa de medição compreende a elaboração de um documento descrevendo como a medição deve ser realizada. Esse documento descreve quatro fases para realização da medição: planejamento, coleta de dados, análise de dados e tratamento estatístico dos dados.

A fase de planejamento tem como resultado um documento descrevendo a forma como a medição deve ser realizada. Esse documento contém informações sobre “o quê”, “onde”, “como”, “quem”, “quando” e a “frequência” do processo de coleta dos dados, assim como a forma de armazená-los, o plano de análise e quem deverá realizar cada tarefa. **O quê** é um conjunto de critérios subjetivos que devem ser transformados em métricas de desempenho. **Onde** é a definição do ambiente de medição. **Como** é a especificação de como realizar as medições, do processo de calibração dos equipamentos e ferramentas utilizadas. **Quem** é a pessoa responsável pela execução da medição. **Quando e a Frequência** são a garantia da representatividade dos dados coletados.

A fase de coleta dos dados corresponde à definição do ambiente de medição, calibração dos equipamentos e ferramentas utilizadas na coleta de dados. Nessa fase define-se o formato em que os dados serão coletados e onde serão armazenados para posterior análise.

A fase de análise dos dados corresponde à aplicação de métodos estatísticos nos dados coletados com o objetivo de fornecer informações precisas a respeito do sistema em avaliação. Os resultados das estatísticas geradas com a análise de dados são a média (μ_D), o desvio padrão (σ_D) e o coeficiente de variação.

A fase de tratamento estatístico dos dados utiliza o *moment matching* [DAJ95] para calcular os dois primeiros momentos da distribuição empírica (dados coletados), a média (μ_D) e o desvio padrão (σ_D) e os associa aos respectivos primeiros momentos da *s-transition*. As estatísticas obtidas permitem a seleção da distribuição expolinomial (hipoexponencial, hiperexponencial ou *erlang*) que melhor se adapta à distribuição empírica. A Figura 4.2 apresenta o diagrama de atividades de medição.

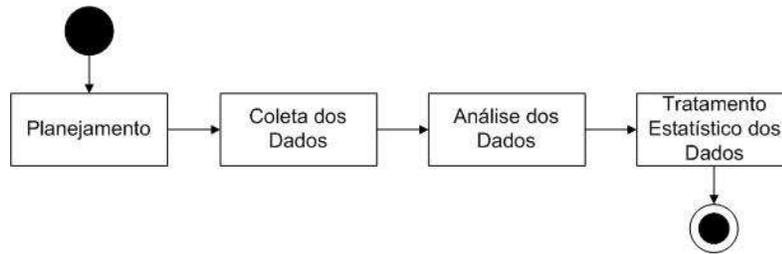


Figura 4.2: Fluxo de Medição de Desempenho

A etapa de geração do modelo abstrato corresponde à geração do modelo de desempenho, que é utilizado para compreender o comportamento do sistema e estimar o desempenho desse sistema quando submetido a diferentes cargas de trabalho ou variações na infraestrutura. Os modelos podem ser expressos em diferentes níveis. A escolha da granularidade determina as análises que podem ser realizadas, dependendo do detalhamento dos componentes do sistema [MA05].

A etapa de análise de propriedades e validação do modelo abstrato tem como objetivo a verificação das propriedades qualitativas [MLC96, Mur89] e a validação do modelo. Essa etapa ocorre após a geração do modelo abstrato. Nessa etapa, são avaliadas propriedades como: alcançabilidade, limitação, segurança, liveness e cobertura. A verificação dessas propriedades permite a validação do modelo gerado. No processo de análise e validação, pode-se observar a necessidade de ajustes no modelo. Após os ajustes, o modelo deve ser novamente analisado e validado. Após a etapa de análise qualitativa e validação do modelo abstrato e da etapa de medição, o modelo refinado deve ser obtido e, posteriormente, analisado qualitativamente e validado.

A etapa de geração do modelo refinado corresponde à geração do modelo de desempenho em função do modelo abstrato e das estatísticas obtidas na fase de medição. Essas estatísticas sugerem o tipo de distribuição expolinomial que melhor se adapta à distribuição empírica (dados coletados). Essa adequação é realizada com o auxílio da técnica de *moment matching* [DAJ95], que calcula os dois primeiros momentos da distribuição empírica, a média (μ_D) e o desvio-padrão (σ_D), e os associa aos dois primeiros momentos da *s-transition*. Dessa forma, primeiro determina-se qual o tipo de distribuição expolinomial que melhor se adapta aos dados coletados, em seguida, encontram-se os valores dos

parâmetros numéricos da distribuição expolinomial escolhida.

Como no modelo abstrato, o modelo refinado deve ser qualitativamente analisado e validado. A verificação de propriedades como alcançabilidade, limitação, segurança, liveness e cobertura permite a validação do modelo refinado [MLC96, Mur89].

O mapeamento das métricas de desempenho corresponde ao processo de representação do conjunto de critérios de avaliação em métricas através de referências aos elementos do modelo refinado obtido.

A etapa de validação quantitativa analisa se os resultados das métricas de desempenho calculadas pelo modelo são comparáveis aos obtidos através de medições no sistema, considerando um erro de exatidão aceitável.

Após validação quantitativa do modelo, as técnicas de avaliação devem ser escolhidas. A escolha do método apropriado depende fundamentalmente das métricas a serem avaliadas e das características do modelo desenvolvido. Para que o modelo seja analisável, é necessário que o modelo tenha algumas propriedades qualitativas (nas etapas anteriores de Análise de Propriedades e Validação - ver Seção 3.1.1).

A análise transiente analisa o comportamento do modelo a partir de uma marcação inicial no tempo zero até um instante especificado. O emprego dessa técnica permite responder perguntas do tipo: qual é a probabilidade de o sistema TEF estar operando após um ano?

Uma análise de estado estacionário computa o desempenho médio do sistema depois que todos os efeitos transientes iniciais passarem e um estado de equilíbrio tenha sido alcançado, se for o caso. Nessa técnica, as medidas de desempenho podem ser usadas para responder perguntas do tipo: qual a utilização da infraestrutura de processamento do servidor TEF em função da taxa de envio de transações comerciais de crédito e débito?

A análise realiza a avaliação de desempenho de forma numérica com base na exploração do gráfico de alcançabilidade (cadeia de Markov) [BGdMT06, Mur89]. A escolha da técnica de avaliação depende também dos recursos computacionais disponíveis, uma vez que a análise exige mais recursos de memória que a simulação, dada a necessidade do armazenamento do espaço de estados. Embora a simulação não demande recursos de armazenamento substanciais, os resultados computados são aproximações, diferentemente da análise que fornece resultados exatos.

A etapa de avaliação de cenários corresponde à análise de diferentes cenários com o objetivo de encontrar configurações adequadas em termos de custo/benefício. A análise pode compreender o estudo do sistema em função de variações na demanda de usuários, nos níveis de carga de trabalho e na infraestrutura dos sistemas.

4.3 MÉTODO PARA AVALIAÇÃO DE DEPENDABILIDADE

A estratégia adotada para avaliação de dependabilidade de sistemas TEF é composta de nove etapas: a compreensão do problema e entendimento do sistema, obtenção dos dados,

geração do modelo abstrato, análise de propriedades e validação do modelo abstrato, geração do modelo refinado, análise de propriedades e validação do modelo refinado, mapeamento das métricas de dependabilidade, escolha do método de avaliação e avaliação de cenários. A Figura 4.3 mostra o diagrama de atividades do método para avaliação de dependabilidade de sistemas TEF.

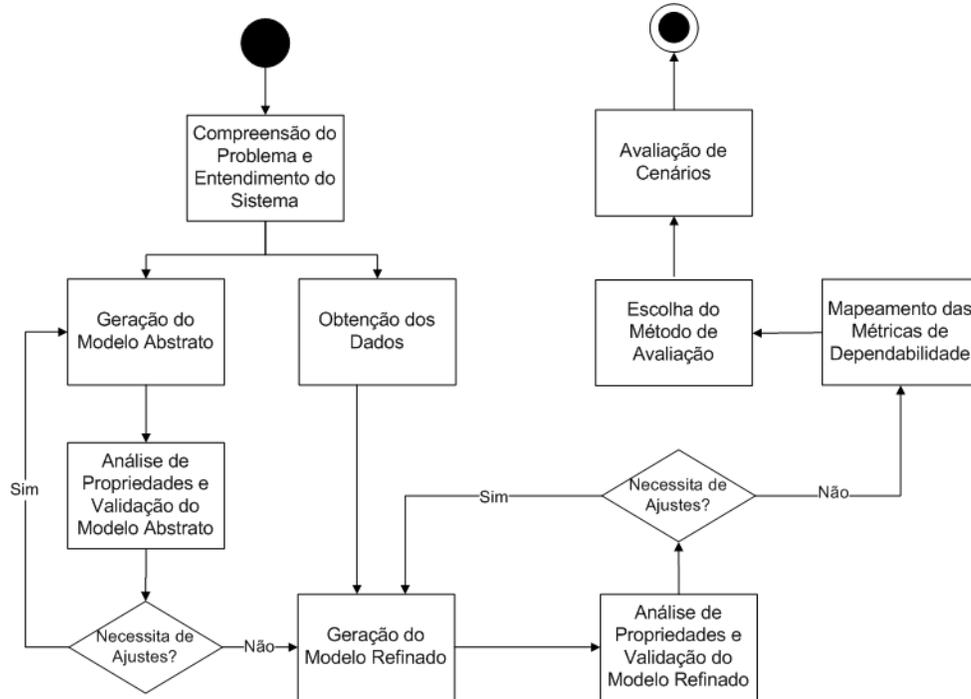


Figura 4.3: Método para Avaliação de Dependabilidade

Na etapa de compreensão do problema e entendimento do sistema, são definidos os requisitos de dependabilidade (disponibilidade e confiabilidade) a serem alcançados e o intervalo de tempo em que esses requisitos devem ser válidos.

A etapa de obtenção dos dados compreende a elaboração de um documento que descreve a técnica utilizada para obtenção dos dados de dependabilidade. Caso a técnica utilizada seja medição, o documento elaborado será similar ao descrito na etapa de medição do método para avaliação de desempenho de sistemas TEF.

A etapa de geração do modelo abstrato corresponde à geração do modelo de dependabilidade. Esse modelo é utilizado para estimar a disponibilidade e a confiabilidade do sistema quando submetido a variações na infraestrutura.

A etapa de geração do modelo refinado corresponde à geração do modelo de dependabilidade em função do modelo abstrato e das estatísticas computadas na etapa de obtenção dos dados.

Como no método para avaliação de desempenho, a etapa de análise das propriedades e validação dos modelos (abstrato e refinado) tem como objetivo a verificação de propriedades qualitativas como alcançabilidade, limitação, segurança, liveness e cobertura

[MLC96, Mur89] e a validação dos modelos.

O mapeamento das métricas de dependabilidade corresponde à definição das métricas de acordo com a particularidade do sistema e do estudo que se pretende realizar.

A etapa de avaliação de dependabilidade compreende a análise do efeito dos eventos de falhas e das atividades de reparo na disponibilidade e confiabilidade dos sistemas. Nessa etapa, técnicas de avaliação são aplicadas levando em consideração os requisitos da infraestrutura disponíveis, a estrutura do modelo refinado e as métricas de dependabilidade que se deseja avaliar.

A etapa de avaliação de cenários corresponde à análise da disponibilidade e confiabilidade de sistemas considerando variações na infraestrutura do sistema durante um intervalo de tempo.

4.4 ESTRATÉGIA DE DECOMPOSIÇÃO E COMPOSIÇÃO

A análise de performabilidade de sistemas TEF é realizada através da composição dos resultados obtidos da avaliação de desempenho e da avaliação de dependabilidade. Uma técnica de decomposição hierárquica divide o modelo de performabilidade em dois modelos distintos, um modelo de desempenho e um modelo de dependabilidade [HMRT01].

A estratégia adotada combina um modelo de dependabilidade de alto nível, o qual considera eventos de falhas e atividades de reparo, a um conjunto de modelos de desempenho de baixo nível [PRS96, STP96].

A análise da performabilidade da infraestrutura de sistemas TEF descreve o efeito da dependabilidade no desempenho do sistema através de métricas de desempenho e dependabilidade. Essas métricas são calculadas, independentemente, a partir dos modelos de desempenho e dependabilidade e posteriormente combinadas para mostrar o efeito da dependabilidade no desempenho do sistema.

A Figura 4.4 apresenta o diagrama de atividades da metodologia para avaliação de performabilidade.

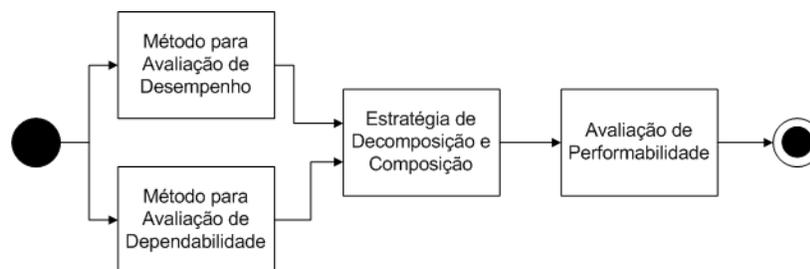


Figura 4.4: Metodologia para Avaliação de Performabilidade

4.5 CONSIDERAÇÕES FINAIS

Neste capítulo, foi apresentada a metodologia para avaliação de performabilidade de sistemas de transferência eletrônica de fundos (TEF), a qual considera o método para avaliação de desempenho de sistemas TEF, o método para avaliação de dependabilidade de sistemas TEF e uma estratégia hierárquica. Os métodos para avaliação de desempenho e dependabilidade são iniciados a partir da compreensão do problema que será analisado, entendimento dos componentes do sistema TEF e identificação dos critérios de desempenho e dependabilidade que serão utilizadas na geração dos modelos abstratos. A partir dos modelos abstratos e das estatísticas obtidas, derivam-se modelos refinados que incluem os aspectos temporais através das estruturas específicas. Esses modelos refinados e as métricas de desempenho e dependabilidade são combinados com o objetivo de analisar o efeito da dependabilidade no desempenho do sistema.

Modelos

MODELOS PROPOSTOS

Este capítulo apresenta os modelos SPN de desempenho e dependabilidade do sistema TEF. Inicialmente, apresentam-se o sistema TEF e seus componentes. Posteriormente, modelos abstratos de desempenho e dependabilidade são propostos e discutidos. Em seguida, apresenta-se a aplicação da metodologia apresentada no Capítulo 4.

5.1 SISTEMA TEF

O sistema TEF gerencia todo o processo de transação comercial realizado entre os pontos de vendas e as autorizadoras de crédito de débito. Esse sistema é composto de aplicativos clientes e de aplicativos de gerenciamento. Os aplicativos clientes são configurados nos terminais de pontos de vendas de diversas empresas e filiais, os aplicativos de gerenciamento são configurados no servidor TEF. O aplicativo cliente é responsável pela interface entre os pontos de vendas e o aplicativo de gerenciamento. O aplicativo de gerenciamento controla todo o processo de transação comercial. As fases desse processo são a exibição de mensagens, leitura do cartão magnético, coleta de senha e impressão do cupom. Além disso, os aplicativos de gerenciamento recebem todas as solicitações geradas pelos pontos de vendas, incluindo todos os parâmetros necessários para a construção das mensagens a serem enviadas pelos terminais de pontos de vendas e, em seguida, encaminham à transação comercial. Dessa forma, o sistema TEF gerencia as operações comerciais das empresas e filiais, controla o status dos pontos de venda e troca de mensagens entre os pontos de venda e as autorizadoras. A Figura 5.1 apresenta o sistema TEF.

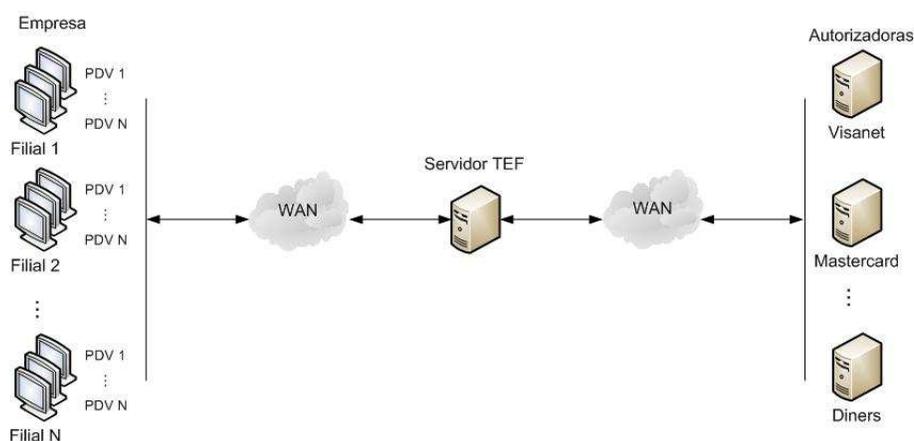


Figura 5.1: Sistema TEF

5.2 MODELO DE DESEMPENHO DE SISTEMAS TEF

Esta seção apresenta o modelo SPN concebido para avaliação de desempenho de sistemas TEF [SMA⁺09b]. O modelo proposto (ver Figura 5.2) é composto de sub-modelos (sub-redes que descrevem o comportamento do sistema TEF). As sub-redes Cliente representam os aplicativos clientes dos terminais de pontos de vendas de diferentes empresas, tais como farmácias, supermercados, postos de gasolina, lojas e *shopping centers*. Essas sub-redes podem ser refinadas para representar uma grande variedade de carga de trabalho (operações comerciais de crédito e débito) relacionadas a um determinado período ou época do ano. Esses modelos podem representar um grande número de pontos de vendas transmitindo transações comerciais de crédito e débito com frequências distintas. Assim, as demandas de serviço podem ser representadas através de diferentes frequências de transferência das transações. As marcações N_i são atribuídas aos lugares Farmácias, Supermercados, Postos de Gasolina, *Shopping Centers* e Lojas das sub-redes Cliente. Essas marcações representam um número específico de pontos de vendas de um determinado tipo. As transições estocásticas genéricas (*s-transitions*) representam a distribuição empírica dos tempos de envio das transações comerciais. A Figura 5.3 descreve uma sub-rede Cliente.

