



Universidade Federal de Pernambuco  
Centro de Informática

Mestrado em Ciência da Computação

**MODELAGEM HIERÁRQUICA E  
HETEROGÊNEA PARA INFRAESTRUTURA  
DE REDES CONVERGENTES E POLÍTICA  
DE MANUTENÇÃO PARA GARANTIA DE  
NÍVEIS DE SERVIÇOS**

Kádna Maria Alves Camboim

DISSERTAÇÃO DE MESTRADO

Recife  
14 de Setembro de 2012

Universidade Federal de Pernambuco  
Centro de Informática

Kádna Maria Alves Camboim

**MODELAGEM HIERÁRQUICA E HETEROGÊNEA PARA  
INFRAESTRUTURA DE REDES CONVERGENTES E POLÍTICA  
DE MANUTENÇÃO PARA GARANTIA DE NÍVEIS DE SERVIÇOS**

*Trabalho apresentado ao Programa de Mestrado em  
Ciência da Computação do Centro de Informática da Uni-  
versidade Federal de Pernambuco como requisito parcial  
para obtenção do grau de Mestre em Ciência da Com-  
putação.*

*Orientador: Prof. Dr. Paulo Romero Martins Maciel*

Recife  
14 de Setembro de 2012

**Catálogo na fonte**  
**Bibliotecária Jane Souto Maior, CRB4-571**

**Camboim, Kádna Maria Alves**

**Modelagem hierárquica e heterogênea para infraestrutura de redes convergentes e política de manutenção para garantia de níveis de serviços / Kádna Maria Alves Camboim. - Recife: O Autor, 2012.**

**xv, 78 folhas: il., fig., tab.**

**Orientador: Paulo Romero Martins Maciel.**

**Dissertação (mestrado) - Universidade Federal de Pernambuco. Cln, Ciência da Computação, 2012.**

**Inclui bibliografia.**

**1. Redes de computadores e Sistemas distribuídos. I. Maciel, Paulo Romero Martins (orientador). II. Título.**

**004.6**

**CDD (23. ed.)**

**MEI2012 – 196**

Dissertação de Mestrado apresentada por **Kádna Maria Alves Camboim** à Pós-Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, sob o título “**Modelagem Hierárquica e Heterogênea para Infraestrutura de Redes Convergentes e Política de Manutenção para Garantia de Níveis de Serviço**” orientada pelo **Prof. Paulo Romero Martins Maciel** e aprovada pela Banca Examinadora formada pelos professores:

---

Prof. Eduardo Antonio Guimarães Tavares  
Centro de Informática / UFPE

---

Prof. Sérgio Mário Lins Galdino  
Departamento Básico da Escola Politécnica / UPE

---

Prof. Paulo Romero Martins Mscirl  
Centro de Informática / UFPE

Visto e permitida a impressão.  
Recife, 14 de setembro de 2012.

---

**Prof. Nelson Souto Rosa**

Coordenador da Pós-Graduação em Ciência da Computação do  
Centro de Informática da Universidade Federal de Pernambuco.

*A Deus.  
Aos meus pais e irmãos.  
Ao Prof. Dr. Paulo Maciel.  
A Baby Lhu.*

## AGRADECIMENTOS

Primeiramente, agradeço a Deus pela vida e pelas pessoas que fazem parte dela. Agradeço-Lhe, também, pelos caminhos que me permitiu escolher; pela força e coragem para enfrentar cada dificuldade durante a realização deste trabalho e por mais essa conquista.

Ao meu orientador, Paulo Romero Martins Maciel, pelo compartilhamento de sua experiência, orientação e paciência ao longo desses anos. Agradeço-lhe pela oportunidade de crescimento e por ter acreditado no meu esforço.

Aos professores Sergio Galdino e Eduardo Tavares, por terem aceitado o convite para compor a banca de defesa do mestrado.

Ao pessoal do grupo MoDCS (*Modeling of Distributed and Concurrent Systems*), pelos bons ensinamentos e divertidos encontros. Agradeço em especial a Almir Guimarães, por toda informação transmitida, a Erica Souza, que não mediu esforços em repassar seus conhecimentos e a Rubens Matos, pela disponibilidade em me ajudar. A Jean, Erico e Marcos, minha equipe de mestrado, pelas divertidas horas que passamos aprendendo juntos. Aos amigos Julian, Gustavo, Jair, Rafael, João e Jamilson por suas importantes contribuições.

A Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), que me promoveu recursos financeiros para execução deste trabalho. Ao Centro de Informática (CIn) da Universidade Federal de Pernambuco (UFPE), que forneceu suporte durante a realização desta pesquisa.

Aos meus amados pais, Davi e Fátima, que superaram a distância e a dor da saudade para que eu pudesse trilhar os meus caminhos. Obrigada pelos conselhos, educação e amor oferecidos sempre. Tudo o que sou, agradeço a vocês. Aos meus irmãos Kadmu e Kennedy pelo carinho e apoio. A toda minha família, que tanto acreditou em mim. Agradeço a Pedrinho, pelas inúmeras palavras de apoio e pelo brilho que me transmitiu.

Agradeço a minha grande amiga, Maria de Lourdes (*Baby Luh*), que desde a minha infância me orientou com seus bons conselhos. A Tatiane, Romário e Séfora pelos anos de boa convivência e amizade. Ricelly e Gabriel que me incentivaram e acreditaram na minha capacidade. As pessoas que conheci no CIn: Brunno, Fabiane, Alan, Andreza, Danilo, Renata, César, Josino, Felipe, José e Pedro, por cada momento de descontração e apoio. Ao pessoal de serviços-gerais e as secretárias da Pós-graduação, pela presteza.

Enfim, aos que amo, que de forma direta ou indireta contribuíram para essa conquista.

A todos, meu sincero agradecimento.

*Tão importante quanto aprender a fazer é aprender que você pode fazer.*

—DAVID NIVEN

## RESUMO

O crescimento contínuo das redes convergentes é influenciado pela demanda por tecnologias modernas de telecomunicações, tais como rádio, televisão, redes de computadores e de telefonia. As redes convergentes permitem a integração de vários tipos de serviços utilizando o protocolo roteável IP (*Internet Protocol*) sobre uma infraestrutura de transmissão unificada. A complexidade introduzida pela heterogeneidade dessas redes traz exigências de um controle eficaz e escalável e de soluções para gerenciamento, principalmente, quando se objetiva alcançar os requisitos definidos nos acordos de níveis de serviços (SLAs). Diferentes técnicas de modelagem são consideradas fundamentais por permitirem representar o comportamento de sistemas e facilitar a avaliação de dependabilidade. Além disso, várias abordagens para políticas de manutenção podem ser utilizadas, resultando em custos de manutenção menores e menos modificações a fazer, além de serem capazes de melhorar o desempenho do sistema em questão. Neste sentido, este trabalho propõe modelos hierárquicos e heterogêneos para avaliação do impacto de políticas de manutenção em infraestrutura de redes convergentes. Adicionalmente, são verificados diferentes níveis de serviços, previamente especificados em SLAs. Duas abordagens são adotadas para a política de manutenção, as quais compreendem a realização de manutenções preventivas e de manutenções corretivas. Na manutenção preventiva, equipes locais são responsáveis por realizar manutenções periódicas, a fim de minimizar ou evitar ocorrência de eventos de falhas e na manutenção corretiva, uma equipe especialista é chamada para recuperar falhas ou substituir componentes. Estudos de caso são realizados para apresentar índices operacionais na verificação dos serviços acordados em SLAs.

**Palavras-chave:** Acordos de Nível de Serviço (SLA), Avaliação de Dependabilidade, Cadeias de Markov de Tempo Contínuo (CTMC), Diagramas de Blocos de Confiabilidade (RBD), Manutenção Corretiva, Manutenção Preventiva, Política de Manutenção, Redes Convergentes, Redes de Petri Estocástica (SPN).

## ABSTRACT

The continuous growth of converged networks is influenced by the demand for modern telecommunications technologies, such as radio, television, computer and telephone networks. Converged networks allow integration of different types of services through the routable protocol IP (Internet Protocol) over a unified transmission infrastructure. The complexity introduced by the heterogeneity of these networks requires an efficient and scalable control, as also as management solutions in order to, above all, reach the conditions defined in the Services Level Agreement (SLAs). Various modeling techniques are considered fundamental for enabling the representation of a systems' behavior, as well as facilitating the dependability evaluation. Moreover, a number of approaches for maintenance policies can be adopted, resulting in lower maintenance costs and less modifications needed, as well as being able to increase the performance of the system. In this sense, this work proposes hierarchical and heterogeneous models to evaluate the impact of maintenance policies in converged network infrastructure. Additionally, different service levels are verified, previously specified in SLAs. Two approaches are adopted for the maintenance policy, which comprehend preventive maintenance and corrective maintenance. In preventive maintenance, local teams are responsible for performing periodic maintenance, in order to minimize or even avoid the occurrence of failure events. In the other hand, in corrective maintenance, a specialized team is called to recover failures or replace components. Case studies are used to present operational rates regarding the services accorded in the SLAs.

**Keywords:** Service Level Agreements (SLA), Dependability Evaluation, Continuous Time Markov Chains (CTMC), Reliability Block Diagram (RBD), Corrective Maintenance, Preventive Maintenance, Maintenance Policy, Converged Networks, Stochastic Petri Nets (SPN).