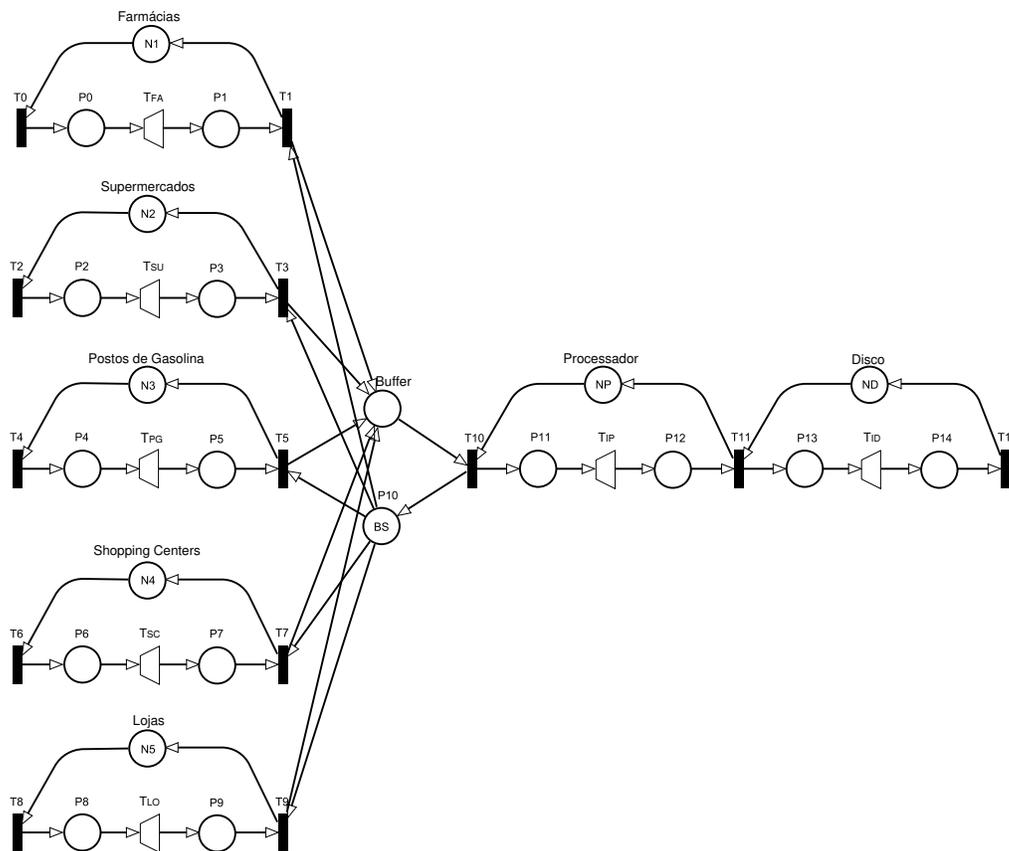


Figura 5.2: Modelo de Desempenho de Sistemas TEF

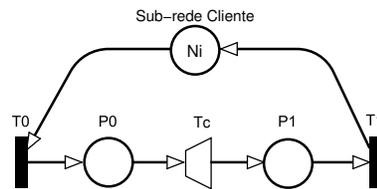


Figura 5.3: Sub-rede Cliente

A sub-rede *Buffer* representa as transações que estão esperando na fila para serem enviadas ao servidor TEF. Essa sub-rede possibilita a variação da quantidade de transações comerciais que podem permanecer na fila antes de serem atendidas, sendo, dessa forma, uma fila limitada. A marcação *BS* do lugar *P10* representa a capacidade de armazenamento das transações na fila do sistema TEF. Cada marcação do lugar *buffer* representa uma operação comercial que será processada pelo servidor TEF. A Figura 5.4 descreve a sub-rede *Buffer*.

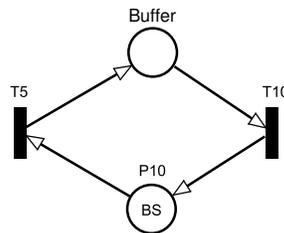


Figura 5.4: Sub-rede *Buffer*

O modelo de gerenciamento representa a infraestrutura de processamento e armazenamento do servidor TEF e é composto pelas sub-redes Processamento de Transações e Armazenamento de Transações (Ver Figura 5.5). A sub-rede Processamento de Transações representa o processamento de transações comerciais, e a sub-rede Armazenamento de Transações representa as operações de leituras e escritas realizadas pelo disco relativas às transações comerciais. A marcação *NP* é atribuída ao lugar Processador para representar a capacidade que o servidor TEF tem de processar as transações, ou seja, o número de transações concorrentes suportadas pelo recurso de processamento (grau de concorrência). A marcação *ND* é atribuída ao lugar Disco para representar o número de transações concorrentes suportadas pelo recurso de armazenamento (grau de concorrência). O tempo associado às transições estocásticas genéricas *TIP* e *TID* (*s-transitions*) representam o tempo de processamento das transações comerciais e o tempo de realização das operações de armazenamento das transações comerciais (leitura e escrita).

Esse modelo possibilita a avaliação de desempenho e o planejamento de capacidade de sistemas TEF através da avaliação dos níveis de utilização das infraestruturas de processamento e armazenamento do servidor TEF. A avaliação de cenários que consideram crescimento na demanda de clientes e flutuações na carga de trabalho auxiliam na definição de configurações adequadas garantindo o cumprimento dos acordos de níveis de serviço.

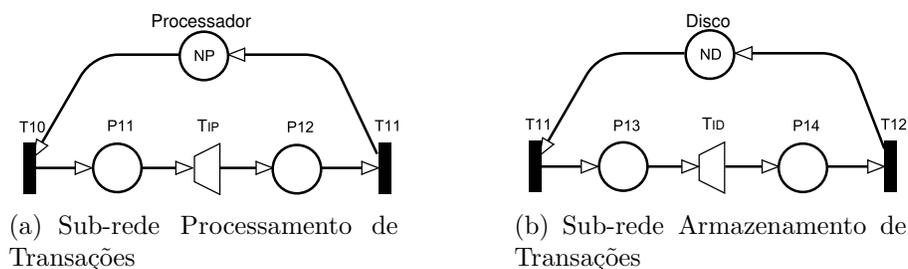


Figura 5.5: Sub-redes Processamento e Armazenamento de Transações

5.3 REFINAMENTO DO MODELO DE DESEMPENHO DE SISTEMAS TEF

Esta seção apresenta um estudo de caso para validar o modelo de desempenho proposto. Neste estudo de caso, o sistema TEF conhecido como SCOPE (Solução Completa para Pagamento Eletrônico) é adotado. O sistema SCOPE gerencia as operações comerciais e troca de mensagens realizadas entre os estabelecimentos comerciais e as empresas autorizadas de crédito e débito. Além disso, o SCOPE também armazena as informações relacionadas às transferências eletrônicas de fundos em um banco de dados permitindo que os demais aplicativos do sistema SCOPE gerem diversos tipos de consultas e relatórios estatísticos, operacionais e financeiros. Este sistema TEF é composto de aplicativos clientes (SCOPE *Client*) e de aplicativos de gerenciamento (SCOPE *Server*, SCOPE *Gateway*, SCOPE Administrador, SCOPE Pendência, SCOPE Configurator, SCOPE Con e SCOPE IBD)[Ita09b].

SCOPE *Client* é responsável pelo interfaceamento entre o SCOPE e os aplicativos dos terminais de pontos de vendas. Este aplicativo controla todo o processo de geração da transação incluindo: a exibição de mensagens, leitura do cartão magnético, coleta de senhas e impressão de cupons. SCOPE *Server* é responsável pelo processo de montagem da mensagem ISO 8583 que será enviada à empresa autorizadora para efetuar a transação eletrônica (Pré-Tef), assim como, também, gerencia todas as conexões clientes dos terminais de pontos de vendas (PDVs) e constrói os blocos de informações que são enviados às empresas autorizadoras. SCOPE *Gateway* é responsável pelo envio e recepção das transações eletrônicas para as empresas autorizadoras. Esse aplicativo também estabelece a conexão e o controle da entrada e saída de pacotes, trabalhando como um roteador de mensagens. O SCOPE Administrador utiliza o banco de dados, tratando e filtrando dados para fornecer informações sobre transações operacionais e gerenciais. Por meio desse aplicativo podem-se visualizar gráficos e planilhas referentes a todo tipo de transação eletrônica. O SCOPE Pendência é responsável pelo gerenciamento de transações pendentes. Uma transação é considerada pendente quando o SCOPE não consegue entregar ao computador da empresa autorizadora uma mensagem de confirmação referente ao sucesso da transação eletrônica. SCOPE Configurator é responsável por toda a configuração do sistema, esse aplicativo gerencia e insere informações no banco de dados, para que a IBD gerencie o restante do sistema. SCOPE Con é responsável pela conciliação dos dados gerados pelas transações eletrônicas no SCOPE e os arquivos gerados

pelas empresas autorizadoras. SCOPE IBD controla todo o acesso ao banco de dados solicitado pelos demais aplicativos, tornando o sistema independente da plataforma de banco de dados e possibilitando o seu acesso local ou remoto.

Esses aplicativos podem estar na mesma máquina ou em máquinas separadas. A escolha da arquitetura é baseada no tamanho da empresa, quantidade de recursos disponíveis, nível de investimento e nas necessidades da empresa.

O aplicativo SCOPE *Server* tem como função garantir o fluxo de transações entre o SCOPE *Client* e o SCOPE *Gateway* pelo provimento de informações ao SCOPE *Client* para realização da Pré-Tef. O SCOPE *Client* solicita informações sobre a transação de TEF a ser executada, o SCOPE *Server* encarrega-se de extrair as informações da base de dados SQL para obter os parâmetros necessários para a montagem da mensagem. O SCOPE *Server* também atualiza as transações executadas na base de dados, mantém as conexões sempre ativas e gerencia os terminais de pontos de vendas a ele conectados, através do controle no fluxo de transações. A Figura 5.6 mostra o fluxo de informações percorridas quando um cliente realiza algum tipo de operação TEF [Ita09b].

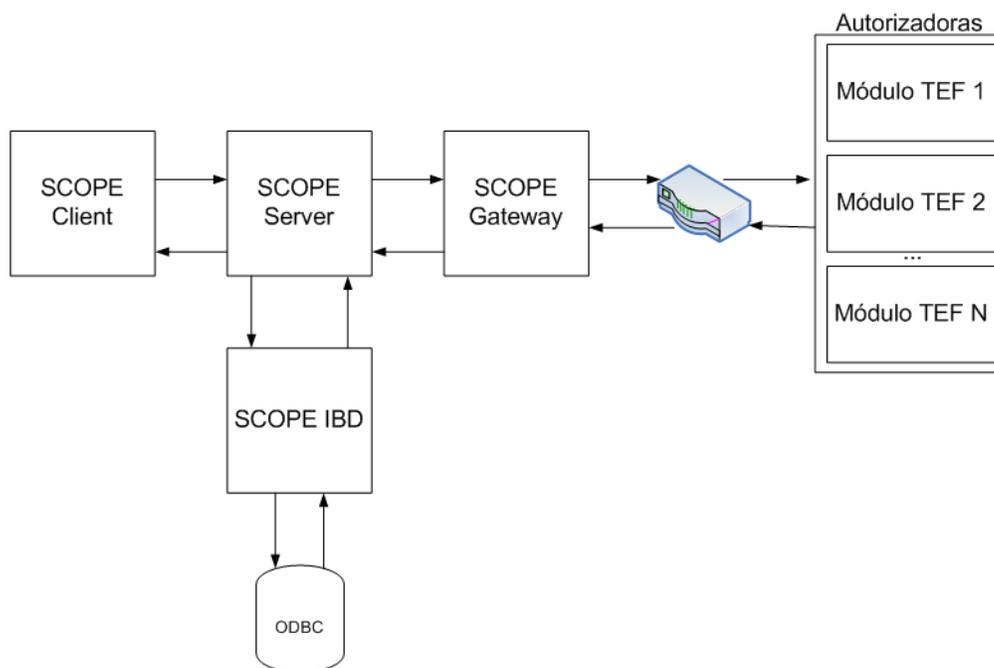


Figura 5.6: Fluxo da Transação TEF

Os experimentos de validação do modelo de desempenho proposto foram realizados no Laboratório de Avaliação de Desempenho CIn-Itautec. Nesse ambiente, os aplicativos clientes foram configurados em três servidores MX201 [Ita09a], e os aplicativos de gerenciamento foram configurados no servidor MX221 (servidor TEF) [Ita09a]. Nesses experimentos, os aplicativos clientes transmitem traces de transações comerciais coletados de clientes do sistema SCOPE. A Figura 5.7 mostra o ambiente onde foram realizados os experimentos.

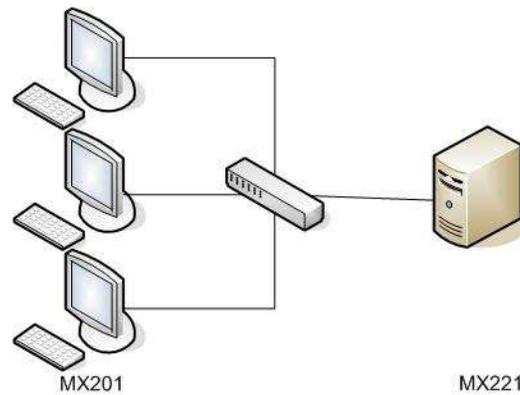


Figura 5.7: Ambiente onde foram Realizados os Experimentos

Os aplicativos clientes foram configurados para representar 3345 pontos de vendas distribuídos em um estabelecimento comercial. Esses pontos de vendas registram as transações comerciais de crédito e débito através dos aplicativos clientes do SCOPE. Os cenários avaliados descrevem um estabelecimento comercial com diversos pontos de vendas onde o registro da demanda ocorreu com 7 diferentes taxas: 100, 200, 300, 400, 500, 600 e 700 tpms (transações por minuto).

As medidas de desempenho foram obtidas no servidor TEF (MX221) através do *Windows Performance Monitor (Perfmon)*. Essas medidas são: percentual de tempo ocioso do processador, percentual de tempo ocioso do disco, transferências de operações de disco por segundo e transferências de operações de leituras por segundo, transferências de operações de escrita por segundo, tempo médio de transferências das operações de disco, tempo médio das operações de leitura do disco, tempo médio das operações de escritas do disco, comprimento médio da fila com as operações de disco, comprimento médio da fila com as operações de leituras de disco e comprimento médio da fila com as operações de escritas de disco [Fri05].

Após a configuração e estabilização do ambiente, as medições das métricas de desempenho foram iniciadas através do medidor de desempenho do *Windows (Perfmon)*. Durante a realização das medições, os processos que não são estritamente necessários para a realização dos experimentos foram encerrados a fim de evitar interferências nos dados coletados [Lil00]. As medições ocorreram durante 12 horas com um intervalo de 1 minuto entre as coletas de dados. O tempo de medição representa o período de funcionamento de um estabelecimento comercial e o intervalo entre as coletas dos dados representa o tempo médio de envio de transações comerciais. Esses dados coletados foram armazenados em *logs* gerados pela própria ferramenta e utilizados para a realização da análise estatística. Os dados coletados foram armazenados em uma partição do disco físico isolada do ambiente de medição a fim de evitar que os dados medidos fossem afetados. Essas medições objetivam a análise do impacto das variações da carga de trabalho (transações comerciais) na infraestrutura de processamento e armazenamento do servidor TEF.

O dados coletados foram tratados estatisticamente com a retirada dos *outliers* através

da ferramenta Minitab 15 [Cam03]. Os *outliers* são valores extremos ou atípicos em relação aos demais valores do conjunto amostral. Os gráficos *boxplots* da ferramenta Minitab 15 indicaram a presença de *outliers* das medidas de desempenho. A representação gráfica através do *boxplot* fornece, entre outras informações, a variabilidade e a simetria dos dados [MR03].

Medidas de desempenho como percentual de tempo de utilização do processador, percentual de tempo de utilização do disco, tempo de serviço do processador e tempo de serviço do disco são estimados indiretamente conforme as métricas obtidas por meio do medidor de desempenho do *Windows (Perfmon)* e as Equações (2.27) e (2.29) apresentadas no Capítulo 2 [Jai91, Tri06].

Entre as medidas de desempenho, o percentual de tempo de utilização do processador, o percentual de tempo de utilização do disco e as transferências de operações de disco por segundo (vazão) foram adotados para avaliar o desempenho da infraestrutura de processamento e armazenamento do servidor TEF. As medidas de desempenho de tempo de serviço do processador e tempo de serviço do disco são adotados para estimar o tempo de processamento e tempo das operações de armazenamento relacionados às transações comerciais.

Os dados medidos (distribuição empírica) foram analisados para decidir qual a distribuição expolinomial que melhor se adapta às operações de processamento e armazenamento (representadas através das transições estocásticas genéricas *TIP* e *TID*). A distribuição expolinomial mais adequada aos dados medidos foi escolhida com base nos cálculos dos valores das médias (μ_D) e desvios padrões (σ_D) dos tempos de processamento e tempos de armazenamento conforme o processo descrito no Capítulo 3 [DAJ95]. Essas transições foram refinadas de acordo com os resultados apresentados na Tabela 5.1.

Tabela 5.1: Média e Desvio Padrão

Recursos Computacionais	μ_D (s)	σ_D (s)	Distribuição Exponencial
Processador	0,001311	0,000508	Hipoexponencial
Disco	0,002756	0,000353	Hipoexponencial

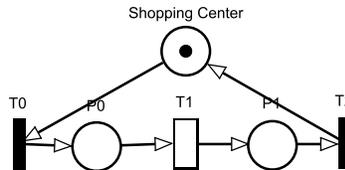
Após a definição da distribuição expolinomial adequada para representar os dados medidos, os parâmetros dessa distribuição foram calculados. Com a escolha da distribuição hipoexponencial para refinar as transições estocásticas genéricas *TIP* e *TID*, os parâmetros μ_1 , μ_2 e γ devem ser calculados. Esses parâmetros foram calculados utilizando as Equações (3.10), (3.13) e (3.14). A Tabela 5.2 mostra os valores dos parâmetros μ_1 , μ_2 e γ para as distribuições hipoexponenciais que refinaram as transições estocásticas genéricas *TIP* e *TID*.

Assim, uma versão refinada do modelo de desempenho de sistemas TEF é gerada. O modelo de desempenho de sistemas TEF refinado é obtido para o estudo de caso proposto. As marcações atribuídas ao lugar Supermercado representam pontos de vendas enviando

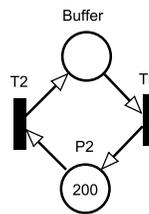
Tabela 5.2: Parâmetros das Distribuições Hipoexponenciais

Transição Estocástica	μ_1 (s)	μ_2 (s)	γ
<i>TIP</i>	0,000080	0,00054	6
<i>TID</i>	0,000001	0,00005	61

transações comerciais de crédito e débito. O envio de transações comerciais é representado pela transição temporizada *T1*. A Figura 5.8 descreve a Sub-rede Cliente Refinada.

**Figura 5.8:** Sub-rede Cliente Refinada

As marcações atribuídas ao lugar *P2* representam a capacidade de armazenamento do *Buffer*. Um grande número de marcações foram atribuídas ao lugar *Buffer* para representar o tamanho dele. Esse lugar permite o armazenamento de um grande número de marcações as quais representam as transações comerciais que o *Buffer* do sistema SCOPE pode armazenar. A definição do tamanho do *Buffer* proporciona a propriedade de limitação do modelo de desempenho de sistemas TEF [MLC96, Mur89]. A Figura 5.9 descreve a Sub-rede *Buffer* Refinada.

**Figura 5.9:** Sub-rede Buffer Refinada

O número de marcações atribuídas ao lugar Processador representa o grau de concorrência relacionada ao número de transações comerciais simultâneas que a infraestrutura de processamento do servidor TEF pode executar. O disparo das transições temporizadas *T5* e *T7* da sub-rede Processamento de Transações representa o tempo médio de processamento das transações comerciais. Da mesma forma, o número de marcações designadas ao lugar Disco representa o grau de concorrência relacionada às transações comerciais que podem ser armazenadas simultaneamente pela infraestrutura de armazenamento do servidor TEF. O disparo das transições temporizadas *T11* e *T13* da sub-rede Armazenamento de Transações representa o tempo médio das operações de leituras e escritas das transações comerciais. A Figura 5.10 descreve as sub-redes Processamento e Armazenamento de Transações Refinadas.

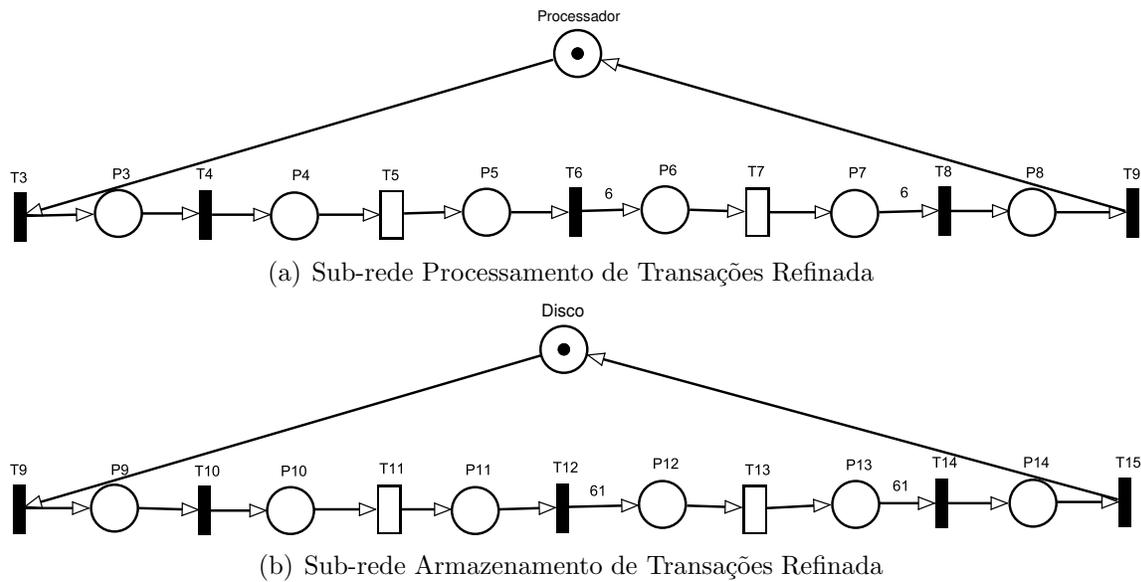


Figura 5.10: Sub-redes de Processamento e Armazenamento de Transações Refinadas

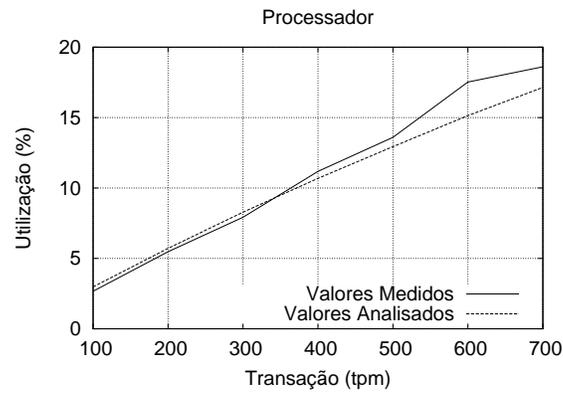
Os níveis de utilização da infraestrutura de processamento e armazenamento, assim como a vazão da infraestrutura de armazenamento, são obtidos através das expressões $UProc = P\{\#Processador = 0\}$, $UDisco = P\{\#Disco = 0\}$ e $DiscoVazao = P\{\#P12 \geq 0\}/(\mu_2)$ as quais são computadas por meio da ferramenta *TimeNET* 4.0 [ZK07]. A Figura 5.11 mostra uma comparação entre os resultados obtidos no processo de medição e na análise estacionária do modelo de desempenho de sistemas TEF proposto. Esses resultados consideram um conjunto de diferentes frequências: 100, 200, 300, 400, 500, 600 e 700 tpms (transações por minuto).

A Figura 5.11(a) mostra a comparação entre os resultados da medição dos níveis de utilização da infraestrutura de processamento e os resultados obtidos através da métrica de desempenho $UProc = P\{\#Processor = Np\}$. O resultado apresenta um erro máximo relativo de 13,60% e um erro médio relativo de 7,36%.

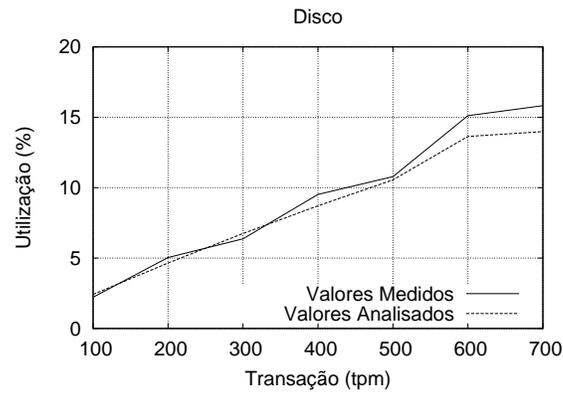
A Figura 5.11(b) descreve uma comparação similar à anterior para os níveis de utilização da infraestrutura de armazenamento. Os valores apresentados são obtidos através do processo de medição e pela avaliação do modelo de desempenho proposto por meio da métrica de desempenho $UDisk = P\{\#Disk = Nd\}$. O resultado apresenta um erro máximo relativo de 11,65% e um erro médio relativo de 6,47%.

A Figura 5.11(c) apresenta uma comparação entre os valores medidos da vazão da infraestrutura de armazenamento e valores obtidos através do modelo de desempenho proposto por meio da métrica de desempenho $Vazao = P\{\#P12 \geq 0\}/(\mu_2)$. O resultado apresenta um erro máximo relativo de 13,67% e um erro médio relativo de 6,70%. A Tabela 5.3 explica as métricas de desempenho.

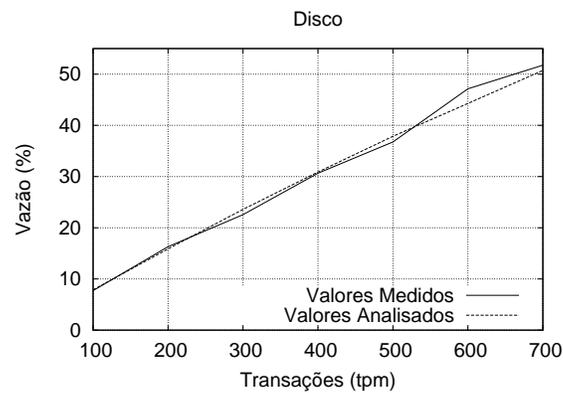
Os resultados obtidos com o modelo e as respectivas medições do sistema foram comparados por meio do Teste T-emparelhado [Lil00]. Pôde-se constatar com 95% de grau de



(a) Utilização do Processador



(b) Utilização do Disco



(c) Vazão do Disco

Figura 5.11: Validação do Modelo de Desempenho de Sistemas TEF

confiança, que os resultados não evidenciam qualquer discrepância entre os dados medidos e os valores obtidos do modelo.

O modelo de desempenho de sistemas TEF é uma abstração do sistema TEF conhecido como SCOPE. O nível de detalhamento desse modelo considera os aspectos de processa-

Tabela 5.3: Métricas de Desempenho

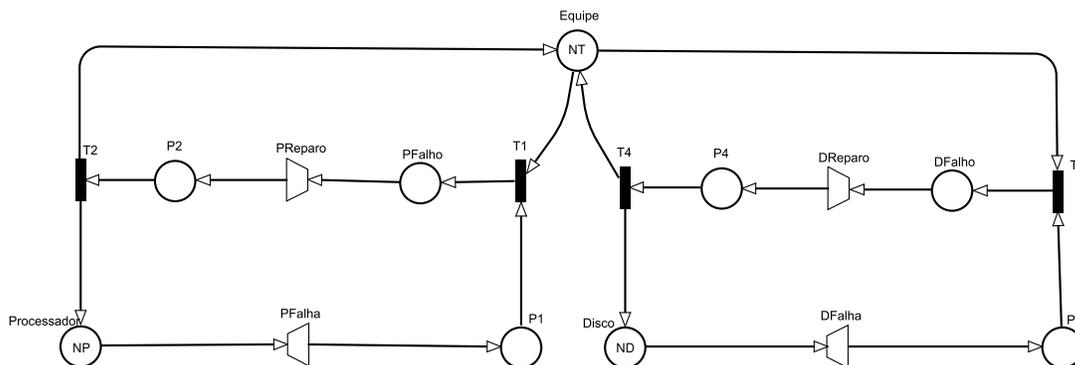
Métrica	Significado
$UProc = P\{\#Processador = 0\}$	O valor calculado é a probabilidade de o lugar Processador não ter marcações.
$UDisco = P\{\#Disco = 0\}$	O valor calculado é a probabilidade de o lugar Disco não ter marcações.
$DiscoVazao = P\{\#P12 \geq 0\}/(\mu_2)$	O valor calculado está associado à transição $T13$ (A taxa de disparo da transição $T13$ depende da marcação no lugar $P12$, este valor é o inverso do tempo associado à transição $T13$).

mento e armazenamento do servidor TEF relacionados às transações comerciais de crédito e débito. O processo de validação é realizado por meio da comparação dos resultados das métricas de desempenho obtidas através do modelo de desempenho de sistemas TEF proposto e dos resultados do processo de medição do servidor TEF. Após a validação do modelo de desempenho de sistemas TEF proposto, ele poderá ser usado como uma ferramenta para estimar índices de desempenho de um sistema TEF considerando várias condições e configurações [MA05].

Os resultados da análise qualitativa do modelo de desempenho de sistemas TEF utilizando a ferramenta INA [INA09] são: alcançabilidade (*reachability*), limitada (*bounded*), segura (*safe*), viva (*live*), cobertura (*coverability*) e ausência de *deadlocks*.

5.4 MODELO DE DEPENDABILIDADE DE SISTEMAS TEF

Esta seção apresenta o modelo SPN concebido para avaliação de dependabilidade de sistemas TEF conforme mostrado na Figura 5.12 [SMAC09].

**Figura 5.12:** Modelo de Dependabilidade de Sistemas TEF

O modelo proposto é composto de sub-modelos (sub-redes que descrevem a infraestr-

tura do sistema TEF). As sub-redes Processamento de Transações e Armazenamento de Transações representam os eventos de falhas e as atividades de reparo relacionadas à infraestrutura de processamento e armazenamento do servidor TEF.

A marcação NP atribuída ao lugar Processador representa o recurso processador em estado operacional, e a marcação NP atribuída ao lugar $PFalho$ representa eventos de falha do recurso processador. A transição estocástica genérica $PFalha$ (*s-transição*) representa a distribuição empírica do MTTF (tempo médio de falhas), ou seja, o tempo médio das ocorrências de eventos de falhas. A transição estocástica genérica $PReparo$ (*s-transição*) representa a distribuição empírica do MTTR (tempo médio de reparos), ou seja, o tempo médio das atividades de reparo. A Figura 5.13 descreve a sub-rede Processamento de Transações.

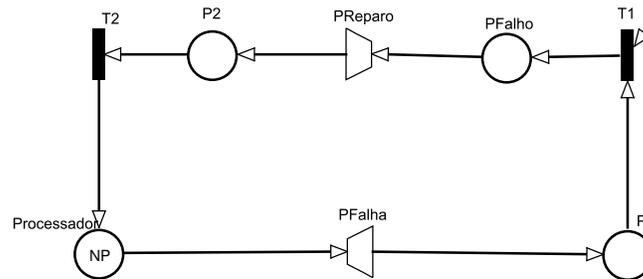


Figura 5.13: Sub-rede Processamento de Transações

De forma similar, a marcação ND atribuída ao lugar Disco representa o recurso disco em estado operacional e a marcação ND atribuída ao lugar $DFalho$ representa os eventos de falha do recurso disco. A transição estocástica genérica $DFalha$ (*s-transição*) representa a distribuição empírica do MTTF (tempo médio de falhas), ou seja, o tempo médio das ocorrências de eventos de falhas. A transição estocástica genérica $DReparo$ (*s-transição*) representa a distribuição empírica do MTTR (tempo médio de reparos), ou seja, o tempo médio das atividades de reparo. A Figura 5.14 descreve a sub-rede Armazenamento de Transações.

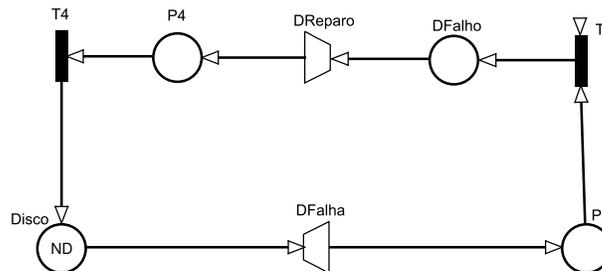


Figura 5.14: Sub-rede Armazenamento de Transações

A marcação NT atribuída ao lugar Equipe (ver Figura 5.15) representa o número de equipes disponíveis para realização de manutenções.

**Figura 5.15:** Lugar Equipe

5.4.1 Refinamento do Modelo de Dependabilidade de Sistemas TEF Considerando Variações de Temperatura

Esta seção apresenta um estudo de caso para refinar o modelo de dependabilidade proposto. Este estudo de caso considera o servidor MX221 [Ita09a] onde estão configurados os aplicativos de gerenciamento para obtenção do MTTF (tempo médio para falhas).

Os tempos médios para falhas do processador e do disco são estimados para o servidor TEF (MX221) conforme a Equação (2.11) apresentada no Capítulo 2. As Tabelas 5.4 e 5.5 mostram os tempos médios para falhas do processador e do disco obtidos por meio da Equação (2.11).

Tabela 5.4: MTTFs do Processador

Temperatura (°C)	Tempo Médio para Falhas (s)
20	17520,0000
25	12388,5108
30	8760,0000
35	6194,2554
40	4380,0000

Tabela 5.5: MTTFs do Disco

Temperatura (°C)	Tempo Médio para Falhas (s)
20	43800,0000
25	30971,2770
30	21900,0000
35	15485,6385
40	10950,0000

Os dados obtidos (distribuição empírica) foram analisados para decidir qual a distribuição exponencial que melhor se adapta aos tempos médios para falhas do processador e do disco (representadas através das transições estocásticas genéricas *PFalha* e *DFalha*). A distribuição exponencial adequada aos dados obtidos foi escolhida com base nos cálculos dos valores das médias (μ_D) e desvios padrões (σ_D) dos tempos médios para falhas do processador e do disco conforme o processo descrito no Capítulo 3 [DAJ95]. Estas transições foram refinadas de acordo com os resultados apresentados na Tabela 5.6.

Tabela 5.6: Média e Desvio Padrão

Recursos Computacionais	μ_D (s)	σ_D (s)	Distribuição Exponencial
Processador	11215,69	4911,38	Hipoexponencial
Disco	28039,22	12278,45	Hipoexponencial

Após a definição da distribuição expolinomial adequada para representar os dados obtidos, os parâmetros dessa distribuição são calculados. Com a escolha da distribuição hipoexponencial para refinar as transições estocásticas genéricas $PFalha$ e $DFalha$, os parâmetros μ_1 , μ_2 e γ devem ser calculados. Esses parâmetros são calculados utilizando as Equações (3.10), (3.13) e (3.14). A Tabela 5.7 mostra os valores dos parâmetros μ_1 , μ_2 e γ para as distribuições hipoexponenciais que refinarão as transições estocásticas genéricas $PFalha$ e $DFalha$.

Tabela 5.7: Parâmetros das Distribuições Hipoexponenciais

Transição Estocástica	μ_1 (s)	μ_2 (s)	γ
P_{Falha}	247,45	10968,23	5
D_{Falha}	618,63	27420,59	5

Assim, uma versão refinada do modelo de dependabilidade de sistemas TEF é gerada. O modelo de dependabilidade de sistemas TEF refinado é obtido para o estudo de caso proposto. O número de marcas atribuídas ao lugar Processador representa o número de processadores em estado operacional processando as transações comerciais. As transições temporizadas $PFalha1$ e $PFalha2$ da sub-rede Processamento de Transações representam o tempo médio para falhas do processador. Da mesma forma, o número de marcas designadas ao lugar Disco representa o número de discos em estado operacional realizando operações de armazenamento relacionadas às transações comerciais. As transições temporizadas $DFalha1$ e $DFalha2$ da sub-rede Armazenamento de Transações representam o tempo médio para falhas do disco. A Figura 5.16 descreve as sub-redes Processamento de Transações e Armazenamento de Transações Refinadas.

5.4.2 Refinamento do Modelo de Dependabilidade de Sistemas TEF Considerando uma Política de Manutenção

Esta seção apresenta uma política de manutenção para refinar o modelo de dependabilidade proposto [SMA09a].

A política de manutenção de sistemas TEF (ESMP) descreve os procedimentos a serem adotados para minimizar as ocorrências de eventos de falhas na infraestrutura de sistemas TEF garantindo a qualidade do serviço oferecido.