# SUMÁRIO

<b>Lista de Figuras</b>	x
<b>Lista de Tabelas</b>	xii
<b>Lista de Abreviaturas</b>	xiii
<b>Capítulo 1—Introdução</b>	1
1.1 Motivação e Justificativa . . . . .	4
1.2 Trabalhos Relacionados . . . . .	5
1.3 Objetivos . . . . .	7
1.4 Estrutura da Dissertação . . . . .	7
<b>Capítulo 2—Fundamentação Teórica</b>	8
2.1 Redes Convergentes para Instituições de Ensino . . . . .	8
2.2 Técnicas para Modelagem Analítica de Dependabilidade . . . . .	9
2.2.1 Cadeias de Markov . . . . .	10
2.2.2 Redes de Petri . . . . .	13
2.2.2.1 Propriedades das Redes de Petri . . . . .	16
2.2.2.2 Rede de Petri Estocástica . . . . .	20
2.2.3 Diagramas de Bloco de Confiabilidade . . . . .	24
2.3 Dependabilidade . . . . .	27
2.4 Política de manutenção . . . . .	32
2.5 Acordo de Nível de Serviço . . . . .	37
2.6 Considerações Finais . . . . .	38
<b>Capítulo 3—Metodologia e Modelos</b>	39

3.1	Introdução . . . . .	39
3.2	Metodologia . . . . .	40
3.3	Descrição da Rede . . . . .	41
3.4	Modelos . . . . .	42
3.4.1	Modelo RBD de Primeiro Nível . . . . .	43
3.4.2	Modelo RBD de Segundo Nível . . . . .	45
3.4.2.1	Modelo CTMC . . . . .	46
3.4.3	Modelo RBD de Terceiro Nível . . . . .	46
3.4.4	Modelo SPN . . . . .	48
3.5	Considerações Finais . . . . .	54
<b>Capítulo 4—Estudo de Caso</b>		<b>57</b>
4.1	Definição dos Custos . . . . .	57
4.1.1	Custo de Mão de Obra das Equipes . . . . .	57
4.1.2	Custo de Reparo dos Componentes . . . . .	57
4.1.3	Custo de um <i>Site</i> . . . . .	58
4.2	Custos de Manutenção . . . . .	59
4.2.1	Custo da Manutenção Preventiva . . . . .	60
4.2.2	Custo da Manutenção Corretiva - Equipe Local . . . . .	60
4.2.3	Custo da Manutenção Corretiva - Equipe Local e Especialista . . . . .	61
4.3	Definição dos Cenários e dos Parâmetros do Contrato SLA . . . . .	61
4.3.1	Avaliação dos Cenários . . . . .	62
4.3.2	Discussão de Resultados . . . . .	66
4.4	Considerações Finais . . . . .	67
<b>Capítulo 5—Conclusões</b>		<b>69</b>
5.0.1	Trabalhos Futuros . . . . .	70
<b>Referências</b>		<b>78</b>

## LISTA DE FIGURAS

2.1	Elementos de uma Rede de Petri . . . . .	14
2.2	Rede de Petri . . . . .	14
2.3	Períodos do Dia . . . . .	15
2.4	Conexão Série com Erlang . . . . .	16
2.5	Rede Estritamente Conservativa . . . . .	18
2.6	Diagramas de Blocos . . . . .	25
2.7	Diagrama de Blocos Série-Paralelo . . . . .	26
2.8	Diagrama de Blocos Paralelo-Série . . . . .	26
2.9	Diagrama de Blocos Série-Paralelo (Possível de Combinação) . . . . .	26
2.10	Diagrama de Blocos Série-Paralelo (Resultado da Combinação) . . . . .	27
2.11	Árvore de Dependabilidade . . . . .	29
2.12	Curva da Banheira . . . . .	30
2.13	Tempo de Inatividade para Manutenção . . . . .	35
3.1	Metodologia . . . . .	40
3.2	Topologia do <i>Site1</i> . . . . .	42
3.3	Topologia dos <i>Sites 2, 3 e 4</i> . . . . .	43
3.4	Modelo Operacional - Componentes de um Computador . . . . .	44
3.5	Modelo RBD de Primeiro Nível . . . . .	45
3.6	Modelo Operacional - 30 Computadores (21/9) . . . . .	45
3.7	Modelo RBD de Segundo Nível . . . . .	45
3.8	Confiabilidade X Inconfiabilidade dos Computadores . . . . .	46
3.9	Modelo CTMC . . . . .	47
3.10	Modelo Operacional - Componentes de um <i>Site</i> . . . . .	47
3.11	Modelo RBD de Terceiro Nível . . . . .	48
3.12	Confiabilidade X Inconfiabilidade do Sistema . . . . .	48

3.13	SPN Tempos para Falha de um <i>Site</i> . . . . .	50
3.14	SPN Manutenção Preventiva pela Equipe Local . . . . .	51
3.15	SPN Manutenção Corretiva . . . . .	51
3.16	SPN Sistema em Estado Operacional . . . . .	53
3.17	SPN Sistema em Estado Não Operacional . . . . .	53
3.18	SPN Equipe Especialista . . . . .	54
3.19	SPN Site1 . . . . .	55
3.20	Modelo SPN Completo . . . . .	56
4.1	Disponibilidade Alcançada com as Avaliações dos Cenários . . . . .	67

## LISTA DE TABELAS

3.1	Informações Estatísticas dos Polos . . . . .	41
3.2	MTTF Processador e Disco . . . . .	43
3.3	MTTF dos Componentes de um PC . . . . .	44
3.4	MTTF dos Componentes de um <i>Site</i> . . . . .	47
3.5	Função de Habilitação para a Manutenção Preventiva . . . . .	51
3.6	Tempo de Deslocamento da Equipe Especialista . . . . .	52
3.7	Função de Habilitação Sistema . . . . .	53
3.8	Número de Pessoas por Equipes . . . . .	54
4.1	Custos de Mão de Obra das Equipes . . . . .	58
4.2	Custos de Aquisição e de Reparo dos Componentes de um <i>Site</i> . . . . .	58
4.3	Custo Total de Reparo dos Componentes de um <i>Site</i> . . . . .	59
4.4	Custos de um <i>Site</i> . . . . .	59
4.5	Parâmetros de um Contrato SLA . . . . .	62
4.6	Parâmetros dos Cenários . . . . .	63
4.7	Resultados das Avaliações dos Cenários . . . . .	64
4.8	Resultados das Avaliações dos Cenários - Custo das Manutenções . . . . .	64

## LISTA DE ABREVIATURAS

**AC** - *Annual Cost.*

**CPU** - *Central Processing Unit.*

**CM** - *Corrective Maintenance.*

**CTMC** - *Continuous Time Markov Chains.*

**DSL** - *Digital Subscriber Line.*

**DTMC** - *Discrete Time Markov Chains.*

**EaD** - *Educação à Distância.*

**GSPN** - *Generalized Stochastic Petri Nets.*

**IDNS** - *Integrated Service Digital Network.*

**IP** - *Internet Protocol.*

**IPTV** - *Internet Protocol Television.*

**LT** - *Local Teams.*

**MDT** - *Mean Downtime.*

**MTBF** - *Mean Time Between Failures.*

**MTTF** - *Mean Time to Failure.*

**MTTR** - *Mean Time to Repair.*

**NTIC** - *Novas Tecnologias da Informação e Comunicação.*

**PC** - *Personal Computer.*

**PM** - *Preventive Maintenance.*

**PN** - *Petri Nets.*

**QoS** - *Quality of Service.*

**RBD** - *Reliability Block Diagram.*

**RG** - *Reliability graph.*

**RI** - *Reliability Importance.*

**SHARPE** - *Symbolic Hierarchical Automated Reliability and Performance Evaluator.*

**SLA** - *Service Level Agreement.*

**SLM** - *Service Level Management.*

**SO** - *Sistema Operacional.*

**SPN** - *Stochastic Petri Nets.*

**ST** - *Specialist Teams.*

**TCO** - *Total Cost Ownership.*

**TCP** - *Transmission Control Protocol.*

**TimeNET** - *Timed Net Evaluation.*

**TPN** - *Timed Petri Nets*.

**UNISANTOS** - Universidade Católica de Santos.

**UNOESTE** - Universidade do Oeste Paulista.

**VoIP** - *Voice over Internet Protocol*.

**WAN** - *Wide Area Network*.

# INTRODUÇÃO

A crescente evolução das novas tecnologias da informação e comunicação (NTICs) e a abrangência da Internet são fatores que influenciam a convergência das redes tradicionais para uma infraestrutura única de transmissão em pacotes.

A convergência pode ser entendida como um processo pelo qual as telecomunicações, tecnologia da informação e os meios de comunicação estão crescendo juntos, representando uma mudança qualitativa que liga dois ou mais mercados existentes, anteriormente distintos. Isso afeta dispositivos de usuário final, serviços e infraestruturas. Outra fonte importante de convergência do mercado é a mudança nas características do produto resultantes de novas tecnologias (convergência do produto) [SJ06].