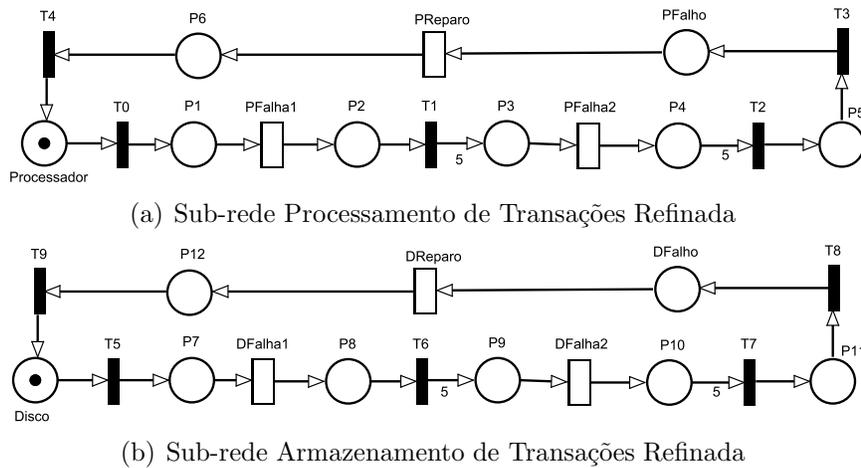


Figura 5.16: Sub-redes de Processamento e Armazenamento de Transações Refinadas

Na ESMP, a manutenção corretiva será realizada após a ocorrência de eventos de falhas e pode resultar em uma atividade de reparo ou na substituição do equipamento. A manutenção preventiva segue os seguintes procedimentos: a primeira atividade de manutenção preventiva será realizada quando as métricas de dependabilidade (disponibilidade e confiabilidade) do sistema TEF atingirem os níveis críticos especificados. Após a primeira atividade de manutenção preventiva, os intervalos entre as manutenções são fixados kT ($k = 1, 2, \dots, N$). T representa o período entre cada manutenção preventiva e N , o número de manutenções preventivas que serão realizadas antes da substituição do equipamento. A substituição do equipamento ocorrerá quando o número de intervalos entre manutenções sequenciais atingir kT ($k = 1, 2, \dots, N + 1$). Nas primeiras N manutenções, o equipamento será removido para a identificação dos dispositivos defeituosos e os reparos evitarão as ocorrências de eventos de falhas. Na $N + 1$ manutenção, o dispositivo deve ser substituído e o processo será reiniciado.

As equipes locais (LT's) e equipes especialistas (ST's) são alocadas para realização das atividades de manutenção e substituição dos equipamentos e a equipe de planejamento de manutenção (MPT) é alocada para o gerenciamento das atividades de manutenção.

As Equipes Locais (LT's) realizarão as manutenções corretivas e as manutenções preventivas na empresa. Essas equipes têm competência para realizar vários tipos de atividades de manutenção em diferentes equipamentos na empresa. Dessa forma, essas equipes não são especializadas em nenhum equipamento. As LT's realizam o controle das atividades de manutenção e substituição dos equipamentos. Essas equipes também realizam o registro do histórico de eventos de falhas, reparos e substituições dos equipamentos.

As Equipes Especialistas (ST's) são chamadas quando os procedimentos são insuficientes para os diagnósticos, reparos ou substituições. As ST's também visitam as empresas, a fim de realizar as análises periódicas mais complexas ou críticas.

A Equipe de Planejamento de Manutenção (MPT) é responsável pelo gerenciamento das informações registradas pelas LT's. Baseadas nessas informações, as MPT's ajustam

os intervalos entre manutenções e o número máximo de reparos antes da substituição de cada dispositivo. O resultado desse planejamento é o controle dos níveis de desempenho e de dependabilidade do sistema TEF.

O tempo médio de reparo (MTTR) dos equipamentos depende da equipe que será responsável pela manutenção. O MTTR mede o tempo entre a interrupção do serviço e a restauração do serviço. O tempo médio de reparo ou substituição da LT (MTFPR) é composto pelo tempo médio de percepção da falha e pelo tempo médio de reparo. O tempo médio de reparo ou substituição da ST (MTFPRA) é composto pelo tempo médio de percepção da falha, pelo tempo médio de reparo e pelo tempo médio de chegada da equipe ao local. Quando a manutenção é realizada pela ST, o MTFPRA é influenciado pela localização da equipe ou dificuldades no diagnóstico. Quando as mudanças são bem documentadas e gerenciadas, o tempo necessário para o diagnóstico correto do problema é afetado e, conseqüentemente, o MTTR é reduzido.

O modelo de manutenção de sistemas TEF foi concebido conforme os procedimentos da política de manutenção de sistemas TEF (ESMP). A marcação NT atribuída ao lugar Equipe representa tipos distintos de equipes de manutenção. Equipes de manutenção com diferentes níveis de qualificação são representadas com lugares específicos e suas marcações iniciais mostram a disponibilidade dessas equipes que têm competências e especialidades distintas. O número de marcações atribuídas ao lugar Equipe representa o grau de concorrência relacionado ao processo de reparo. A ESMP habilita as manutenções corretivas através da função de habilitação $\{\#P4 > 0\}$ na transição imediata $T5$ e da função de habilitação $\{\#P9 > 0\}$ na transição imediata $T12$. De forma semelhante, as manutenções preventivas são habilitadas por meio da função de habilitação $\{\#P0 > 0\}$ nas transições imediatas $T0, T2, T3, T5, T7, T9, T10$ e $T12$. A ESMP é adotada para dar maior prioridade ao processo de reparo da infraestrutura de processamento em relação à infraestrutura de armazenamento. Essa prioridade é motivada devido à infraestrutura de armazenamento ter um número maior de dispositivos que a infraestrutura de processamento, permitindo períodos longos de manutenção. Para representar essa prioridade no processo de reparo da infraestrutura de processamento em relação à infraestrutura de armazenamento, pesos são atribuídos às transições imediatas $T5$ e $T12$ [ZK07], e o peso com maior valor é atribuído à transição imediata $T5$. A Figura 5.17 descreve o modelo de dependabilidade considerando a política de manutenção de sistemas TEF.

A sub-rede Manutenção (ver Figura 5.18) representa a política de manutenção adotada. A marcação atribuída ao lugar Manutenção representa o número de manutenções que devem ser realizadas antes da substituição do equipamento. As transições estocásticas genéricas $MTBM$ e $MTBR$ representam os tempos médios entre manutenções preventivas e os tempos médios entre substituições dos equipamentos, respectivamente. A função de habilitação $\{\#P0 > 0\}$ [ZK07] é atribuída às transições imediatas $T0, T2, T3, T5, T7, T9, T10$ e $T12$ para representar os intervalos entre manutenções preventivas especificados na política de manutenção.

Os custos anuais de manutenção são obtidos por meio da Equação (5.1). N_i representa o número de equipes de um tipo específico realizando atividades de manutenção.

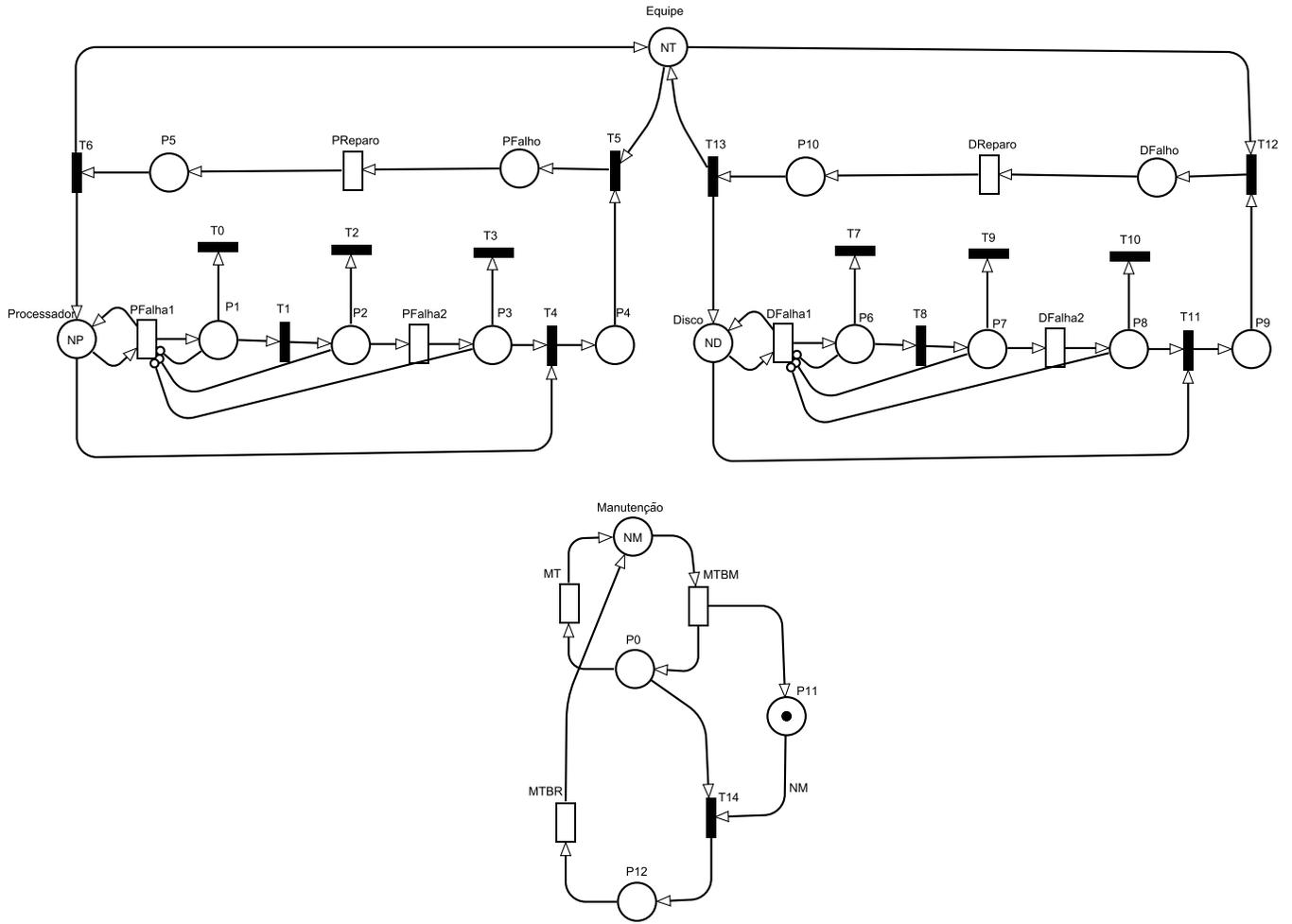


Figura 5.17: Modelo de Manutenção

$C_i(Equipe_i)$ representa o custo da manutenção por horas de trabalho de um tipo específico de equipe de manutenção. $Equipe_i$ é lugar que representa um tipo específico de equipe de manutenção. RA_j é a transição temporizada que representa as atividades de manutenção. $TP(RA_j)$ representa a vazão da transição RA_j . $C_j(RA_j)$ representa o custo do material utilizado nas atividades de manutenção. T representa o período de cálculo do custo da manutenção.

$$C_a = [(\sum_{i=1}^m N_i \times C_i(Equipe_i)) + (\sum_{j=1}^n TP(RA_j) \times C_j(RA_j)) \times T] \quad (5.1)$$

As métricas disponibilidade e confiabilidade dos recursos de processamento e armazenamento são descritas através das seguintes expressões: $UPProc = P\{\#Processador = 1\}$ e $UPDisco = P\{\#Disco = 1\}$. Essas expressões são computadas por meio da ferramenta *TimeNET* 4.0 [ZK07]. A Tabela 5.8 explica as métricas de dependabilidade.

Os resultados da análise qualitativa do modelo de dependabilidade de sistemas TEF

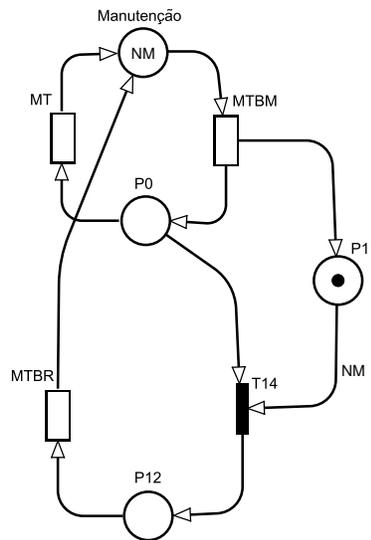


Figura 5.18: Sub-rede Manutenção

Tabela 5.8: Métricas de Dependabilidade

Métrica	Significado
$UPProc = P\{\#Processador = 1\}$	O valor computado é a probabilidade de o lugar Processador ter uma marcação.
$UPDisco = P\{\#Disco = 1\}$	O valor computado é a probabilidade de o lugar Disco ter uma marcação.

utilizando a ferramenta INA [INA09] são: alcançabilidade (*reachability*), limitada (*bounded*), segura (*safe*), viva (*live*) e ausência de *deadlocks*.

5.5 AVALIAÇÃO DE PERFORMABILIDADE

A avaliação de performabilidade de sistemas TEF é realizada através do modelo para avaliação de desempenho e do modelo para avaliação de dependabilidade de sistemas TEF. As métricas de desempenho e de dependabilidade são calculadas independentemente através dos modelos de desempenho e dependabilidade, em seguida são combinadas para mostrar o efeito da dependabilidade no desempenho.

5.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou a aplicação da metodologia descrita no Capítulo 4 para modelagem e refinamento dos modelos de desempenho e dependabilidade de sistemas TEF. A partir da especificação do sistema TEF e seus componentes, são obtidos os modelos

abstratos. Esses modelos foram obtidos por meio do método para avaliação de desempenho do sistema TEF e do método para avaliação de dependabilidade do sistema TEF. Posteriormente, definiu-se um ambiente de medição do sistema TEF a ser avaliado, para que o modelo de desempenho de sistemas TEF seja refinado e modelos mais precisos sejam gerados. O refinamento do modelo de dependabilidade de sistemas TEF foi realizado através de uma política de manutenção e por meio de dados obtidos através da Equação (2.11).

Estudo de Caso

ESTUDO DE CASO

Este capítulo apresenta um cenário para avaliação do impacto das variações da carga de trabalho no desempenho. Posteriormente, um conjunto de cenários são apresentados para avaliação do impacto de uma política de manutenção (ESMP) e dos efeitos das variações de temperatura na dependabilidade. Finalmente, os resultados das avaliações de desempenho e dependabilidade são combinados para obtenção dos resultados de performabilidade de acordo com a metodologia de performabilidade adotada.

6.1 ESTUDO DE CASO DE DESEMPENHO

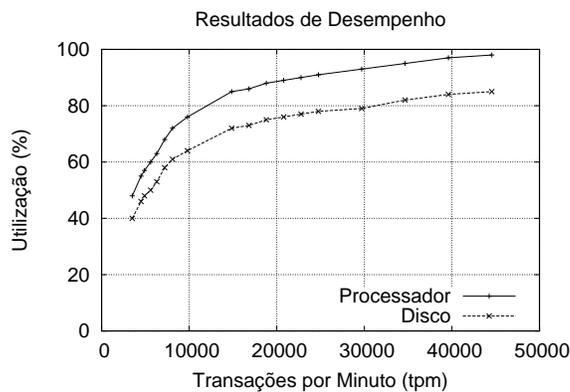
O sistema SCOPE descrito no Capítulo 5 foi adotado para ilustrar a metodologia de avaliação de performabilidade e os modelos propostos no Capítulo 4. O estudo de caso de desempenho e planejamento de capacidade descreve um estabelecimento comercial com diversos pontos de vendas registrando operações comerciais de crédito e débito. As taxas de transações adotadas nesse estudo de caso foram 3500, 4500, 4900, 5600, 6300, 7200, 8100, 9801, 14850, 16830, 18810, 20790, 22770, 24750, 29700, 34650, 39600 e 44550 tpms (transações por minuto). Essas taxas foram adotadas com o objetivo de encontrar os níveis de utilização que indicam a saturação das infraestruturas de processamento e armazenamento do sistema SCOPE. A Tabela 6.1 mostra o impacto da variação das frequências de envio de transações comerciais na utilização da infraestrutura de processamento e armazenamento do servidor TEF.

Quando o servidor TEF é submetido à frequência 16830 tpms, o nível de utilização do processador é 86% e o nível de utilização do disco é 73%. Esses resultados mostram que o nível de utilização do processador ultrapassa 85%, assim, este recurso deve ser cuidadosamente avaliado para determinação das possíveis causas desse alto nível de utilização. Um processador com um alto nível de utilização pode não responder a todas as requisições da carga de trabalho (transações comerciais), assim, um crescimento na demanda da carga de trabalho resultaria na degradação do nível de serviço oferecido pelo sistema TEF [AAB⁺03]. A Figura 6.1 representa as variações das utilizações do processador e disco considerando diferentes frequências de envio de transações comerciais. Esses resultados mostram o efeito de flutuações da carga de trabalho nos níveis de utilização do processador e disco considerando cada frequência de transação comercial. A utilização do processador e disco são obtidos por meio das seguintes expressões: $UProc = (P\{\#Processor = 0\})$ e $UDisk = (P\{\#Disk = 0\})$ [ZK07], respectivamente, conforme mostradas no Capítulo 5.

Com o objetivo de evitar a degradação no nível de serviço oferecido pelo sistema TEF causada por eventuais picos e flutuações da carga de trabalho e para prevenir a saturação,

Tabela 6.1: Utilização dos Recursos

Frequência (tpms)	Processador	Disco
3500	48%	40%
4500	55%	46%
4900	57%	48%
5600	60%	50%
6300	63%	53%
7200	68%	58%
8100	72%	61%
9801	76%	64%
14850	85%	72%
16830	86%	73%
18810	88%	75%
20790	89%	76%
22770	90%	77%
24750	91%	78%
29700	93%	79%
34650	95%	82%
39600	97%	84%
44550	98%	85%

**Figura 6.1:** Utilização do Processador e do Disco

uma nova infraestrutura de processamento foi adicionada ao servidor TEF. Adicionando essa nova infraestrutura de processamento, o nível de utilização do processador é 42,67%, considerando a frequência 16830 tpm. Dessa forma, o emprego de duas infraestruturas de processamento reduz drasticamente a utilização e previne eventuais gargalos que podem ocorrer devido a oscilações na carga de trabalho. As duas infraestruturas de processamento representam o grau de concorrência relacionado ao processamento das transações comerciais de crédito e débito, ou seja, o número de transações comerciais que podem ser simultaneamente processadas pelas aplicações de gerenciamento.

6.2 ESTUDOS DE CASO DE DEPENDABILIDADE

Os estudos de caso 1, 2 e 3 têm o objetivo de avaliar o impacto da política de manutenção (ESMP) na dependabilidade da infraestrutura de processamento e armazenamento de sistemas TEF. O Estudo de caso 1 adota a ESMP com kT ($k = 1, 2, \dots, 10$) onde $T = 4380$ horas. O Estudo de caso 2 adota a ESMP com kT ($k = 1, 2, \dots, 10$) onde $T = 1752$ horas. O Estudo de caso 3 adota a ESMP com kT ($k = 1, 2, \dots, 10$) onde $T = 876$ horas. O tempo médio de reparo (MTTR) da infraestrutura de processamento do servidor TEF é 16 horas quando uma equipe (composta possivelmente por uma pessoa) é alocada para a execução das atividades de manutenção. De forma similar, o tempo médio de reparo (MTTR) da infraestrutura de armazenamento do servidor TEF é 18 horas quando uma equipe (composta possivelmente por uma pessoa) é alocada para a execução das atividades de manutenção. O tempo médio de falha (MTTF) da infraestrutura de processamento do servidor TEF é 17520 horas, e o tempo médio de falha (MTTF) da infraestrutura de armazenamento do servidor TEF é 43800 horas. As Tabelas 6.2, 6.3 e 6.4 mostram os parâmetros dos Estudos de caso 1, 2 e 3, respectivamente. Os MTTFs do processador e disco são baseados nos tempos médios entre falhas (MTBF) desses recursos [Ita09a]. Se as infraestruturas de processamento e armazenamento falharem simultaneamente, a equipe de manutenção deverá levar em consideração os requisitos da política de manutenção adotada.

Tabela 6.2: Estudo de Caso 1

Recurso	MTBM (hs)	MTBR (hs)	$MTTR$ (hs)	$MTTF$ (hs)
Processador	4380	43800	16	17520
Disco	4380	43800	18	43800

Tabela 6.3: Estudo de Caso 2

Recurso	MTBM (hs)	MTBR (hs)	$MTTR$ (hs)	$MTTF$ (hs)
Processador	1752	17520	16	17520
Disco	1752	17520	18	43800

Tabela 6.4: Estudo de Caso 3

Recurso	MTBM (hs)	MTBR (hs)	$MTTR$ (hs)	$MTTF$ (hs)
Processador	876	8760	16	17520
Disco	876	8760	18	43800

As Figuras 6.2, 6.3 e 6.4 descrevem os resultados de dependabilidade (disponibilidade e confiabilidade) dos Estudos de caso 1, 2 e 3 obtidos através de avaliação transiente. Essas figuras apresentam a análise da disponibilidade e da confiabilidade das unidades de processamento e de armazenamento do servidor TEF considerando uma equipe de

manutenção realizando atividades de manutenção e substituição de acordo com a ESMP. Essas métricas de dependabilidade foram obtidas durante um período de 43800 horas. Esses resultados são obtidos por meio das expressões $UPProc = (P\{\#Processor = 1\})$ e $UPDisk = (P\{\#Disk = 1\})$ [ZK07], respectivamente, conforme apresentadas no Capítulo 5.

As Figuras 6.2(a), 6.3(a) e 6.4(a) mostram o impacto das ocorrências de eventos de falhas e das atividades de manutenção corretiva e preventiva nos resultados da disponibilidade. As Figuras 6.2(b), 6.3(b) e 6.4(b) representam o impacto dos intervalos entre manutenções e intervalos entre substituições da política de manutenção adotada nos resultados da confiabilidade.

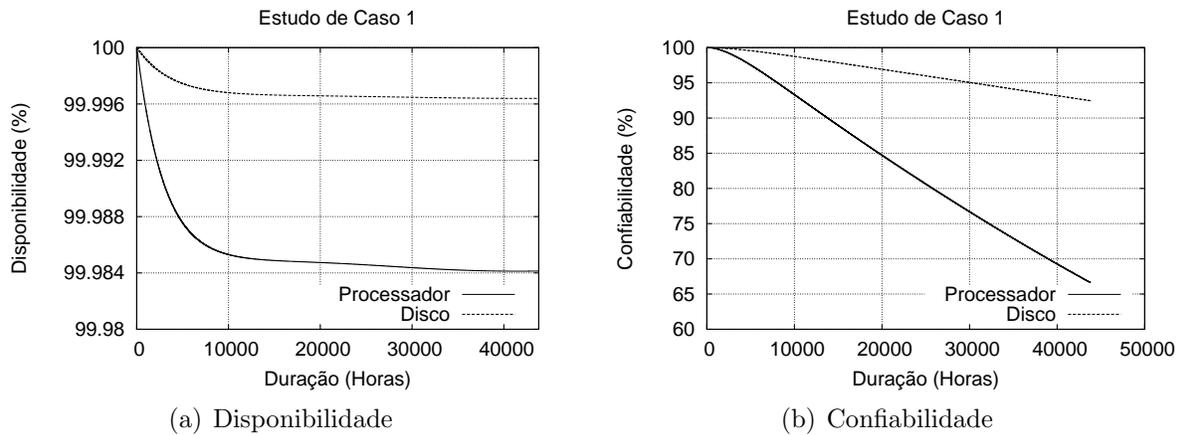


Figura 6.2: Estudo de Caso 1

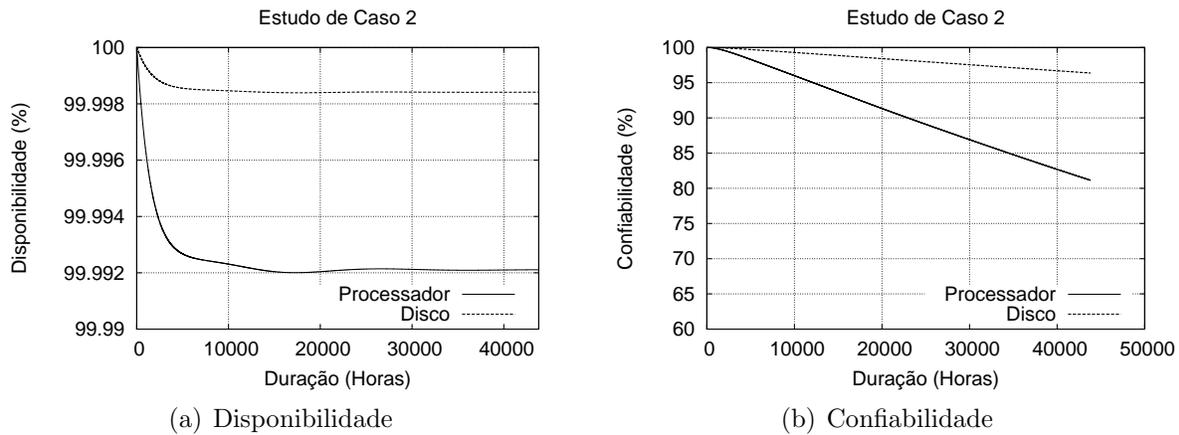


Figura 6.3: Estudo de Caso 2

Os resultados de disponibilidade apresentados nas Figuras 6.2(a), 6.3(a) e 6.4(a) mostram uma redução da disponibilidade antes dos primeiros intervalos entre manutenções preventivas (MTBM) de cada estudo de caso e uma estabilização da disponibilidade após o *MTBM*. Essa redução da disponibilidade é ocasionada por eventos de falhas.

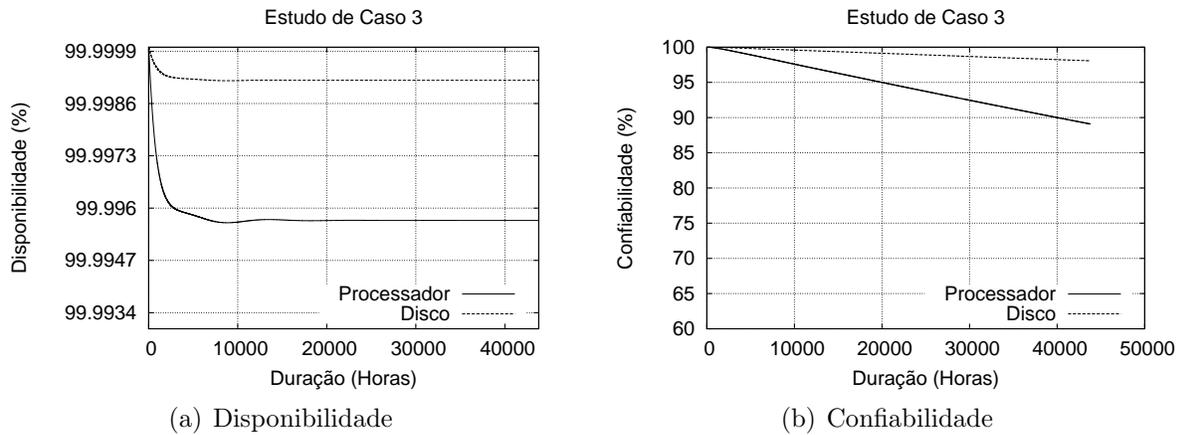


Figura 6.4: Estudo de Caso 3

Antes do MTBM, as manutenções serão realizadas após a ocorrência de um evento de falha (manutenções corretivas). Após o primeiro MTBM, sucessivas manutenções preventivas serão realizadas e, conseqüentemente, a disponibilidade instantânea tenderá à estabilização. Os resultados de confiabilidade apresentados nas Figuras 6.2(b), 6.3(b) e 6.4(b) mostram uma redução da confiabilidade proporcional ao aumento do intervalo entre manutenções preventivas (MTBM) (Ver Figura 6.5). A redução do *MTBM* proporciona uma identificação mais rápida de componentes defeituosos, evitando a ocorrência de eventos de falhas. Assim, intervalos mais curtos entre manutenções preventivas proporcionam uma menor degradação física dos equipamentos, retardando a substituição. Esse fator aumenta a disponibilidade e confiabilidade da infraestrutura do sistema TEF. Isso explica a alta disponibilidade e confiabilidade do Estudo de caso 3.

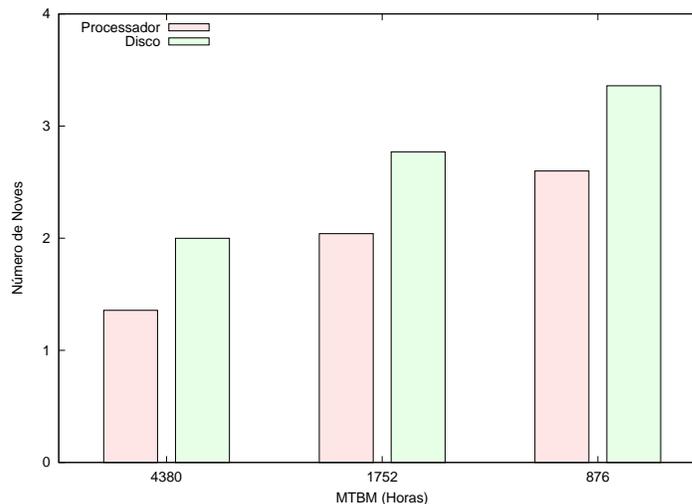


Figura 6.5: Confiabilidade X MTBM

As Tabelas 6.5 e 6.6 mostram a disponibilidade e confiabilidade dos Estudos de casos 1, 2 e 3, respectivamente. Esses valores de disponibilidade e confiabilidade são obtidos

por meio de análise estacionária [ZK07], e os resultados mostram o efeito da política de manutenção adotada (ESMP) nos três estudos de caso. Os altos níveis de disponibilidade e confiabilidade do Estudo de caso 3 reafirmam o fato de que um menor intervalo entre manutenções preventivas (MTBM) aumenta a disponibilidade e a confiabilidade da infraestrutura do sistema TEF.

Tabela 6.5: Disponibilidade

Recurso	Estudo de Caso 1	Estudo de Caso 2	Estudo de Caso 3
Processador	99,9842957438082%	99,9921175742861%	99,995709644024%
Disco	99,9964365019902%	99,9984163505058%	99,9991811218724%

Tabela 6.6: Confiabilidade

Recurso	Estudo de Caso 1	Estudo de Caso 2	Estudo de Caso 3
Processador	95,6030244122151%	99,0883678225921%	99,7486448713673%
Disco	98,9979643140155%	99,8301744691485%	99,9562723622485%

A Tabela 6.7 mostra o custo anual da manutenção para os Estudos de caso 1, 2 e 3. Estes resultados são baseados na Equação (5.1) e nos parâmetros $C_i(Equipe_{LT})=R\$100,00$ para a equipe de manutenção local e $C_j(RA)=R\$3300,00$ para o servidor TEF. O custo mais alto da manutenção anual do Estudo de caso 3 é compensado pelos altos níveis de disponibilidade e confiabilidade oferecidos.

Tabela 6.7: Custo Anual da Manutenção

Estudos de Caso	Custo da Manutenção (R\$)
1	R\$ 939,31
2	R\$ 1164,80
3	R\$ 1262,16

As Figuras 6.6(a) e 6.6(b) mostram a relação entre o número de nove da disponibilidade e confiabilidade com os custos da manutenção, respectivamente, para os Estudos de casos 1, 2 e 3.

6.2.1 Estudos de Caso de Dependabilidade com Temperatura

Os estudos de caso 4, 5, 6, 7 e 8 têm o objetivo de avaliar o efeito de variações de temperatura na dependabilidade da infraestrutura de processamento e armazenamento de sistemas TEF. Esses estudos de caso analisaram a disponibilidade e a confiabilidade do sistema SCOPE considerando diferentes temperaturas, conforme a Equação (2.11). Os tempos médios de falha (MTTF) da infraestrutura de processamento e da infraestrutura de armazenamento variam proporcionalmente às variações na temperatura. O tempo

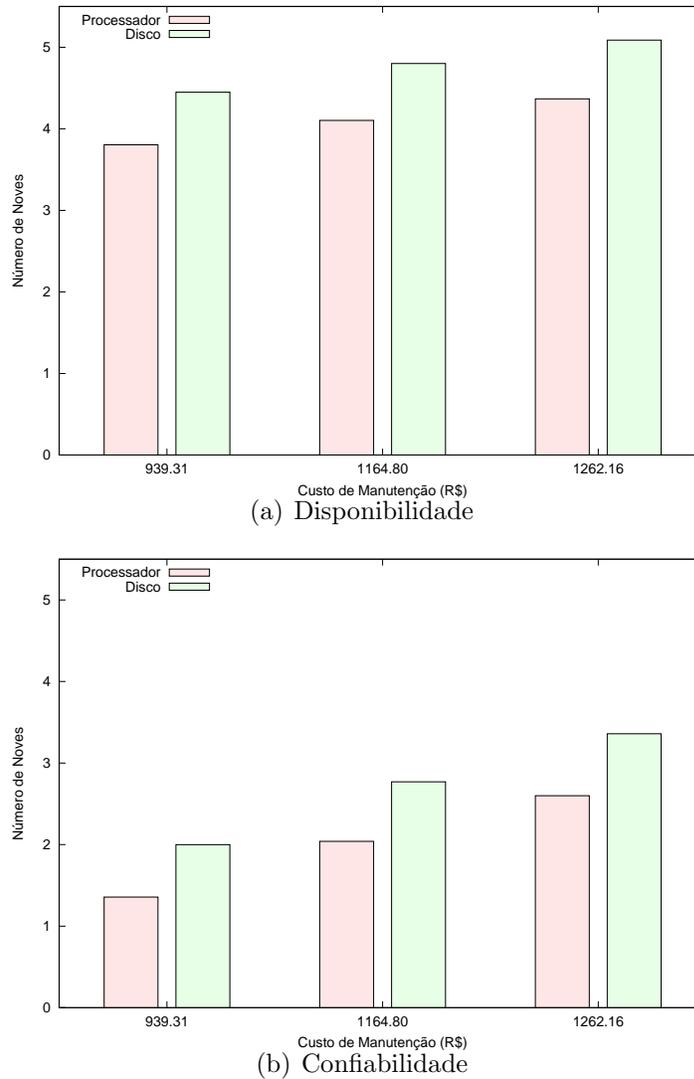


Figura 6.6: Número de Naves X Custo de Manutenção

médio de reparo (MTTR) da infraestrutura de processamento do servidor TEF é 16 horas quando uma equipe (composta possivelmente por uma pessoa) é alocada para a execução das atividades de manutenção. De forma similar, o tempo médio de reparo (MTTR) da infraestrutura de armazenamento do servidor TEF é 18 horas quando uma equipe (composta possivelmente por uma pessoa) é alocada para a execução das atividades de manutenção. As Tabelas 6.8, 6.9, 6.10, 6.11 e 6.14 mostram os parâmetros dos Estudos de caso 4, 5, 6, 7 e 8.

As Figuras 6.7, 6.8, 6.9 e 6.10 mostram os resultados de dependabilidade dos Estudos de caso 4, 5, 6 e 7 obtidos através de análise transiente. Essas figuras apresentam a análise da disponibilidade e da confiabilidade das infraestruturas de processamento e de armazenamento do servidor TEF considerando um período de 1000 horas para a análise da disponibilidade e 43800 horas para a análise da confiabilidade. Esses resultados

Tabela 6.8: Estudo de Caso 4

Recurso	<i>MTTR</i> (hs)	<i>MTTF</i> (hs)	Temperatura (°C)
Processador	16	17520	20
Disco	18	43800	20

Tabela 6.9: Estudo de Caso 5

Recurso	<i>MTTR</i> (hs)	<i>MTTF</i> (hs)	Temperatura (°C)
Processador	16	12388,51081	25
Disco	18	30971,27702	25

Tabela 6.10: Estudo de Caso 6

Recurso	<i>MTTR</i> (hs)	<i>MTTF</i> (hs)	Temperatura (°C)
Processador	16	8760	30
Disco	18	21900	30

Tabela 6.11: Estudo de Caso 7

Recurso	<i>MTTR</i> (hs)	<i>MTTF</i> (hs)	Temperatura (°C)
Processador	16	6194,255403	35
Disco	18	15485,63851	35

são obtidos através das expressões $UPProc = (P\{\#Processor = 1\})$ e $UPDisk = (P\{\#Disk = 1\})$ [ZK07], respectivamente, conforme apresentadas no Capítulo 5.

As Figuras 6.7(a), 6.8(a), 6.9(a) e 6.10(a) descrevem o impacto das ocorrências de eventos de falhas e das atividades de reparo nos resultados da disponibilidade. As Figuras 6.7(b), 6.8(b), 6.9(b) e 6.10(b) representam o efeito das ocorrências de eventos de falhas nos resultados da confiabilidade.

As Figuras 6.7(a), 6.8(a), 6.9(a) e 6.10(a) mostram uma redução da disponibilidade proporcional ao aumento da temperatura. Um aumento na temperatura de $10^{\circ}C$ diminui o tempo médio para falha (*MTTF*) pela metade, aumentando assim os eventos de falhas. Esses resultados de disponibilidade representam o impacto do aumento dos eventos de falhas e das atividades de reparo. Os resultados de confiabilidade apresentados nas Figuras 6.7(b), 6.8(b), 6.9(b) e 6.10(b) mostram uma redução da confiabilidade proporcional ao aumento da temperatura. Esses resultados de confiabilidade demonstram que um aumento gradual da temperatura proporciona a degradação dos componentes dos equipamentos, ocasionando uma maior quantidade de eventos de falhas. Esse fator é responsável pela degradação da disponibilidade e da confiabilidade do sistema TEF.