As redes convergentes realizam a fusão de tecnologias modernas de telecomunicações, tais como rádio, televisão, redes de computadores e de telefonia, baseadas no protocolo TCP/IP sobre uma plataforma de rede unificada, capaz de prover os serviços antes prestados pelas diversas redes. Com isso, simplificam o gerenciamento e facilitam o desenvolvimento de aplicações multimídia, reduzindo despesas de investimentos e custos de operação [SJ06].

Estamos testemunhando o desdobramento mundial de aplicações de vídeo e áudio na Internet. Milhares de *sites*, incluindo CNN, Rhapsody, Napster, MSN, AOL, Yahoo, disponibilizam o envio de conteúdo de áudio e vídeo. O YouTube e outros sites de compartilhamento de vídeo permitem que os usuários vejam (por pedido) vídeos que foram carregados por outros usuários [KR10].

Milhões de usuários usam o *Skype* regularmente para suas necessidades relacionadas à telefonia e videoconferência. E alguns canais tradicionais de televisão estão sendo transmitidos pela Internet, permitindo que os usuários da Internet assistam aos canais de televisão de todos os lugares do mundo. Esse crescimento explosivo das aplicações multimídia da Internet impulsiona a convergência tecnológica e pode ser visto como resultado do aumento da penetração do acesso residencial devido à banda larga e do acesso sem fio de alta velocidade (como *WiFi*) [KR10].

Um dos primeiros exemplos de convergência foi entre as redes de voz e as redes de dados, antes, através da tecnologia IDNS (*Integrated Service Digital Network*) e, mais recentemente, pela tecnologia xDSL (*Digital Subscriber Line*) [SJ06]. Com a inclusão dos serviços de vídeo e multimídia, muitos outros serviços já nasceram convergentes, como é o caso do IPTV [KR10]. Kurose e Ross [KR10] citam três classes de multimídia abrangentes, que englobam diversos tipos de tecnologias, são elas:

- Áudio e vídeo de fluxo contínuo armazenados - Nesta classe de aplicações, clientes requisitam sob demanda, arquivos de áudio e vídeo comprimidos que estão armazenado em servidores.
- Áudio e vídeo de fluxo contínuo ao vivo - É semelhante a transmissão tradicional de rádio e televisão, exceto que a transmissão é realizada pela Internet. Essas aplicações são reconhecidas como rádio da Internet e IPTV. O fluxo contínuo de áudio e vídeo não é armazenado.
- Áudio e vídeo interativos em tempo real - Permite que pessoas usem áudio e vídeo para se comunicar em tempo real. Áudio interativo em tempo real pela Internet é frequentemente denominado telefone por Internet (VoIP), pois, da perspectiva do usuário, é semelhante ao tradicional serviço telefônico por comutação de circuitos. Hoje, há muitos produtos de telefonia por Internet disponíveis. Por exemplo, usuários do *Skype* podem fazer chamadas por voz de computador para telefone e de computador para computador a um custo muito baixo. Com vídeo interativo em tempo real, também denominado de videoconferência, indivíduos se comunicam visual e oralmente.

Uma vídeo-aula é um exemplo de aplicação que se encaixa na classe multimídia de áudio e vídeo de fluxo contínuo armazenados. As vídeo-aulas são gravadas e distribuídas em forma de vídeo e servem de auxílio para obtenção de conhecimentos e fixação de conteúdos. Ultimamente, essa tecnologia vem sendo bastante utilizada em cursos de educação à distância (EAD), devido também à difusão da Internet.

VoIP e videoconferência são exemplos de aplicações de áudio e vídeo interativos em tempo real. O VoIP torna possível a conversação entre usuários através da troca de mensagens de áudio, simulando uma ligação telefônica. Usuários conectados podem fazer ligações uns aos outros e manter comunicação em tempo real, para isso, basta apenas possuir instaladas em suas máquinas ferramentas que possibilitem a transmissão e o recebimento da voz. Assim, a voz é enviada ao seu destino na forma de pacote de dados. Um fator que favorece o uso mais eficaz da infraestrutura, bem como a redução dos custos, é a possibilidade que o VoIP tem de operar numa mesma rede (com uso de adaptadores), com diferentes aplicações por meio de telefones e computadores.

A videoconferência é uma tecnologia que permite uma discussão em grupo ou pessoa a pessoa através do contato visual e sonoro entre participantes que se encontram em locais diferentes, mas podem ver e ouvir uns aos outros, dando a sensação de que os interlocutores se encontram no mesmo local. A comunicação em tempo real entre grupos de pessoas só é possível graças aos sistemas interpessoais de videoconferência, que possibilitam trabalhar de forma cooperativa, compartilhando informações e materiais independentemente de suas localizações geográficas.

Além das tecnologias citadas, incluindo rádio da Internet e IPTV, existem ainda as tradicionais aplicações elásticas, como *webmail* para envio e recebimento de mensagens,

navegação na *web*, *login* remoto, *download* e compartilhamento de arquivo. O uso de todas essas aplicações em uma só infraestrutura de rede só foi possível graças à convergência das redes.

A digitalização de conteúdos e redes é o pré-requisito para convergência, de maneira a prover serviços que, anteriormente, requeriam equipamentos, canais de comunicação, protocolos e padrões independentes, agora em uma infraestrutura unificada [SJ06]. Isto torna possível enviar os diferentes serviços através de redes integradas em uma extensão muito maior do que antes. Em particular, diferentemente das aplicações elásticas, as aplicações multimídia são muito mais sensíveis a atrasos ponto a ponto e a variações de atrasos. Nesse contexto, o projeto para redes convergentes deve ser bem elaborado, de forma que possam dar adequada vazão ao tráfego requerido pelos diferentes serviços utilizados.

Quando comparadas com redes independentes, as redes de comunicação convergentes, com interfaces abertas e capacidade para transmitir voz, dados, imagens, som e vídeo, exploram plenamente as tecnologias de ponta para oferecer serviços sofisticados e aumentar as receitas dos operadores, reduzindo despesas de investimentos e custos de operação [NM05]. Adicionalmente, a integração de recursos e a convergência do tráfego permitem o compartilhamento da operação, a administração, a manutenção de equipamentos e facilidades para o desenvolvimento de aplicações multimídia. Assim, é possível ampliar fontes de receitas e atingir os requisitos definidos em acordos de níveis de serviços (SLA), com base nas vantagens oferecidas.

Uma rede convergente pode ser considerada como uma associação de redes autônomas, operadas e mantidas por autoridades diferentes, cada qual com sua própria forma de disponibilizar e recuperar o serviço. Os componentes da rede que contribuem para disponibilizar um serviço não são tratados necessariamente pela mesma autoridade provedora. Muitas vezes, são contratados servidores terceirizados para cuidar da manutenção dos equipamentos, o que implica que diferentes níveis de serviços têm que ser estabelecidos entre as partes envolvidas no contrato. Os níveis de serviços, por sua vez, devem ser previamente especificados e posteriormente poderão ser melhorados.

Na topologia convergente, quanto maior a diversidade de serviços associados, maior será a quantidade de elementos e a sua complexidade. Então, para que a convergência se torne cada vez mais abrangente, é importante apresentar soluções eficientes para implementação, controle, dimensionamento e funcionamento, a fim de resolver problemas relacionados ao projeto e gestão, bem como fornecer recursos adequados para garantir os serviços contratados em SLAs.

A ocorrência de eventos de falhas nos componentes da rede é um fator capaz de degradar seu desempenho, torná-la indisponível e inclusive causar prejuízos às organizações que delas necessitam. Assim, muitas empresas e instituições se preocupam em utilizar diferentes soluções e abordagens capazes de evitar a ocorrência de eventos de falhas, para manter e prolongar o funcionamento da rede. As soluções podem ser determinadas na fase de planejamento, possibilitando minimizar o número de reparos, os custos de operação, antecipar e evitar a ruptura dos serviços e melhorar o tempo de restabeleci-

mento dos serviços, seja por meio de um processo de coleta de eventos ou uma constante monitoração.

Da convergência para uma infraestrutura inteligente e altamente eficiente, derivarão acesso universal e uma gama de novas oportunidades para tecnologias, aplicações e serviços. Neste contexto, a tecnologia IP está se tornando a pedra angular para a qual várias tecnologias de acesso estão convergindo [Mis06], além de que os usuários podem obter diversas informações e desfrutar de vários serviços a qualquer hora e em qualquer lugar [ZLYY09].

## 1.1 MOTIVAÇÃO E JUSTIFICATIVA

Ultimamente, devido aos serviços baseados na Internet e sua necessidade de operação ininterrupta, questões relacionadas à dependabilidade, como alta disponibilidade, confiabilidade e manutenibilidade, estão bastante em foco, pois, caso esses serviços sofram paradas não esperadas, o seu fornecimento será comprometido, o que pode ocasionar prejuízos financeiros elevados.

O emprego de processos, técnicas, métodos e modelos são muito comuns em projetos de avaliação de dependabilidade de sistemas [Jai91], [Wil04], [MAD05], por já serem amplamente consolidados na literatura.

A avaliação de dependabilidade pode ser considerada uma atividade essencial para promover a melhoria da qualidade do serviço prestado e para o planejamento de infraestruturas. Além disso, é possível verificar atributos que não correspondem a níveis adequados de serviços e, então, propor melhorias para maximizar o tempo de atividade de sistemas em geral.

As atividades de melhoria envolvem os meios pelos quais a dependabilidade é atingida, a fim de encontrar soluções para prevenção, tolerância e remoção de falhas, que podem incluir verificação da importância da confiabilidade dos componentes, técnicas para a redundância de componentes e a adoção de políticas de manutenção. Essa última tem um impacto fundamental sobre a disponibilidade e confiabilidade de sistemas, pois permite estender a vida útil do equipamento ou, pelo menos, o tempo médio para a próxima falha.

Eventos de falhas e atividades de reparo de um componente específico podem não resultar na interrupção do serviço, mas são capazes de afetar o seu desempenho. Dessa forma, é de extrema importância a adoção de uma política de manutenção para evitar o risco de paradas não programadas.

A manutenção correta e planejada dos equipamentos e sistemas reduz custos com reparos, troca de peças, perda de horas de produção e permite que os equipamentos operem em plena carga e com qualidade. No entanto, as empresas e instituições brasileiras não têm disseminada a cultura da manutenção preventiva, que aparentemente pode apresentar custos elevados, mas que são compensados com a não parada dos componentes em momentos críticos.

A maioria das manutenções é realizada apenas quando os componentes falham, pois, ainda existe a impressão de que, assim, se gasta menos. Mas analisando de uma maneira mais abrangente, as ações corretivas têm um custo muito alto, uma vez que, componentes ou sistemas totalmente parados, em caráter emergencial, possuem diversos custos indiretos, o que acaba se tornando mais caro.

Devido às inúmeras vantagens oferecidas pelas redes convergentes, quando comparadas às redes convencionais, muitas empresas e instituições de ensino têm seus negócios e pesquisas baseados nos diversos serviços multimídia, tais como: correio eletrônico, mensagens instantâneas, VoIP, videoconferência, vídeo-aula, etc. No entanto, as questões não técnicas, envolvendo gestão e controle das redes convergentes tendem a ser difíceis.

Exigências de soluções de gerenciamento para lidar com a complexidade das redes convergentes são necessárias, pois, se a rede ficar indisponível, as operações podem ser interrompidas completamente, agregando prejuízos financeiros que podem ser elevadíssimos. Portanto, serviços de manutenção são essenciais para manter a rede em estado operacional, com níveis de disponibilidade e confiabilidade adequados.

## 1.2 TRABALHOS RELACIONADOS

Nos últimos anos, muitos trabalhos com diferentes abordagens e técnicas de modelagem têm sido conduzidos para lidar com questões relacionadas com a disponibilidade de redes de computadores. Pesquisadores têm utilizado abordagens diferentes para lidar com esses problemas, incluindo técnicas de análise de sensibilidade e o desenvolvimento de mecanismos de redundância. Os trabalhos descritos a seguir apresentam temas relacionados ao desta pesquisa.

Palade [Pal10] analisa a confiabilidade e a disponibilidade das redes de comunicações, com a suposição de que as falhas são simultâneas e dinâmicas, e podem ocorrer em ambos *hardware* e *software*. O modelo apresentado é uma combinação entre um modelo combinacional (Diagrama de Bloco de Confiabilidade - RBD) e um modelo de estado (Cadeias de Markov). O estudo de caso demonstra a aplicabilidade dos modelos para o projeto e apoio de uma rede IP, para melhor comunicação.

Guimarães *et al.* [GMMC11b] focam em aspectos de dependabilidade de redes redundantes de comunicação, usando modelos de Redes de Petri Estocásticas (SPN) para avaliação analítica de cenários complexos. Um estudo de caso foi realizado para avaliar a disponibilidade de redes empresariais em arquiteturas que consideram diferentes mecanismos de redundância. Uma combinação de SPN e RBD foi utilizada para analisar a disponibilidade do sistema de acordo com os componentes mais importantes. Também são mostradas as variações da disponibilidade do sistema em conformidade com o Tempo Médio para Falha (MTTF) em diferentes arquiteturas.

Em outro trabalho de Guimarães *et al.* [GMMC11a], os autores construíram diferentes cenários para calcular variáveis e avaliar o impacto da disponibilidade do sistema

em infraestrutura de redes de computadores, usando um modelo de desempenho SPN para analisar a variação do tráfego em cenários reais. O objetivo era determinar o nível adequado de disponibilidade do sistema e o Custo Total de Propriedade (TCO - *Total Cost Ownership*). Além disso, em Matos *et al.* [MGC<sup>+</sup>11], uma abordagem de modelagem em cadeias de Markov analisa a sensibilidade e tempos de reparo de diferentes componentes para verificar a disponibilidade de redes de computadores com mecanismos de redundância.

Zou *et al.* [ZJKK07] discutem métodos algorítmicos para calcular a disponibilidade da rede para uma dada topologia e apresentam duas ferramentas para o cálculo de disponibilidade de rede em redes grandes e complexas. Semaan *et al.* [Sem08] discutem diversos temas relacionados com a disponibilidade da rede. Além disso, os autores apresentam alguns dos elementos que impactam a disponibilidade e discutem como os projetistas de redes podem calcular a disponibilidade, oferecendo meios para determinar o nível ótimo de disponibilidade.

Trivedi *et al.* [TKRM09a] apresentam uma nova classificação de modelos de segurança e de dependabilidade para sistemas e redes. Adicionalmente, os autores apresentam vários tipos de modelos individuais, como modelos de disponibilidade, confidencialidade, integridade, desempenho, confiabilidade, capacidade de sobrevivência, segurança e manutenção. E ainda, mostram que os tipos de modelo individuais podem ser combinados para formar modelos de dependabilidade. Os modelos de confiabilidade e segurança podem ser representados como modelos combinatoriais, modelos de espaço de estados e modelos hierárquicos.

Sesmun e Turner [ST00] analisam desempenho e confiabilidade de redes de comunicação, a fim de criar uma metodologia para projeto de redes tolerantes a falhas, capazes de oferecer o melhor desempenho médio ao longo de um período de tempo. Os resultados obtidos mostram que a utilização da metodologia em comparação com os métodos de concepção convencional gera uma rede de comunicação mais confiável.

Diferentemente dos trabalhos apresentados, este trabalho visa fornecer modelos para representação de infraestruturas de redes convergentes para instituições de ensino, aplicando uma política de manutenção que envolve diferentes abordagens para garantia de níveis de serviços. A modelagem e verificação da disponibilidade são características de interesse comum em relação aos trabalhos citados. Como resultado deste, espera-se que além da conscientização da importância da realização de manutenções preventivas, as instituições de ensino possam investir em uma política de manutenção para melhores condições de funcionamento e elevada produtividade.

Apesar de o estudo ser demonstrado para uma infraestrutura particular, a abordagem é genérica o suficiente para avaliar várias outras infraestruturas. Mas é necessário encontrar uma solução que seja apropriada para resolução do problema que se deseja resolver. Assim, adaptações sobre técnicas, abordagens e os modelos aqui apresentados podem ser necessárias a uma solução que seja eficiente o bastante para extrair detalhes e características do sistema envolvido e possa gerar um resultado que seja apropriado.

### 1.3 OBJETIVOS

Este trabalho propõe modelos para avaliar o impacto de políticas de manutenção sobre a dependabilidade de redes convergentes. Especificando-se melhor, o trabalho ora posto, possui os seguintes objetivos:

- Propor modelos hierárquicos e heterogêneos para infraestrutura de redes convergentes;
- Avaliar o impacto de uma política de manutenção, considerando abordagens de manutenções preventivas e corretivas.

### 1.4 ESTRUTURA DA DISSERTAÇÃO

A estrutura da dissertação é detalhada a seguir: O Capítulo 2 apresenta vantagens da utilização de redes convergentes para instituições de ensino. Também, são definidas técnicas para modelagem de sistemas, que incluem cadeias de Markov de tempo contínuo, redes de Petri estocásticas e diagramas de blocos de confiabilidade. Em seguida, são dadas definições para dependabilidade, política de manutenção e acordo de nível de serviços. Essa fundamentação é essencial para melhor entendimento deste estudo. O Capítulo 3 apresenta a metodologia, a descrição da rede utilizada para o estudo de caso e os modelos concebidos para avaliação de dependabilidade. O Capítulo 4 apresenta o estudo de caso, baseados nos modelos propostos, para verificação de diferentes níveis de serviços e avaliação do impacto da política de manutenção. Por fim, o Capítulo 5 apresenta as conclusões obtidas com o desenvolvimento deste trabalho, assim como as principais contribuições e, posteriormente, são apresentadas propostas para trabalhos futuros.

## CAPÍTULO 2

# FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma introdução sobre redes convergentes para instituições de ensino, apresentando exemplos de aplicações e vantagens. Em seguida, na seção 2.2 são introduzidas as técnicas para modelagem de sistemas, tais como: cadeias de Markov de tempo contínuo (CTMC), redes de Petri (PN), em especial redes de Petri estocásticas (SPN), e diagramas de blocos de confiabilidade (RBD). Na seção 2.3 são abordados atributos de dependabilidade de interesse deste trabalho. Política de manutenção e suas abordagens são apresentadas na seção 2.4. A seção 2.5 explica o que são acordos de níveis serviços. Por fim, algumas considerações finais são apresentadas.