A Tabela 6.12 apresenta a disponibilidade dos Estudos de caso 4, 5, 6 e 7. Esses valores de disponibilidade são obtidos por meio de avaliação estacionária [ZK07] e representam o efeito da variação da temperatura nos quatro estudos de casos. Os altos níveis de

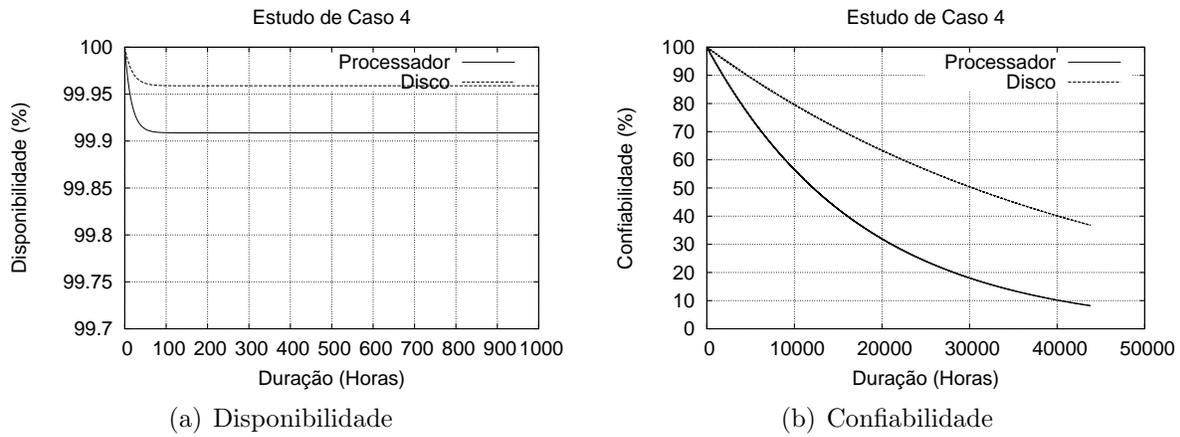


Figura 6.7: Estudo de Caso 4

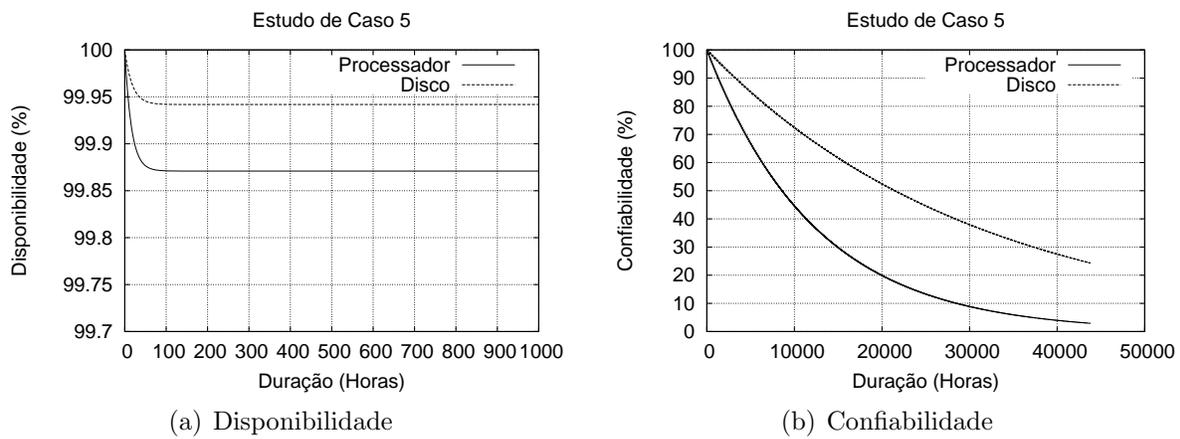


Figura 6.8: Estudo de Caso 5

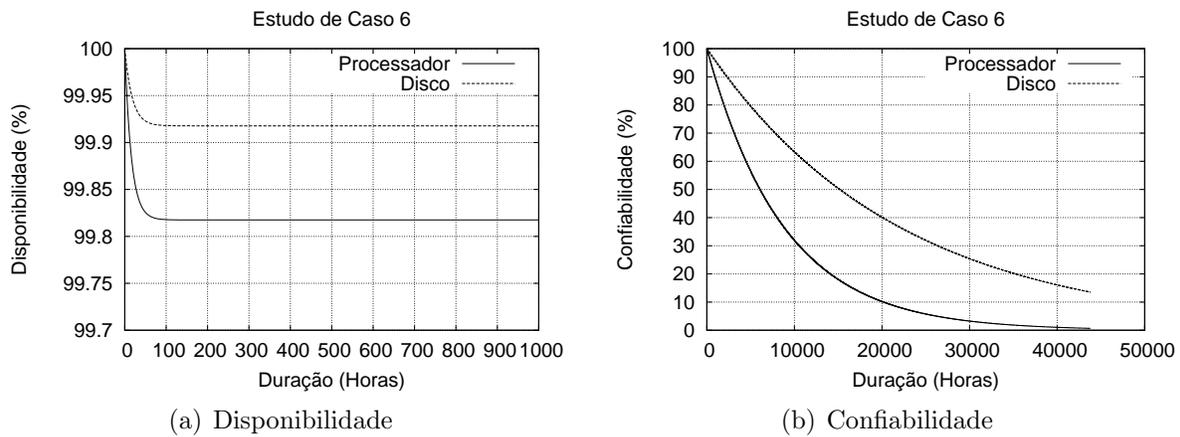
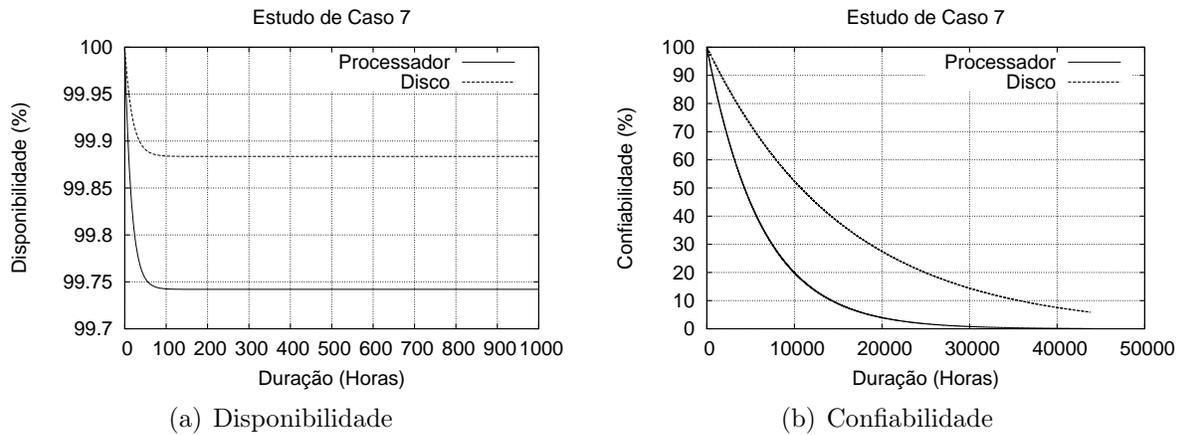


Figura 6.9: Estudo de Caso 6

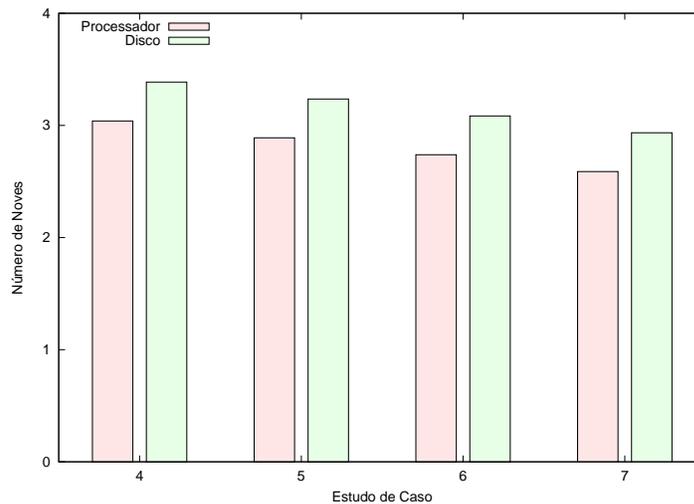
**Figura 6.10:** Estudo de Caso 7

disponibilidade do Estudo de caso 4 demonstram o efeito do aumento da temperatura na disponibilidade da infraestrutura do sistema TEF.

Tabela 6.12: Disponibilidade

Recurso	Estudo de Caso 4	Estudo de Caso 5	Estudo de Caso 6	Estudo de Caso 7
Processador	99,9087170014252%	99,8709305002133%	99,8175164991047%	99,7420261056561%
Disco	99,9588876872446%	99,941848830585%	99,9177427268929%	99,8836326102443%

A Figura 6.11 mostra o número de noves da disponibilidade dos Estudos de casos 4, 5, 6 e 7.

**Figura 6.11:** Número de Noves

O Estudo de caso 8 analisa o efeito do aumento da temperatura na disponibilidade e a influência da política de manutenção adotada (ESMP) na redução desse efeito. A

Tabela 6.13 mostra as disponibilidades dos recursos do servidor TEF quando o ambiente é submetido à temperatura de 20°C, e a Tabela 6.14 mostra as disponibilidades dos recursos do servidor TEF quando o ambiente é submetido à temperatura de 40°C.

Tabela 6.13: Estudo de Caso 4

Recurso	<i>MTTR</i> (hs)	<i>MTTF</i> (hs)	Temperatura (°C)	Disponibilidade	Número de Noves
Processador	16	17520	20	99,9087170014252%	3,04
Disco	18	43800	20	99,9588876872446%	3,39

Tabela 6.14: Estudo de Caso 8

Recurso	<i>MTTR</i> (hs)	<i>MTTF</i> (hs)	Temperatura (°C)	Disponibilidade	Número de Noves
Processador	16	4380	40	99,6353635254142%	2,59
Disco	18	10950	40	99,8353560412057%	2,93

A Tabela 6.14 mostra que a disponibilidade do processador é 99,6353635254142% e do disco é 99,8353560412057%. Esses resultados de disponibilidade representam o efeito dos 40°C de temperatura nos recursos do servidor TEF. Para minimizar esses efeitos e atingir as disponibilidades apresentadas, na Tabela 6.13 será aplicada a política de manutenção adotada (ESMP). Para atingir as disponibilidades apresentadas na Tabela 6.13, considerando uma temperatura de 40°C, será necessário um intervalo entre manutenções preventivas (MTBM) de 219 horas.

6.3 ESTUDOS DE CASO DE PERFORMABILIDADE

Os resultados de performabilidade são obtidos pela combinação dos resultados das métricas de dependabilidade e dos resultados das métricas de desempenho do sistema TEF. Esses resultados são uma combinação dos níveis de disponibilidade e dos níveis de confiabilidade com a utilização do processador e do disco considerando as frequências: 3500, 4500, 4900, 5600, 6300, 7200, 8100, 9801, 14850, 16830, 18810, 20790, 22770, 24750, 29700, 34650, 39600 e 44550 tpms (transações por minuto). Esses resultados de performabilidade apresentam a degradação do desempenho das infraestruturas do servidor TEF para os Estudos de caso 1, 2 e 3.

As Tabelas 6.15 e 6.16 mostram os efeitos das falhas, atividades de manutenção corretiva e atividades de manutenção preventiva na utilização do processador e disco, respectivamente, para os Estudos de caso 1, 2 e 3 considerando uma única equipe de manutenção. Essas tabelas representam os resultados da combinação da disponibilidade e da utilização do processador e do disco considerando um conjunto de frequências de envio de transações comerciais. Nessas tabelas, a primeira coluna apresenta a utilização dos recursos de processamento e armazenamento e as outras colunas representam a redução da utilização desses recursos em função da ocorrência de eventos de falhas e atividades de manutenção.

Tabela 6.15: Utilização do Processador para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Disponibilidade

Frequência (tpms)	Utilização	Estudo de Caso 1	Estudo de Caso 2	Estudo de Caso 3
3500	48,4704100000%	48,4627980826%	48,4665893559%	48,4683304469%
4500	54,8245500000%	54,8159402122%	54,8202284956%	54,8221978316%
4900	57,0853200000%	57,0763551751%	57,0808202921%	57,0828708366%
5600	59,9811500000%	59,9717304065%	59,9764220304%	59,9785765951%
6300	63,3841500000%	63,3741959907%	63,3791537915%	63,3814305943%
7200	68,2340900000%	68,2233743437%	68,2287114985%	68,2311625146%
8100	71,6649500000%	71,6536955527%	71,6593010636%	71,6618753185%
9801	75,9839000000%	75,9719672937%	75,9779106255%	75,9806400202%
14850	84,5334800000%	84,5202046457%	84,5268167112%	84,5298532128%
16830	86,1891700000%	86,1756346319%	86,1823762027%	86,1854721778%
18810	88,0330200000%	88,0191950690%	88,0260808626%	88,0292430701%
20790	89,1646000000%	89,1505973628%	89,1575716666%	89,1607745213%
22770	90,3336500000%	90,3194637722%	90,3265295171%	90,3297743648%
24750	91,4079600000%	91,3936050598%	91,4007548355%	91,4040382731%
29700	92,7834200000%	92,7688490540%	92,7761064158%	92,7794392610%
34650	94,9875900000%	94,9726729055%	94,9801026738%	94,9835146943%
39600	96,7786000000%	96,7634016407%	96,7709714987%	96,7744478536%
44550	97,7030600000%	97,6877164612%	97,6953586289%	97,6988681909%

O Estudo de caso 3 apresenta uma degradação menor na utilização dos recursos que os Estudos de caso 1 e 2.

As Tabelas 6.17 e 6.18 mostram os efeitos dos eventos de falhas e atividades de manutenção preventiva na utilização do processador e disco, respectivamente. Essas tabelas representam os resultados da combinação da confiabilidade e da utilização do processador e do disco considerando uma equipe de manutenção e diferentes taxas de transações comerciais. Os resultados demonstram uma degradação menor à medida que os intervalos entre manutenções diminuem assim o Estudo de caso 3 apresenta os melhores resultados em relação aos Estudos de caso 1 e 2.

As Figuras 6.12(a) e 6.12(b) mostram o nível de degradação da utilização do processador e do disco, respectivamente. Essas figuras mostram os níveis de degradação da utilização dos recursos de processamento e armazenamento quando são comparados os resultados das utilizações obtidas para as taxas das transações comerciais 3500, 4500, 4900, 5600, 6300, 7200, 8100, 9801, 14850, 16830, 18810, 20790, 22770, 24750, 29700, 34650, 39600 e 44550 tpms e as utilizações considerando, os efeitos da disponibilidade dos Estudos de caso 1, 2 e 3

As Figuras 6.13(a) e 6.13(b) mostram o nível de degradação da utilização dos recursos processador e do disco quando são considerados os efeitos dos níveis de confiabilidade dos Estudos de caso 1, 2 e 3.

Tabela 6.16: Utilização do Disco para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Disponibilidade

Frequência (tpms)	Utilização	Estudo de Caso 1	Estudo de Caso 2	Estudo de Caso 3
3500	40,4223700000%	40,4209295496495%	40,4217298513420%	40,4220389900534%
4500	45,8644000000%	45,8627656230188%	45,8636736686614%	45,8640244264600%
4900	47,7533900000%	47,7516883088977%	47,7526337536808%	47,7529989579341%
5600	50,2718300000%	50,2700385643385%	50,2710338704185%	50,2714183349798%
6300	53,3411100000%	53,3392091906068%	53,3402652637813%	53,3406732013172%
7200	57,5340000000%	57,5319497770550%	57,5330888631000%	57,5335288666581%
8100	60,5208200000%	60,5186633417838%	60,5198615623402%	60,5203244082424%
9801	64,2216400000%	64,2193514631367%	64,2206229543230%	64,2211141030369%
14850	71,6090200000%	71,6064682139975%	71,6078859641170%	71,6084336093978%
16830	73,1039200000%	73,1013149432657%	73,1027622901407%	73,1033213679887%
18810	74,8079700000%	74,8053042194779%	74,8067853039615%	74,8073574138960%
20790	75,8357100000%	75,8330075959834%	75,8345090281622%	75,8350889979579%
22770	76,9567700000%	76,9540276470326%	76,9555512745011%	76,9561398178428%
24750	78,0386100000%	78,0358290956858%	78,0373741419475%	78,0379709588916%
29700	79,4415600000%	79,4387291015905%	79,4403019241369%	79,4409094704409%
34650	81,7010500000%	81,6981385847093%	81,6997561417349%	81,7003809679715%
39600	83,7456100000%	83,7426257268544%	83,7442837630708%	83,7449242255169%
44550	84,7420300000%	84,7390102194475%	84,7406879832705%	84,7413360660515%

As Tabelas 6.15, 6.16, 6.17 e 6.18 descrevem os efeitos da política de manutenção adotada (ESMP) para os Estudos de caso 1, 2 e 3 quando uma equipe de manutenção é alocada. Essas tabelas demonstram que um menor intervalo entre manutenções preventivas (MTBM) para o Estudo de caso 3 em relação aos Estudos de casos 1 e 2 promove uma identificação mais rápida dos componentes defeituosos dos equipamentos, evitando eventos de falhas e reduzindo a degradação de desempenho da infraestrutura do sistema. Esses resultados podem ser adotados para a identificação das causas da degradação de desempenho do sistema, garantindo, assim, a qualidade do serviço oferecido.

6.3.1 Estudos de Caso de Performabilidade com Temperatura

Esses resultados de performabilidade são obtidos pela combinação dos níveis de disponibilidade e da utilização do processador e disco considerando as frequências: 3500, 4500, 4900, 5600, 6300, 7200, 8100, 9801, 14850, 16830, 18810, 20790, 22770, 24750, 29700, 34650, 39600 e 44550 tpms (transações por minuto).

As Tabelas 6.19 e 6.20 mostram os efeitos das variações de temperatura na degradação da utilização do processador e do disco. Essas tabelas representam os resultados da combinação dos níveis de disponibilidade e da utilização do processador e do disco considerando diferentes frequências de envio de transações comerciais.

Tabela 6.17: Utilização do Processador para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Confiabilidade

Frequência (tpms)	Utilização	Estudo de Caso 1	Estudo de Caso 2	Estudo de Caso 3
3500	48,4704100000%	46,3391779050%	48,0285381459%	48,3485771386%
4500	54,8245500000%	52,4139279204%	54,3247517611%	54,6867456818%
4900	57,0853200000%	54,5752924154%	56,5649118543%	56,9418331205%
5600	59,9811500000%	57,3437934772%	59,4343425362%	59,8303843033%
6300	63,3841500000%	60,5971643980%	62,8063196932%	63,2248306882%
7200	68,2340900000%	65,2338537202%	67,6120460796%	68,0625801153%
8100	71,6649500000%	68,5138596435%	71,0116292559%	71,4848164727%
9801	75,9839000000%	72,6429064664%	75,2912063180%	75,7929105704%
14850	84,5334800000%	80,8165635209%	83,7628455956%	84,3210007626%
16830	86,1891700000%	82,3994532358%	85,4034417928%	85,9725291009%
18810	88,0330200000%	84,1622296014%	87,2304826629%	87,8117444893%
20790	89,1646000000%	85,2440543051%	88,3517468155%	88,9404802050%
22770	90,3336500000%	86,3617014619%	89,5101393796%	90,1065917378%
24750	91,4079600000%	87,3887743135%	90,5746556239%	91,1782014046%
29700	92,7834200000%	88,7037556731%	91,9375764880%	92,5502041153%
34650	94,9875900000%	90,8110088563%	94,1216525650%	94,7488338210%
39600	96,7786000000%	92,5232685838%	95,8963351416%	96,5353420255%
44550	97,7030600000%	93,4070803033%	96,8123674667%	97,4574783479%

As Figuras 6.14(a) e 6.14(b) mostram o nível de degradação da utilização do processador e do disco, respectivamente, provocada por variações na temperatura. Esses resultados são obtidos pela combinação da utilização dos recursos de processamento e armazenamento com a disponibilidade desses recursos quando submetidos a variações de temperatura.

As Tabelas 6.19 e 6.20 descrevem os efeitos das variações de temperatura para os Estudos de caso 4, 5, 6 e 7. Um acréscimo de $10^{\circ}C$ na temperatura reduz o tempo médio para falha pela metade. Esse fato resulta em uma maior degradação dos componentes dos equipamentos e conseqüentemente uma maior ocorrência de eventos de falhas na infraestrutura do sistema TEF. Os estudos de caso de performabilidade com temperatura mostram que alterações significativas na temperatura resultam na degradação do nível da disponibilidade da infraestrutura do sistema TEF conseqüentemente, uma degradação nos níveis de performabilidade. Esses resultados podem ser empregados para identificação de possíveis causas da degradação dos equipamentos.

6.4 CONSIDERAÇÕES FINAIS

Este capítulo apresentou estudos de caso de desempenho, dependabilidade e performabilidade no sistema SCOPE. Neste capítulo, cenários foram definidos e nove estudos de

Tabela 6.18: Utilização do Disco para os Estudos de Caso 1, 2 e 3 Considerando a Métrica de Confiabilidade

Frequência (tpms)	Utilização	Estudo de Caso 1	Estudo de Caso 2	Estudo de Caso 3
3500	40,4223700000%	40,0173234275%	40,3537224956%	40,4046942525%
4500	45,8644000000%	45,4048223448%	45,7865105392%	45,8443445813%
4900	47,7533900000%	47,2748839909%	47,6722925519%	47,7325085706%
5600	50,2718300000%	49,7680883234%	50,1864555978%	50,2498473163%
6300	53,3411100000%	52,8066130425%	53,2505231768%	53,3177851926%
7200	57,5340000000%	56,9574887884%	57,4362925791%	57,5088417409%
8100	60,5208200000%	59,9143797861%	60,4180401962%	60,4943556751%
9801	64,2216400000%	63,5781162491%	64,1125752589%	64,1935573939%
14850	71,6090200000%	70,8914720652%	71,4874096016%	71,5777070671%
16830	73,1039200000%	72,3713926337%	72,9797708798%	73,0719533827%
18810	74,8079700000%	74,0583674446%	74,6809269678%	74,7752582419%
20790	75,8357100000%	75,0758091231%	75,7069216029%	75,8025488354%
22770	76,9567700000%	76,1856357018%	76,8260777568%	76,9231186224%
24750	78,0386100000%	77,2566352790%	77,9060805163%	78,0044855593%
29700	79,4415600000%	78,6455272193%	79,3066479490%	79,4068220824%
34650	81,7010500000%	80,8823763232%	81,5623007581%	81,6653240608%
39600	83,7456100000%	82,9064491024%	83,6033885733%	83,7089900230%
44550	84,7420300000%	83,8928846184%	84,5981163977%	84,7049743121%

caso foram realizados. Inicialmente, um estudo de caso de desempenho considerou diferentes frequências de envio de transações comerciais. O objetivo do estudo de caso de desempenho é a análise do impacto das variações das frequência de envio de transações comerciais na utilização dos recursos do servidor TEF. Em seguida, três estudos de caso de dependabilidade foram realizados para análise dos efeitos da política de manutenção adotada (ESMP) nos níveis disponibilidade e confiabilidade do sistema TEF e quatro estudos de caso de dependabilidade foram realizados para análise do impacto das variações de temperatura nos níveis de disponibilidade e confiabilidade do sistema TEF. Finalmente, estudos de caso de performabilidade combinaram os resultados das métricas de desempenho e os resultados das métricas de dependabilidade para a análise dos níveis de degradação do desempenho do processador e disco considerando os efeitos da política de manutenção adotada e os efeitos da variação de temperatura.

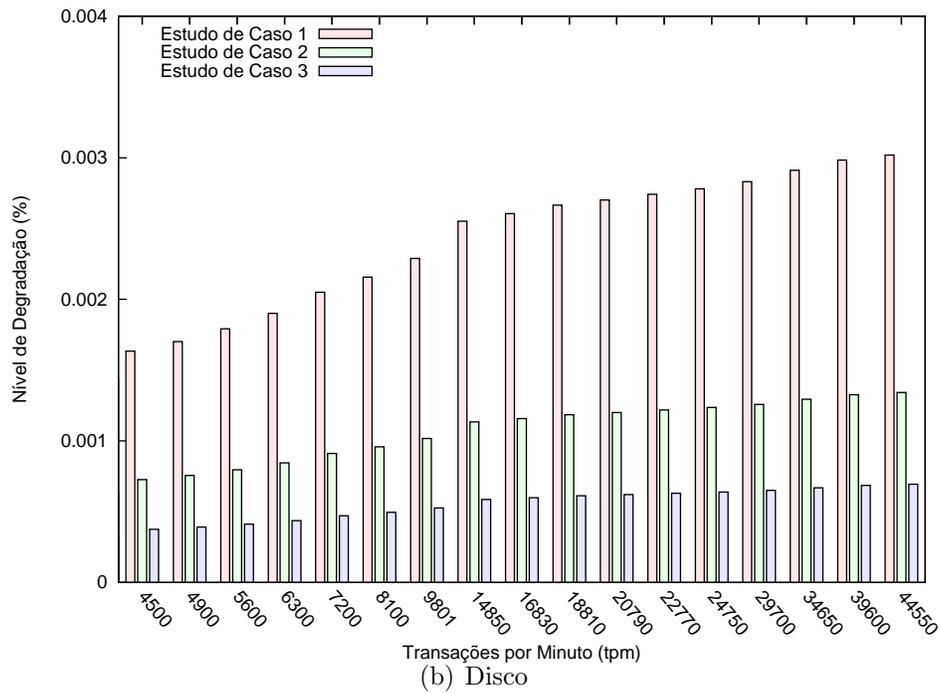
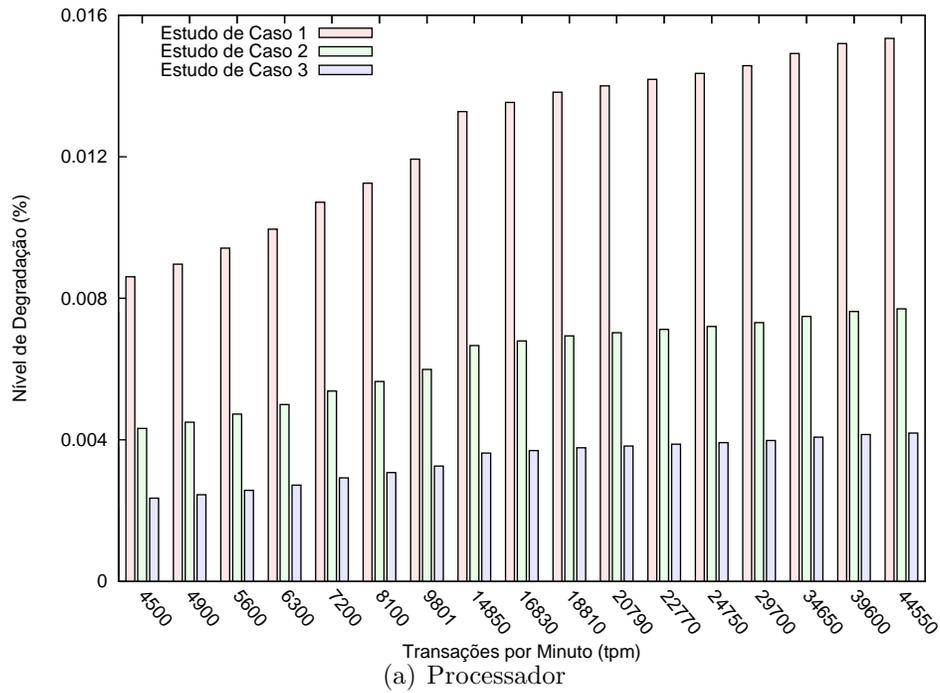


Figura 6.12: Nível de Degradação

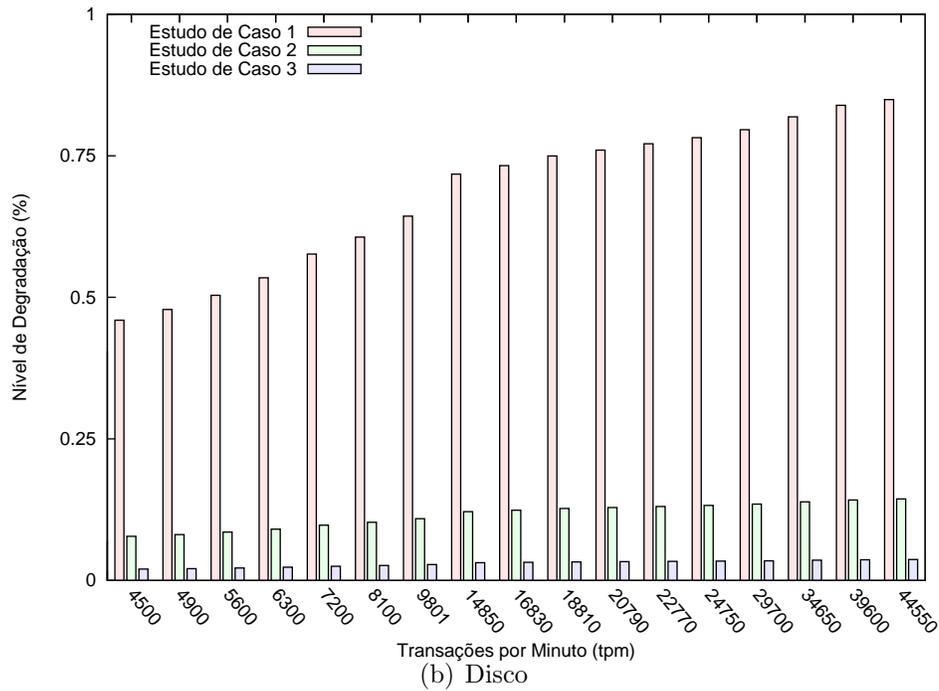
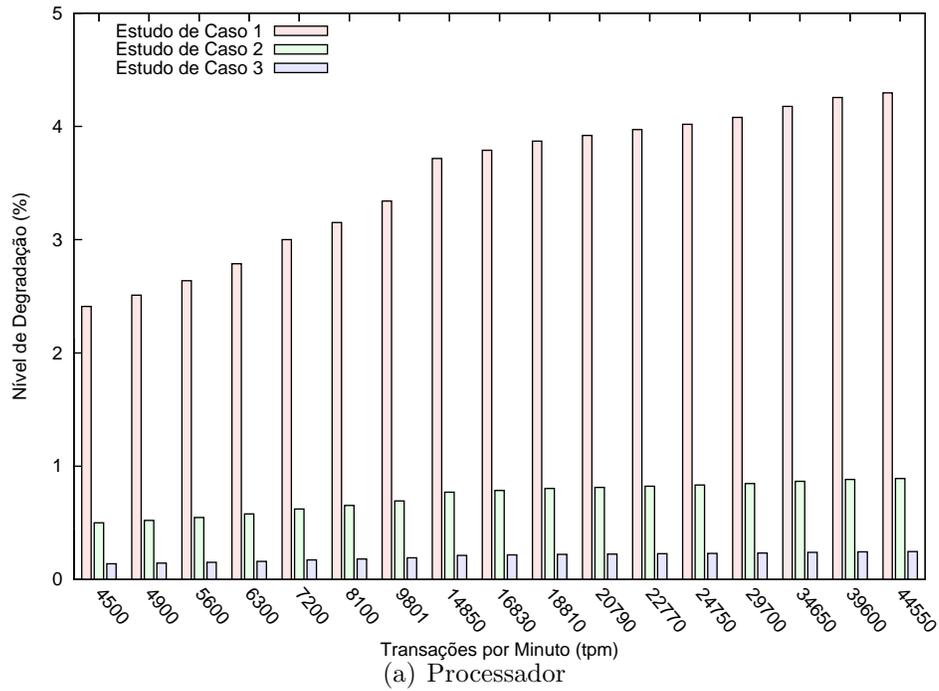


Figura 6.13: Nível de Degradação

Tabela 6.19: Utilização do Processador para os Estudos de Caso 4, 5, 6 e 7 Considerando a Métrica de Disponibilidade

Frequência (tpms)	Estudo de Caso 4	Estudo de Caso 5	Estudo de Caso 6	Estudo de Caso 7
3500	48,4261647563%	48,4078494843%	48,3819594989%	48,3453689957%
4500	54,7745045068%	54,7537882276%	54,7245042418%	54,6831169733%
4900	57,0332108082%	57,0116402630%	56,9811487096%	56,9380547769%
5600	59,9263974077%	59,9037326297%	59,8716942976%	59,8264142915%
6300	63,3262910473%	63,3023403947%	63,2684843841%	63,2206354398%
7200	68,1718038766%	68,1460206014%	68,1095740438%	68,0580638608%
8100	71,5995320847%	71,5724524075%	71,5341732903%	71,4800731376%
9801	75,9145396176%	75,8858279604%	75,8452419192%	75,7878813741%
14850	84,4563153047%	84,4243730602%	84,3792203463%	84,3154056896%
16830	86,1104939412%	86,0779260694%	86,0318889852%	85,9668244416%
18810	87,9526608196%	87,9193962214%	87,8723742632%	87,8059177900%
20790	89,0832078795%	89,0495156968%	89,0018893164%	88,9345786090%
22770	90,2511907356%	90,2170568098%	90,1688059930%	90,1006127652%
24750	91,3245200732%	91,2899802033%	91,2411555545%	91,1721513258%
29700	92,6987245120%	92,6636649039%	92,6141055669%	92,5440629981%
34650	94,9008824796%	94,8649899927%	94,8142533204%	94,7425468149%
39600	96,6902575919%	96,6536883451%	96,6019950226%	96,5289364767%
44550	97,6138737171%	97,5769551492%	97,5247680356%	97,4510116112%

Tabela 6.20: Utilização do Disco para os Estudos de Caso 4, 5, 6 e 7 Considerando a Métrica de Disponibilidade

Frequência (tpms)	Estudo de Caso 4	Estudo de Caso 5	Estudo de Caso 6	Estudo de Caso 7
3500	40,3854712486%	40,3988639191%	40,3486058441%	40,3753315432%
4500	45,8225336004%	45,8377293151%	45,7807050372%	45,8110287949%
4900	47,7097992737%	47,7256208453%	47,6662479421%	47,6978206265%
5600	50,2259403661%	50,2425963430%	50,1800922047%	50,2133299836%
6300	53,2924186353%	53,3100915208%	53,2437712751%	53,2790383426%
7200	57,4814812396%	57,5005433062%	57,4290099426%	57,4670491860%
8100	60,4655747807%	60,4856264354%	60,4103794889%	60,4503935015%
9801	64,1630165613%	64,1842943653%	64,1044461030%	64,1469069539%
14850	71,5436531393%	71,5673785175%	71,4783453533%	71,5256904526%
16830	73,0371885497%	73,0614092156%	72,9705174075%	73,0188508765%
18810	74,7396830418%	74,7644682906%	74,6714577974%	74,7209179180%
20790	75,7664848899%	75,7916106478%	75,6973223415%	75,7474619638%
22770	76,8865215527%	76,9120187383%	76,8163365919%	76,8672174155%
24750	77,9673740167%	77,9932296357%	77,8962024124%	77,9477985065%
29700	79,3690433619%	79,3953638039%	79,2965922601%	79,3491159302%
34650	81,6264708317%	81,6535398840%	81,5519590637%	81,6059766207%
39600	83,6691644960%	83,6969109485%	83,5927880790%	83,6481574196%
44550	84,6646749340%	84,6927515186%	84,5873897769%	84,6434179117%

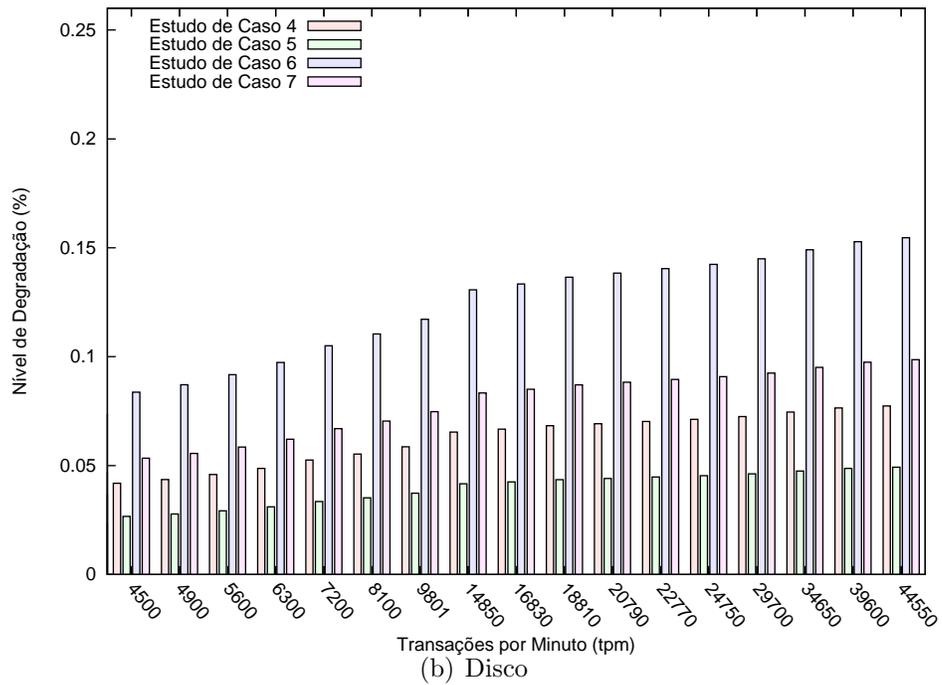
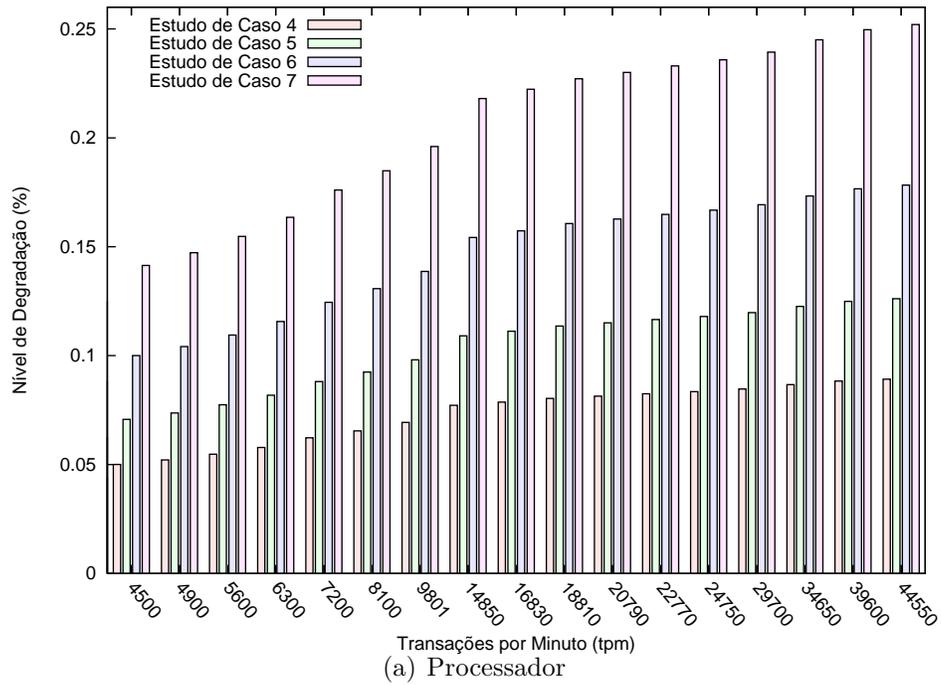


Figura 6.14: Nível de Degradação

Conclusões e Trabalhos Futuros

CONCLUSÕES E TRABALHOS FUTUROS

O uso de moedas, cédulas e cheques está sendo substituído, cada vez mais, por cartões de crédito e débito. As instituições financeiras, bancos e um crescente número de estabelecimentos comerciais oferecem aos seus clientes cartões que podem ser usados na compra de grande número de bens e serviços, inclusive em estabelecimentos comerciais virtuais através da internet. A justificativa para essa substituição está no fato de que o uso de cartões de crédito e débito como forma de pagamento é muito mais barato e seguro que outros meios, como o cheque ou mesmo o dinheiro em espécie.