### 2.1 REDES CONVERGENTES PARA INSTITUIÇÕES DE ENSINO

Devido à flexibilidade, programabilidade e custo relativamente baixo, a Internet está modificando a maneira como se faz educação, permitindo facilidades para pesquisas *online* e a utilização de conteúdos multimídias e interativos. Porém, algumas limitações como estreita largura de banda diminui a sua usabilidade. Para tentar resolver este problema, recentes avanços nas tecnologias de transmissão de dados e investimentos por parte de empresas e governos visam proporcionar alta velocidade de transmissão de dados, através da Internet banda larga, que por sua vez, tende a alcançar no curto e médio prazo o consumidor final.

Com o aumento na exigência por formação ou qualificação e a diversidade das novas tecnologias no mercado, é importante que as empresas e comunidades científicas estejam sempre atentas para identificar os impactos e as oportunidades para a formação profissional, principalmente, no que diz respeito ao processo educacional. Nesse contexto, as redes convergentes podem ser aliadas à educação, pois utilizam sistemas capazes de integrar comunicação de voz, vídeo e aplicações centradas na web e na rede local de dados, juntamente com mobilidade e conectividade, apoiando ainda, o conhecimento através da modalidade de educação à distância (EAD).

A convergência possibilita comunicação unificada e engloba uma variedade de tecnologias, mídias, meios e dispositivos de comunicação, permitindo que os usuários se comuniquem em tempo real com qualquer pessoa em qualquer lugar. Assim, muitas instituições de ensino objetivam aprimorar e ampliar o processo de ensino, difundindo a prática de multimídia, teleconferência e interação didática; valorizando a capacidade de produção, transmissão e consumo de novos conhecimentos por meio das redes convergentes.

A comunicação unificada (convergência) é uma mudança tecnológica da qual as instituições de ensino podem participar para ampliar a área de ensino e atingir patamares

nunca alcançados, garantindo um aprendizado no qual o limite de tempo e espaço se concentre até aonde a tecnologia pode chegar [Oli11].

A convergência para educação reúne aplicações de correio eletrônico, VoIP, videoconferência, videoaula, mensagem instantânea, fórum de discussão, espaço *wiki*, *chat* e *blog*, as quais podem auxiliar as instituições educacionais, impulsionando diversas funções tecnológicas, tais como: reuniões virtuais, gravação de chamadas, roteamento por nível de serviço e notificação de solução de pedidos. Esses recursos contribuem largamente para as demais áreas, principalmente do ensino à distância.

Adicionalmente, as redes convergentes integram tempo e espaço, em que professor e aluno, ligados através das hipermídias unificadas podem realizar seus trabalhos como se estivessem no mesmo ambiente, bem como em tempo distinto. Projetos que envolvem a construção de ambientes integrados de comunicação aliados ao ensino trazem vantagens para o meio educacional, como aumento da produtividade e da satisfação dos alunos, redução de custos com ligações, ampliação da mobilidade EaD, facilidade para trabalhar remotamente, etc.

No segmento de educação, a convergência tem contribuído de forma valiosa para a melhoria e disponibilidade das redes institucionais, promovendo infraestrutura adequada para o acesso de alunos, professores, administradores e visitantes das instituições, de forma a integrar diversos campi. Este cenário se torna o diferencial, contribuindo fortemente com a capacitação e formação profissional e acadêmica.

As instituições de ensino, em geral, se encontram em diferentes estágios em relação à convergência, algumas já adotaram suas redes convergentes e outras se defrontam com implicações administrativas e organizacionais. Mas ainda é preciso melhores transformações no meio corporativo para agregar valor à educação moderna. Entretanto, os responsáveis podem se apoiar em melhores práticas e métodos capazes de ajudar à essa transição [San08].

Algumas instituições de ensino modernizaram suas infraestruturas de comunicação, possibilitando tráfego convergente, a exemplo, tem-se a Universidade do Oeste Paulista (UNOESTE), de Presidente Prudente (interior de São Paulo), e a Universidade Católica de Santos (UNISANTOS), na cidade de Santos. Essas instituições fizeram investimentos para suportar a convergência de voz, vídeo e dados, criando, assim, uma base sólida para o futuro. Seus projetos foram feitos pela empresa *TDec Network Group*, presente no mercado brasileiro desde 1991, que provê soluções inovadoras de infraestrutura de redes IP [TDe12]. Este trabalho adota para estudo de caso, uma infraestrutura de redes convergentes de instituições de ensino que interligam quatro instituições do nordeste brasileiro, separadas geograficamente.

## 2.2 TÉCNICAS PARA MODELAGEM ANALÍTICA DE DEPENDABILIDADE

A representação de sistemas por meio de técnicas de modelagem permite a obtenção de informações úteis sobre a estrutura e o comportamento dinâmico do sistema, fornecendo,

assim, suporte para análise de dependabilidade. No entanto, diferentes níveis de abstração podem ser alcançados na modelagem dos sistemas. Porém, cada nível é mais adequado para a resolução de um problema específico [Jai91] e deve ser escolhido aquele que melhor representa o problema em questão. Os modelos, por sua vez, podem apresentar-se como protótipos ou como modelos matemáticos e as técnicas de modelagem podem ser classificadas como técnicas analíticas e técnicas baseadas em simulação [Lil00].

As simulações baseiam-se em modelos abstratos do sistema, logo, não exigem que o sistema esteja totalmente implantado para que sejam aplicadas. 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 podem variar para diferentes sistemas. Durante a simulação, controlam-se com maior eficiência os valores assumidos por parâmetros [Lil00].

Já a modelagem analítica utiliza um conjunto de equações e funções matemáticas para descrever o comportamento de um sistema, considerando determinados parâmetros que podem ser adaptados para outros. Além disso, a modelagem analítica permite uma análise ampla e aprofundada em relação aos efeitos causados pelos parâmetros definidos nas equações sobre a aplicação, apresentando menor custo de execução [Lil00].

Algumas ferramentas acadêmicas e comerciais permitem a modelagem de sistemas e a descrição de seu comportamento, seja através de Diagramas de Bloco de Confiabilidade (*Reliability block diagram* - RBD), Grafo de Alcançabilidade (*Reliability graph* - RG), Cadeias de Markov (*Markov Chain*), Redes de Petri Estocásticas (*Stochastic Petri net* - SPN) ou outras. Dentre as várias ferramentas existentes para modelagem de sistemas, destacam-se: Sharpe [STP96], Block-Sim [Rel10], TimeNet [Zim12], Asto Tool [SMT<sup>+</sup>10], GreatSPN [CFGR95] e SPNP [CMT89].

Os modelos criados para realização deste trabalho compreendem a diagramas de blocos de confiabilidade, cadeias de Markov de tempo contínuo e redes de Petri estocásticas, utilizando as ferramentas Sharpe [STP96], Asto Tool [SMT<sup>+</sup>10] e TimeNet 4.1 [Zim12]. A seguir, são apresentados alguns conceitos sobre cadeias de Markov, redes de Petri, com foco em redes de Petri estocásticas e diagramas de blocos de confiabilidade.

### 2.2.1 Cadeias de Markov

Uma cadeia de Markov é um processo estocástico que apresenta a propriedade markoviana em que os estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido. Um processo estocástico proporciona uma relação entre os elementos de uma família de variáveis aleatórias, em que  $X(t), t \in T$  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  $X(t_0), X(t_1), \dots, X(t_{n-1})$  [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.

Uma maneira simples de visualizar um tipo específico de cadeia de Markov é através de uma máquina de estados finitos. Se algo está no estado  $y$  no tempo  $n$ , então a probabilidade de que se mova para o estado  $x$  no tempo  $n + 1$  não depende de  $n$ , e somente depende do estado atual  $y$  em que está. Assim, em qualquer tempo  $n$ , uma cadeia de Markov finita pode ser caracterizada por uma matriz de probabilidades, cujo elemento  $(x, y)$  é dado por  $P(X_{n+1} = x | X_n = y)$ , e é independente do tempo  $n$ . Estes tipos de cadeia de Markov finitas e discretas podem também ser descritas por meio de um grafo dirigido, onde cada aresta é rotulada com as probabilidade de transição de um estado a outro, sendo esses estados representados como os nós conectados pelas arestas [MAMF95, KM86, CL08, BM02].

A propriedade markoviana (ausência de memória) destaca que eventos futuros não estão condicionados a eventos passados, ou seja, são dependentes apenas do estado presente [MAMF95]. Em outras palavras, o sistema, tendo a propriedade acima, o passado afeta o futuro somente através do presente; ou, afirmado ainda de outra maneira, o estado atual do sistema contém todas as informações relevantes necessárias para produzir o futuro em sentido probabilístico [Gal09].

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. 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 (2.1).

$$Q = \begin{pmatrix} q_{ii} & q_{ij} \\ q_{ji} & q_{jj} \end{pmatrix} \quad (2.1)$$

As cadeias de Markov nos permitem descrever arquiteturas e sistemas complexos, sem restrições nos esquemas de sincronização envolvidos. No entanto, para o uso de cadeias de Markov explicitamente na modelagem, precisamos enumerar todos os possíveis estados do sistema e todas as possíveis transições entre eles. Essa abordagem, além de ser muito suscetível a erros, se torna inviável para sistemas com mais de uma centena de estados.

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*). Na sequência só será feita a discussão de cadeias de Markov de tempo contínuo, utilizadas neste trabalho.

Para os modelos em CTMC, a matriz de taxas, também denominada de gerador infinitesimal, cada elemento não diagonal da linha  $i$  e coluna  $j$  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 zero.

Em uma CTMC, as probabilidades de transição de estado são calculadas por meio da Equação (2.2).

$$p_{i,j}(s, t) = P\{X(t) = j | X(s) = i\} \quad (2.2)$$

O comportamento do processo markoviano é descrito pela Equação de probabilidade de transição de Chapman-Kolmogoroff (2.3). Nessa equação,  $i$ ,  $r$  e  $j$  são estados;  $s$ ,  $u$ , e  $t$ , instantes de tempo [Her01].

$$p_{i,j}(s, t) = \sum_{\text{todor}} p_{i,r}(s, u) \cdot p_{r,j}(u, t), s \leq u \leq t \quad (2.3)$$

Desenvolvendo-se essa equação (na notação matricial), obtém-se a Equação diferencial (2.4).

$$\frac{d\pi(t)}{dt} = \pi(t)Q \quad (2.4)$$

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, 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), consequentemente,  $\pi'(t) = 0$ , resultando nas Equações (2.5) e (2.6).

$$\sum_{i=1}^N \pi_i = 1 \quad (2.5)$$

$$\pi Q = 0 \quad (2.6)$$

A Equação (2.5) é a condição de normalização, adicionada para assegurar que a solução obtida é um único vetor de probabilidade. A Equação (2.6) tem um conjunto de soluções indeterminadas. Normalizando as soluções, chega-se a um único vetor de probabilidades.

### 2.2.2 Redes de Petri

Redes de Petri (PN) constituem um formalismo matemático, que apresenta uma notação gráfica capaz de representar sistemas de forma intuitiva, exibindo características de concorrência e sincronização. Segundo Peterson [Pet77], o principal uso das redes de Petri é a modelagem de sistemas de eventos em que é possível que alguns eventos ocorram simultaneamente, mas há restrições sobre a precedência ou a frequência dessas ocorrências.

A representação formal de um modelo PN é dado pela 5-tupla  $PN = \{P, T, F, W, \mu_0\}$ , onde:

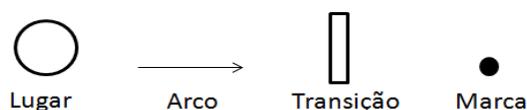
- $P$  é o conjunto finito de lugares;
- $T$  é o conjunto finito de transições,  $P \cap T = \emptyset$ ;
- $F \subseteq (P \times T) \cup (T \times P)$  é o conjunto de arcos;
- $W : F \rightarrow \mathbb{R}^+ \cup \{0\}$  é a função de atribuição de peso aos arcos;
- $\mu_0 : P \rightarrow \mathbb{N}$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

As redes de Petri foram originalmente desenvolvidas e usadas para o estudo das propriedades qualitativas de sistemas. A PN original não tinha qualquer noção de tempo, seu objetivo era desenvolver um modelo em que as máquinas de estado fossem capazes de se comunicar. Elas são assim chamadas em homenagem ao seu criador, Carl Adam Petri [Pet66].

Redes de Petri permitem a modelagem e análise de sistemas de eventos discretos que são demasiado complexo para serem descritos por autômatos ou modelos de filas [Rei85]. A aplicabilidade das Redes de Petri como ferramenta para estudo de sistemas é muito importante, pois, além de permitir representação matemática e a análise dos modelos, também pode fornecer informações úteis sobre a estrutura e o comportamento dinâmico dos sistemas modelados.

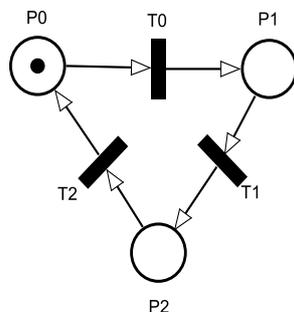
Atualmente, é possível encontrar variações que surgiram a partir do modelo original. Estas são usadas em diversas áreas (sistemas de manufatura, desenvolvimento de *software*, sistemas administrativos, entre outras) para ajudar no estudo do comportamento e desempenho de diferentes sistemas.

O aspecto estrutural de um modelo PN equivale a um grafo direcionado composto por alguns elementos básicos, conhecidos como estados (lugares), ações (transições), arcos e marcas (*tokens*). Os lugares correspondem às variáveis de estado, as transições às ações ou eventos realizados pelo sistema e os arcos ligam os lugares às transições e vice-versa. Observe na Figura 2.1 como são representados graficamente os elementos que compõem uma rede de Petri.



**Figura 2.1:** Elementos de uma Rede de Petri

A realização de uma ação está associada a alguma pré-condição, ou seja, existe uma relação entre os lugares e as transições que possibilitam ou não a realização de uma determinada ação. Após a realização de uma determinada ação, alguns lugares terão suas informações alteradas e podem criar uma pós-condição. Os arcos representam o fluxo das marcas pela rede e as marcas representam o estado em que o sistema se encontra em determinado momento. A Figura 2.2 mostra um exemplo de rede de Petri com marcação inicial no lugar  $P_0$ .



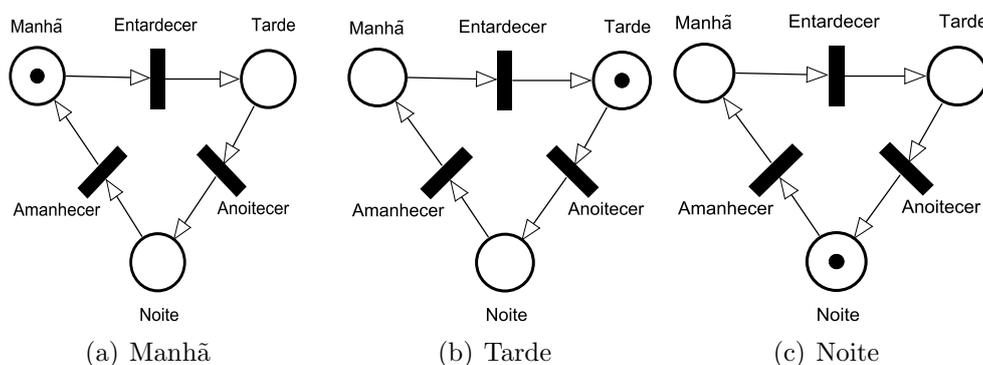
**Figura 2.2:** Rede de Petri

Considerando a Figura 2.2, inicialmente a transição  $T_0$  está habilitada, isto é, pode ser disparada devido à presença de uma marca no lugar  $P_0$ . Após o disparo da transição  $T_0$ , a marca será retirada do lugar  $P_0$  e será depositada no lugar  $P_1$ , habilitando, assim, a transição  $T_1$ . Após o disparo da transição  $T_1$ , a marca será inserida no lugar  $P_2$ , o que habilita a transição  $T_2$ . Quando a transição  $T_2$  for disparada, a rede volta a ter a marca no lugar  $P_0$ .

Para exemplificar melhor, a Figura 2.3 [MLC96] representa 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 manhã, o anterior à noite é 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 foram criados três lugares, os quais representam os três períodos do dia, e três transições, as quais representam as três mudanças de período do dia, observe a Figura 2.3.

O modelo gerado tem o seu estado inicial com uma marca (*token*) no lugar Manhã (Figura 2.3(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 (observe a Figura 2.3(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 2.3(c)). Uma marca no lugar Noite possibilita a ocorrência do evento amanhecer, este representado pela transição amanhecer. A execução desse evento reinicia o processo.



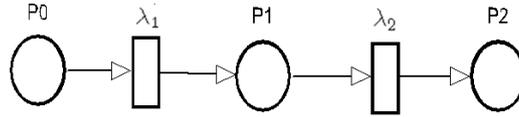
**Figura 2.3:** Períodos do Dia  
Adaptado de [MLC96]

Inicialmente, os modelos de redes de Petri não incluíam nenhuma noção de tempo devido ao efeito que o tempo pode ter sobre o comportamento. Depois, algumas extensões hierárquicas surgiram e adicionaram tempo ao formalismo de Petri. Várias são as extensões que permitem uma conveniência gráfica e de fácil entendimento para o usuário, entre elas estão as Redes de Petri Temporizadas (TPN), Redes de Petri Coloridas (CPN), Redes de Petri Estocásticas (SPN) e Redes de Petri Estocásticas Generalizadas (GSPN). Grupos de pesquisas, em todo o mundo, têm Redes de Petri como tema, estes, por sua vez, vêm desenvolvendo estudos sobre seus aspectos teóricos e suas aplicações [MLC96].

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.

A Figura 2.4 descreve uma *throughput subnet* formada por duas transições exponenciais em série com os parâmetros  $\lambda_1$  e  $\lambda_2$ , 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 (2.7).



**Figura 2.4:** Conexão Série com Erlang

$$f_{\tau}(t) = (f_{\tau_1} * f_{\tau_2})(t) = \frac{\lambda_1 \lambda_2 (\exp^{-\lambda_1 t} - \exp^{-\lambda_2 t})}{\lambda_2 - \lambda_1}, t \geq 0 \quad (2.7)$$

\* é o operador de convolução. Para o caso onde  $\lambda_1, \lambda_2 = \dots = \lambda_n$ , a função densidade é dada pela Equação (2.8).

$$f_{\tau}(t) = \frac{\lambda^n t^{n-1} \exp^{-\lambda t}}{(n-1)!}, t > 0 \quad (2.8)$$

Essa expressão representa uma distribuição do tipo Erlang de ordem  $N$ . Uma distribuição do tipo Erlang é especificada por dois parâmetros  $\lambda > 0$  e  $n > 0$ . O modelo SPN gerado para este trabalho, apresenta uma Erlang de duas fases para representar os tempos de falha de cada *site* (Ver Seção 3.4.4).