Atualmente, existem cerca de 341 milhões de cartões de crédito e débito, e a taxa de crescimento deste mercado tem sido superior a 20% ao ano. Além disso, a expectativa é de que o crescimento continue em ritmo forte, principalmente pelo processo de substituição de outras formas de pagamento, como o cheque e o dinheiro, e a incorporação de facilidades.

O sistema de transferência eletrônica de fundos (TEF) permite que clientes efetuem pagamentos a estabelecimentos comerciais por meio de troca de mensagens eletrônicas entre os pontos de vendas (PDVs) distribuídos em diversas empresas ou filiais e empresas autorizadas. A expansão do mercado de transferência eletrônica de fundos exige confiabilidade, alta disponibilidade, escalabilidade e segurança das empresas que promovem esse serviço a preços acessíveis.

Esse trabalho propôs uma metodologia para avaliação de performabilidade baseada em modelos estocásticos expolinomiais. A metodologia para avaliação de performabilidade foi aplicada à estudos de caso que consideraram traces de transações reais coletados de usuários do sistema SCOPE. Os resultados da avaliação de performabilidade foram obtidos por meio da combinação dos níveis de disponibilidade e dos níveis de confiabilidade das unidades de processamento e de armazenamento com as utilizações desses recursos considerando um conjunto de diferentes frequências de envio de transações comerciais. Os resultados das métricas de desempenho foram obtidos através de análise estacionária, e os resultados das métricas de dependabilidade foram obtidos através de análise estacionária e análise transiente. As avaliações foram realizadas utilizando o *software* TimeNET 4.0, que permite a criação de modelos SPN.

Os modelos estocásticos expolinomiais propostos demonstram, através dos resultados apresentados, que podem ajudar no planejamento da infraestrutura do sistema TEF, garantindo a qualidade do serviço oferecido. A avaliação baseada em modelos possibilita a análise de desempenho e de dependabilidade de sistemas sem a necessidade de interferência no sistema real. Avaliar um sistema em operação pode não ser uma tarefa de fácil implementação ou ainda pode prejudicar sua operação. O uso de modelos como

mecanismo de avaliação também torna possível a avaliação de cenários complexos, cuja implementação real pode ser inviável devido aos custos.

Este trabalho também apresentou a análise do impacto de uma política de manutenção na performabilidade de sistemas TEF, a análise dos efeitos das variações da temperatura na performabilidade de sistemas TEF e análise do impacto de uma política de manutenção na redução dos efeitos das variações de temperatura. A avaliação de um conjunto de cenários ilustrou a metodologia de performabilidade adotada. Através da ferramenta TimeNET 4.0, vários experimentos foram realizados a fim de contemplar os efeitos de variações na temperatura e da política de manutenção adotada no desempenho do sistema TEF.

7.1 CONTRIBUIÇÕES

As principais contribuições desse trabalho são as proposições de:

- uma metodologia para avaliação de performabilidade de sistemas TEF, composta por um método para avaliação de desempenho, um método para avaliação de dependabilidade e uma técnica de decomposição e composição. Essa técnica tem o objetivo de avaliar o desempenho de sistemas TEF considerando a ocorrência de eventos de falhas e atividades de reparo. O processo adotado considera um conjunto de estados estruturais significativos do sistema TEF, em que cada estado corresponde a uma configuração que resulta em um determinado desempenho do sistema TEF. A avaliação de performabilidade do sistema TEF foi realizada através de modelos estocásticos expolinomiais tais como o modelo de desempenho de sistemas TEF e o modelo de dependabilidade de sistemas TEF;
- um modelo estocástico expolinomial para avaliação de desempenho. O modelo de desempenho de sistemas TEF permite a avaliação do impacto de flutuações da carga de trabalho (transações comerciais de crédito e débito) no desempenho do servidor TEF. Essa avaliação previne a degradação do desempenho garantindo, assim, a qualidade do serviço oferecido pelo sistema TEF. Ela proporciona também o planejamento das configurações de *hardware* e *software* adequadas de forma a garantir os níveis de desempenho acordados;
- um modelo estocástico expolinomial para avaliação de dependabilidade. O modelo de dependabilidade de sistemas TEF permite a avaliação das variações estruturais do sistema TEF;
- uma política de manutenção que proporciona através da definição dos seus parâmetros, o planejamento dos níveis de disponibilidade e confiabilidade que o sistema TEF poderá oferecer. Além disso, o modelo de dependabilidade de sistemas TEF também pode ser adotado em conjunto com a política de manutenção, proporcionando o ajuste dos intervalos entre as manutenções preventivas (MTBM), dos intervalos entre as substituições dos equipamentos (MTBR), do número máximo de manutenções

preventivas antes da substituição do equipamento, do tempo médio para reparo, do dimensionamento das equipes de manutenção e dos custos relacionados às manutenções, garantindo, assim, o cumprimento dos acordos de nível de serviço.

7.2 TRABALHOS FUTUROS

Como trabalhos futuros, a política de manutenção detalhará o processo de treinamento da equipe de manutenção local (LT) e especificará a duração do processo de diagnóstico. O modelo de dependabilidade de sistemas TEF será adotado para representação do processo de treinamento da equipe de manutenção e do processo de diagnóstico.

Adicionalmente, pode-se incluir a implementação de uma ferramenta que automatize a avaliação da performabilidade, considerando aspectos de variações na temperatura e da política de manutenção adotada nesse trabalho. Essa ferramenta possibilitará que usuários sem conhecimento específico de redes de Petri possam planejar os melhores tempos entre manutenções dos equipamentos e melhores tempos entre substituições dos equipamentos de forma a minimizar os custos. Além disso, o planejamento desses tempos levará em consideração variações na temperatura.

REFERÊNCIAS BIBLIOGRÁFICAS

- [AAB⁺03] C. Aubley, J. Ayala, C. Banson, P. Bijaoui, S. Daily, K. Delaney, T. Landry, D. Mar-Elia, T. Redmond, J. Rudich, and Zhou T. A Guide to Windows Performance Tuning. Winternals and Windows .NET Magazine eBooks, 2003.
- [ABE09] ABECS. *Brazilian Association of Business and Credit Card Services*. Disponível em <http://www.abecs.org.br/mercadocartoes.asp>, Acessado em 20/06/2009.
- [AR90] RJ Allen and WJ Roesch. *Reliability Prediction: The Applicability of High Temperature Testing*. 1990.
- [ASM⁺09] C. Araújo, E. Sousa, P. Maciel, F. Chicout, and E. Andrade. *Performance Modeling for Evaluation and Planning of Electronic Funds Transfer Systems with Bursty Arrival Traffic. Intensive Applications and Services, International Conference on*, 0:65–70, 2009.
- [ASRM07] R.D. Arteiro, F.N. Souza, N.S. Rosa, and P.R.M. Maciel. *Utilizando Redes de Petri para Modelagem de Desempenho de Middleware Orientado a Mensagem. WPerformance*, pages 1–21, 2007.
- [Bal01] G. Balbo. *Introduction to Stochastic Petri Nets. Lectures on Formal Methods and Performance Analysis: First EEF/Euro Summer School on Trends in Computer Science, Berg en Dal, The Netherlands, July 3-7, 2000: Revised Lectures*, 2001.
- [BCB09] BCB. *Banco Central do Brasil*. Disponível em <http://www.bcb.gov.br/?SPB>, Acessado em 10/05/2009.
- [BGdMT06] G. Bolch, S. Greiner, H. de Meer, and K.S. Trivedi. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. Wiley-Interscience, 2006.
- [BM03] W. Blichke and D. Murthy. *Case Studies in Reliability and Maintenance*. John Wiley, 2003.
- [Cam03] M.S. Campos. *Desvendando o Minitab*. Qualitymark, 2003.
- [CL08] C. Cassandras and S. Lafortune. *Introduction to Discrete Event Systems*. Springer, 2008.

- [DAJ95] A.A. Desrochers and R.Y. Al-Jaar. *Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis*. IEEE Press, 1995.
- [dSLJ+06] A.N. da Silva, F.A.A. Lins, J.C.S. Júnior, N.S. Rosa, N.C. Quental, and P.R.M. Maciel. *Avaliação de Desempenho da Composição de Web Services Usando Redes de Petri*. Brazilian Symposium on Computer Networks. Curitiba, Paraná, Brazil, 2006.
- [Ebe04] C.E. Ebeling. *An introduction to Reliability and Maintainability Engineering*. McGraw Hill, 2004.
- [Fri05] F. Friedman. *Microsoft Windows Server 2003 Performance Guide*. Microsoft Press Redmond, WA, USA, 2005.
- [Ger00] R. German. *Performance Analysis of Communication Systems with Non-Markovian Stochastic Petri Nets*. John Wiley & Sons, Inc. New York, NY, USA, 2000.
- [GLR+03] V. Gupta, V. Lam, H.G.V. Ramasamy, W.H. Sanders, and S. Singh. *Dependability and Performance Evaluation of Intrusion-tolerant Server Architectures*. *Lecture Notes in Computer Science*, pages 81–101, 2003.
- [GPT00] K. Goseva-Popstojanova and K. Trivedi. *Stochastic Modeling Formalisms for Dependability, Performance and Performability*. *Lecture Notes in Computer Science*, pages 403–422, 2000.
- [Gra91] WK Grassmann. *Finding Transient Solutions in Markovian Event Systems Through Randomization*. *Numerical Solution of Markov chains*, pages 357–371, 1991.
- [Gro08] D. Gross. *Fundamentals of Queueing Theory*. Wiley India Pvt. Ltd., 2008.
- [Her01] U. Herzog. *Formal Methods for Performance Evaluation*. *Lecture Notes in Computer Science*, 2090:1–37, 2001.
- [HMRT01] B.R. Haverkort, R. Marie, G. Rubino, and K.S. Trivedi. *Performability Modelling: Techniques and Tools*. John Wiley & Sons Inc, 2001.
- [HMS02] L. Higgins, K. Mobley, and R. Smith. *Maintenance Engineering Handbook*. McGraw-Hill Professional, 2002.
- [INA09] INA. *Integrated Net Analyzer*. Disponível em <http://www2.informatik.hu-berlin.de/~starke/ina.html>, Acessado em 30/01/2009.
- [Ita09a] Itaotec. *Itaotec*. Disponível em <http://www.itaotec.com.br>, Acessado em 10/04/2009.

- [Ita09b] Itaotec. Guia de Referência do Scope - Sistema de Pagamento Eletrônico de Fundos. Acessado em 25/06/2009.
- [Jai91] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons New York, 1991.
- [LAK92] J.C.C. Laprie, A. Avizienis, and H. Kopetz. *Dependability: Basic Concepts and Terminology*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1992.
- [Lil00] D.J. Lilja. *Measuring Computer Performance: A Practitioner's Guide*. Cambridge University Press, 2000.
- [lin09] linhacodigo. *Linha de Código*. Disponível em <http://www.linhadecodigo.com.br/Artigo.aspx?id=835>, Acessado em 15/05/2009.
- [LYT⁺03] M. Lanus, L. Yin, KS Trivedi, M. Inc, and AZ Tempe. *Hierarchical Composition and Aggregation of State-based Availability and Performability Models*. *IEEE Transactions on Reliability*, 52(1):44–52, 2003.
- [MA05] D.A. Menascé and V.A.F. Almeida. *Performance by Design: Computer Capacity Planning by Example*. Prentice Hall PTR, 2005.
- [MAFM00] D.A. Menascé, V.A.F. Almeida, R. Fonseca, and M.A. Mendes. *Business-oriented Resource Management Policies for E-commerce Servers*. *Performance Evaluation*, 42(2-3):223–239, 2000.
- [MBC⁺98] M.A. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. *ACM SIGMETRICS Performance Evaluation Review*, 26(2), 1998.
- [MLC96] P.R.M. Maciel, R.D. Lins, and P.R.F. Cunha. *Introduction of the Petri Net and Applied*. *X Escola de Computação, Campinas, SP*, 1996.
- [MM09] Amip Shah Ratnesh Sharma Tom Christian Virgilio Almeida Carlos Araújo Erica Souza Gustavo Callou Bruno Silva Sérgio Galdino Jose Pires Manish Marwah, Paulo Maciel. *Quantifying the Sustainability Impact of Data Center Availability*. 2009.
- [MR03] D.C. Montgomery and G.C. Runger. *Estatística Aplicada e Probabilidade para Engenheiros*. Livros Técnicos e Científicos, 2003.
- [MRI01] S. Mahevas, G. Rubino, and R. Irisa. *Bound Computation of Dependability and Performance Measures*. *IEEE Transactions on Computers*, 50(5):399–413, 2001.
- [Mun09] A.L.R. Munhoz. *Transferência Eletrônica de Fundos*. Disponível em <http://www.linhadecodigo.com.br/Artigo.aspx?id=864>, Acessado em 12/06/2009.

- [Mur89] T. Murata. *Petri Nets: Properties, Analysis and Applications. Proceedings of the IEEE*, 77(4):541–580, 1989.
- [NGB⁺05] K. Nagaraja, G. Gama, R. Bianchini, RP Martin, W. Meira Jr, and TD Nguyen. *Quantifying the Performability of Cluster-based Services. IEEE Transactions on Parallel and Distributed Systems*, 16(5):456–467, 2005.
- [PRS96] A. Puliafito, S. Riccobene, and M. Scarpa. *Evaluation of Performability Parameters in Client-server Environments. The Computer Journal*, 39(8):647–662, 1996.
- [PSST08] G. Pacifici, W. Segmuller, M. Spreitzer, and A. Tantawi. *CPU Demand for Web Serving: Measurement Analysis and Dynamic Estimation. Performance Evaluation*, 65(6-7):531–553, 2008.
- [Rei90] A.L. Reibman. *Modeling the Effect of Reliability on Performance. IEEE Transactions on Reliability*, 39(3):314–320, 1990.
- [RK03] M. Rabah and K. Kanoun. *Performability Evaluation of Multipurpose Multiprocessor Systems: the "Separation of Concerns" Approach. IEEE Transactions on Computers*, 52(2):223–236, 2003.
- [SARM06] F.N. Souza, R.D. Arteiro, N.S. Rosa, and P.R.M. Maciel. *Using Stochastic Petri Nets for Performance Modelling of Application Servers. Performance Modelling, Evaluation, and Optimisation of Parallel and Distributed Systems*, pages 1–8, 2006.
- [Shi04] V.T.S. Shi. *Evaluating the Performability of Tactical Communications Networks. IEEE Transactions on Vehicular Technology*, 53(1):253–260, 2004.
- [SMA09a] E. Sousa, P. Maciel, and C. Araújo. Performability Evaluation of EFT Systems using Expolynomial Stochastic Models. Proceedings of ICDS, 2009.
- [SMA⁺09b] E. Sousa, P. Maciel, C. Araújo, G. Alves, and F. Chicout. Performance Modeling for Evaluation and Planning of Electronic Funds Transfer Systems. IEEE Proceeding of ISCC, 2009.
- [SMAC09] E. Sousa, P. Maciel, C. Araújo, and F. Chicout. *Performability Evaluation of EFT Systems for SLA Assurance. Parallel and Distributed Processing Symposium, International*, 0:1–8, 2009.
- [Smi05] D.J. Smith. *Reliability, Maintainability and Risk: Practical Methods for Engineers*. Butterworth-Heinemann, 2005.
- [ST00] A. Sesmun and LF Turner. *Using Performability in the Design of Communication Networks. IEE Proceedings-Computers and Digital Techniques*, 147(5):355–363, 2000.

- [STP96] R. Sahner, K.S. Trivedi, and A. Puliafito. *Performance and Reliability Analysis of Computer Systems: An Example-based Approach Using the SHARPE Software Package*. Kluwer Academic Publishers, 1996.
- [TCMS93] K.S. Trivedi, G. Ciardo, M. Malhotra, and R.A. Sahner. *Dependability and Performability Analysis. Performance/SIGMETRICS Tutorials*, pages 587–612, 1993.
- [THGF96] K.S. Trivedi, S. Hunter, S. Garg, and R. Fricks. *Reliability Analysis Techniques Explored Through a Communication Network Example*. In *International Workshop on Computer-Aided Design, Test, and Evaluation for Dependability*, 1996.
- [TM93] K.S. Trivedi and M. Malhotra. *Reliability and Performability Techniques and Tools: A Survey. Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen*, pages 27–48, 1993.
- [Tri06] K.S. Trivedi. *Probability and Statistics with Reliability, Queuing, and Computer Science Applications*. John Wiley and Sons, 2nd Edition, 2006.
- [WP06] H. Wang and H. Pham. *Reliability and Optimal Maintenance*. Springer Verlag, 2006.
- [XDP04] M. Xie, Y.S. Dai, and K.L. Poh. *Computing System Reliability: Models and Analysis*. Kluwer Academic/Plenum Publishers, 2004.
- [ZK07] A. Zimmermann and M. Knoke. *A Software Tool the Performability Evaluation with Stochastic and Colored Petri Nets*. Technische Universitt Berlin. Real-Time Systems and Robotics Group, 2007.