### 2.2.2.1 Propriedades das Redes de Petri

A análise do modelo pode ser realizada através do estudo das propriedades comportamentais (dependentes de marcação inicial) e das propriedades estruturais (não dependentes de marcação) [Mur89, MLC96].

As propriedades comportamentais são aquelas que dependem apenas da marcação inicial da rede de Petri. Para análise comportamental, são verificadas as propriedades de alcançabilidade, limitação, segurança, *liveness*, cobertura, persistência, reversibilidade e justiça. Descritas a seguir:

- 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, leva 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 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  $CA(R; M_0) = \{M' \in IN^m | \exists s, M_0[s > M']\}$ , onde  $m$  é a cardinalidade do conjunto de lugares da rede.

A análise da alcançabilidade de uma marcação consiste em determinar uma dada marcação  $M' \in CA(R; M_0)$  da rede marcada  $RM$ . Em alguns casos, desejam-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*.

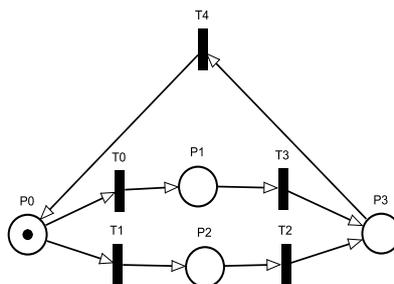
Uma rede  $RM = (R; M_0)$  é viva (*live*) se, para toda  $M \in CA(R; M_0)$ , for possível disparar qualquer transição de  $RM$  através do disparo de alguma sequência de transições.

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

- Uma rede é dita persistente se, para qualquer par de transições habilitadas, o disparo de uma transição não desabilita o disparo da outra transição, ou seja, esta transição continua habilitada até o seu disparo.
- Uma rede é reversível se, para cada marcação  $M_i$ , no conjunto das marcações acessíveis, a marcação inicial pode ser novamente alcançada.
- O conceito de justiça ou *fairness* apresenta diferentes pontos de vista, os dois principais são: justiça limitada (*bounded-fairness* ou *B-fair*) e justiça incondicional. Sob o ponto de vista de justiça limitada, duas transições  $t_i$  e  $t_j$  são classificadas como *B-fair*, se o número de vezes que uma delas dispara, enquanto a outra não dispara, é limitado. Já sob o ponto de vista de justiça incondicional, uma sequência de transições  $s_1$  é classificada como *fair* incondicional, se essa sequência é finita ou se todas as transições da rede aparecem um número infinito de vezes nessa sequência.

Já as propriedades que estão apenas relacionadas com a estrutura do modelo incluem limitação estrutural, conservação, repetitividade e consistência [MLC96].

- Uma rede de Petri  $R = (P, T, F, W, \mu_0)$  é classificada como estruturalmente limitada, se for limitada para qualquer marcação inicial.
- A conservação é uma importante propriedade das redes de Petri permitindo, por exemplo, a verificação da não destruição de recursos através da simples conservação de marcas. Na figura 2.5 é apresentada uma rede na qual observamos esta propriedade. O disparo de qualquer transição desta rede não altera o número de marcas, ou seja, recursos não são criados e nem destruídos.



**Figura 2.5:** Rede Estritamente Conservativa  
Adaptado de [MLC96]

- Uma rede marcada é classificada como repetitiva se, para uma marcação e uma sequência de transições disparáveis, para esta marcação, todas as transições da rede são disparadas ilimitadamente. Seja  $N = (R; M_0)$  uma rede marcada e  $s$  uma sequência de transições.  $N$  é dita repetitiva, se existe uma sequência  $s$ , tal que  $M_0[s > M_i$  e toda  $t_i \in T$  dispara um número infinito de vezes em  $s$ .

- 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, no qual 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].

### 2.2.2.2 Rede de Petri Estocástica

As Redes de Petri Estocásticas (SPN) são uma subclasse das Redes de Petri, usadas para a descrição de sistemas dinâmicos de eventos discretos, cujo comportamento dinâmico pode ser representado por meio de Cadeias de Markov de Tempo Contínuo (CMTC) [FFS91].

Em 1982, M. K. Molloy [Mol82] apresentou as redes de Petri estocásticas (*Stochastic Petri Nets* - SPN) como uma técnica capaz de especificar sistemas e apresentar uma análise probabilística dos mesmos. Elas surgiram a partir do formalismo de Redes de Petri Temporizadas (TPN), que tem como sua principal característica a associação de um atraso fixo para cada transição do modelo. Molloy [Mol82] definiu que todas as transições em uma SPN eram temporizadas (*timed*) e que possuíam um retardo exponencialmente distribuído.

As SPNs oferecem possibilidade de unir a habilidade do formalismo de Redes de Petri para descrever sincronização e concorrência com um modelo estocástico, permitindo a descrição de um comportamento dinâmico na modelagem de desempenho (*performance*) e dependabilidade (*dependability*) de sistemas [MAMC86].

Uma SPN é definida [GKZH95] 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 à marcação do lugar a 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  $\lambda_t$  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  $\lambda_t$  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.

Alguns formalismos permitem tanto análises numérica quanto simulação, e SPN é um dos modelos mais importantes de tal classe. A escolha da granularidade determina as análises que podem ser realizadas, dependendo do detalhamento dos componentes do sistema [MA05].

Nas SPNs, quando múltiplas transições estão habilitadas em uma mesma marcação  $m$ , a transição que tem a maior probabilidade de disparo é a transição que possuir menor tempo de atraso associado a ela [Bal01]. Quando uma transição de uma rede de Petri estocástica é disparada, assim como no formalismo de redes de Petri, uma nova marcação pode ser gerada. Essa nova marcação pode conter transições que já se encontravam habilitadas na marcação anterior, mas não foram disparadas. Por causa da propriedade de ausência de memória (*memorelles* - não utilização da memória do passado) da distribuição exponencial, pode-se assumir que a atividade associada a cada transição é recomeçada para qualquer nova marcação.

O comportamento das variáveis pode ser descrito através de um processo estocástico, juntamente com a utilização de *memorelles*. Dessa forma, se um evento produzir um disparo da transição  $t$  e transformar a marcação  $M1$  em  $M2$ , a evolução futura das transições que estavam sensibilizadas por  $M1$  antes do disparo de  $t$  deverá ser idêntica àquela que as transições sofreriam se viessem a ser sensibilizadas por  $M2$ . Somente as distribuições geométricas e exponenciais verificam este fato.

As redes de Petri estocásticas são as redes em que as durações de sensibilização associadas às transições são definidas por distribuições geométricas e exponenciais, a fim de

poder construir um processo markoviano equivalente e, assim, analisar o comportamento da rede [CV97]. Neste sentido, as SPNs são isomórficas as cadeias de Markov de tempo contínuo [Mur89].

As transições possuem tempo, disparo e retardo de transição através de variáveis aleatórias distribuídas exponencialmente. Esta característica de tempo nas transições permite diferentes tipos de funcionamento para o disparo, entre eles, o disparo em três fases e o disparo atômico.

O disparo em três fases supõe que as marcas (*tokens*) são consumidas do lugar de entrada quando a transição é habilitada e que as marcas são produzidas no lugar de saída após a decorrência do intervalo de tempo associado à transição. Já o disparo atômico supõe que as marcas são consumidas do lugar de entrada e geradas no lugar de saída apenas após o disparo da transição.

As transições em SPNs podem ser imediatas ou temporizadas. As transições imediatas possuem tempos associados, que são exponencialmente distribuídos. As transições temporizadas possuem tempo associado igual a zero, nessas, o período de habilitação corresponde ao período de execução da atividade e o disparo corresponde ao término da atividade. Diferentes níveis de prioridade podem ser atribuídos às transições. A prioridade de disparo das transições imediatas é superior à das transições temporizadas. As probabilidades de disparo associadas às transições imediatas podem solucionar situações de conflito [Bal01, MAMF95].

As transições temporizadas podem ser caracterizadas por diferentes políticas de memória tais como *Resampling*, *Enabling memory* e *Age memory* [MAMC86], descritas a seguir:

- *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 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 em que 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* [MAMC86], que possuem as seguintes definições:

- *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.
- *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.
- *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 [MAMC86]. A construção de modelos SPN realistas serve para avaliar sistemas e encontrar métricas de dependabilidade e desempenho. Os modelos SPN possuem dois tipos de estados: os estados voláteis (*vanish*) e os estados tangíveis (*tangible*).

Os estados voláteis são criados em decorrência da marcação dos lugares que são pré-condições de habilitação de uma transição imediata. O termo volátil 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 [MAMF95].

A vantagem oferecida para uso de modelos estocásticos está na possibilidade de adicionar valores para determinados parâmetros de entrada, a fim de obter valores de saída. Além disso, há ainda a simplicidade oferecida por ferramentas de interface amigável para modelagem de sistemas, como por exemplo, a ferramenta TimeNet [GKZ<sup>+</sup>95, Zim12].

### 2.2.3 Diagramas de Bloco de Confiabilidade

Diagrama de bloco de confiabilidade (*Reliability Block Diagram - RBD*) é uma das técnicas mais usadas para a análise de confiabilidade de sistemas. Sua grande vantagem é a facilidade de analisar a confiabilidade de sistemas [XPD04].

A chamada Importância da Confiabilidade (*Reliability Importance - RI*) tem como objetivo identificar a fraqueza em componentes e quantificar o impacto de falhas através da análise da confiabilidade [KZ03]. O estudo da RI possibilita identificar quais componentes têm maior importância para a confiabilidade do sistema, cujo componente com maior RI é o que apresenta maior risco para o sistema, caso sua falha ocorra.

Um RBD serve para representar as conexões lógicas de componentes necessários para cumprir uma função do sistema especificado através de blocos de subsistemas ou componentes ligados de acordo com suas funções ou uma relação de confiabilidade [RH04a]. Além disso, permite encontrar instalações ou equipamentos que podem causar falhas no sistema. Dessa forma, fica mais fácil propor melhorias como equipamentos de apoio, redundância ou substituição de dispositivos.

Através de modelos RBD, é possível representar um componente físico em modo de operação por um bloco, estimar a confiabilidade de cada bloco individualmente, representar uma falha de um componente, removendo o bloco correspondente e, assim, calcular a confiabilidade de um sistema. Se os blocos são removidos em um RBD para interromper a ligação entre pontos de entrada e saída, o sistema falha [KZ03].

RBD é um formalismo usado para calcular métricas de dependabilidade, tais como disponibilidade, confiabilidade e manutenibilidade. Modelos RBDs contêm uma entrada (ou fonte - no lado esquerdo) e uma saída (ou destino - do lado direito). Entre os nós de entrada e saída, o sistema pode ser constituído por estruturas de blocos em série, paralelo, ponte (*bridge*), blocos  $k$  de  $n$  (*k-out-of-n*) de estruturas de rede em geral ou combinações destas estruturas [KZ03].

Um diagrama que tem componentes conectados em série exige que cada componente esteja funcionando para ter o sistema em estado operacional. Um diagrama que tem componentes conectados em paralelo exige que apenas um componente esteja funcionando para que o sistema seja operacional [TM93]. 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 confiabilidade de dois blocos conectados em série é obtida através da Equação (2.9).

$$R_S = R_1 \times R_2 \quad (2.9)$$

onde:

$R_1$  descreve a confiabilidade do bloco 1.

$R_2$  descreve a confiabilidade do bloco 2.

Para estruturas em série com  $n$  componentes, a confiabilidade do sistema é representada pela Equação 2.10.

$$R_s(t) = \prod_{i=1}^n R_i(t) \quad (2.10)$$

Onde  $R_i(t)$  corresponde à confiabilidade do bloco  $b_i$  no instante de tempo  $t$ . Similarmente, outras métricas probabilísticas podem ser calculadas para estruturas em série.

A confiabilidade de dois blocos conectados em paralelo é obtida através da Equação (2.11).

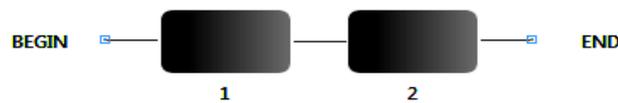
$$R_P = 1 - \prod_{i=1}^2 (1 - R_i) \quad (2.11)$$

A confiabilidade de  $n$  blocos conectados em paralelo é obtida através da Equação (2.12).

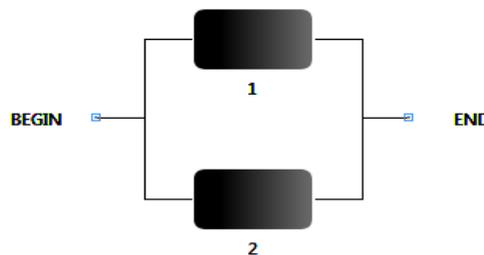
$$R_P = 1 - \prod_{i=1}^n (1 - R_i(t)) \quad (2.12)$$

Onde  $R_i(t)$  corresponde à confiabilidade do bloco  $b_i$  no instante de tempo  $t$ .

A Figura 2.6 mostra a conexão de blocos em série (Figura 2.6(a)) e a conexão de blocos em paralelo (Figura 2.6(b)).



(a) Série

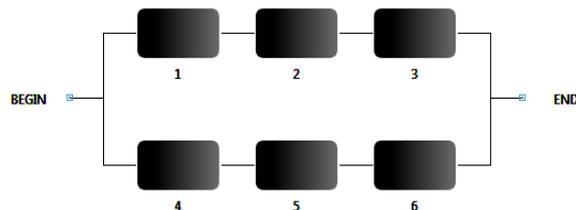


(b) Paralelo

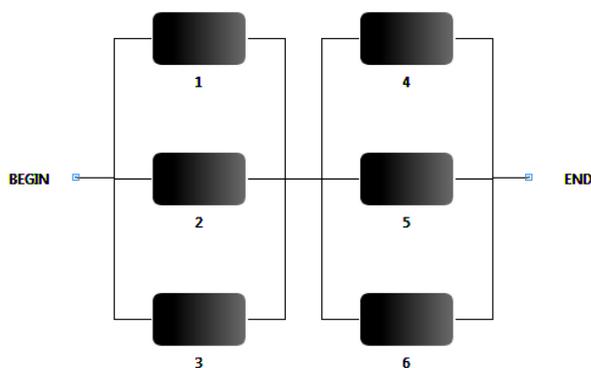
**Figura 2.6:** Diagramas de Blocos

Blocos conectados em série e blocos conectados em paralelo podem ser combinados para formar um novo bloco. A Figura 2.7 apresenta uma combinação série-paralelo e

a Figura 2.8 apresenta uma combinação paralelo-série. A confiabilidade desses blocos é obtida através das Equações (2.10) e (2.12).

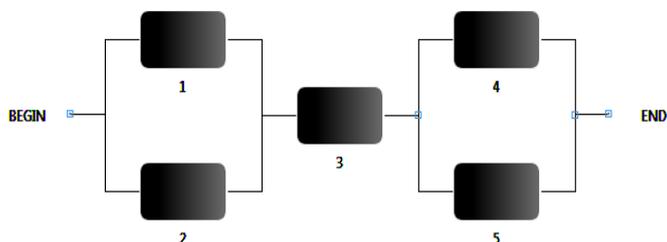


**Figura 2.7:** Diagrama de Blocos Série-Paralelo



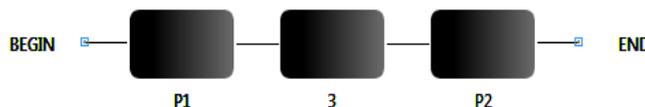
**Figura 2.8:** Diagrama de Blocos Paralelo-Série

A Figura 2.9 mostra a conexão de blocos em série e paralelo, que são combinados para representar um sistema. Os blocos 1 e 2 conectados em paralelo são combinados e formam o bloco  $P1$  e sua confiabilidade é obtida por meio da Equação (2.13). Os blocos 4 e 5 são combinados e formam o bloco  $P2$ , cuja confiabilidade é obtida através da Equação (2.14). A Figura 2.10 mostra o resultado da combinação desses blocos.



**Figura 2.9:** Diagrama de Blocos Série-Paralelo (Possível de Combinação)

$$R_{P1} = 1 - \prod_{i=1}^2 (1 - R_i) \quad (2.13)$$



**Figura 2.10:** Diagrama de Blocos Série-Paralelo (Resultado da Combinação)

$$R_{P2} = 1 - \prod_{i=4}^5 (1 - R_i) \quad (2.14)$$

Blocos  $k - out - of - n$  representam estruturas em que o subsistema pode funcionar se  $k$  ou mais componentes estão no estado operacional [XPD04]. Por exemplo, uma estrutura em que haja cinco componentes e necessita-se de três funcionando para prover o serviço esperado, temos uma estrutura  $3-out-of-5$  (ou 3 de 5). As estruturas em série e paralela são casos especiais de estruturas  $k - out - of - n$ , uma estrutura em série é uma  $n - out - of - n$  e uma estrutura em paralelo é uma estrutura  $1-out-of-n$  [KZ03, ST87]. Para a definição matemática da confiabilidade deste arranjo lógico, é necessária a definição da variável aleatória discreta  $X$ , que define o número de blocos que não apresenta falhas, em um determinado intervalo de tempo. Os eventos probabilísticos de dependabilidade são independentes para cada bloco da configuração  $k$  de  $n$  e todos os  $n$  blocos possuem a mesma taxa de falha [XPD04].

RBDs são utilizados, principalmente, em sistemas modulares que consistam de muitos módulos independentes, em que cada um pode ser facilmente representado por um bloco de confiabilidade. Alguns *softwares* são utilizados para representação de sistemas através da construção de modelos RBD, tais como: Astro Tool [SMT<sup>+</sup>10] e Blocksim7 [Rel12]. Mais detalhes sobre RBDs podem ser encontrados em [KZ03, RH04b]. A sessão a seguir apresenta alguns conceitos importantes sobre dependabilidade e seus atributos.

## 2.3 DEPENDABILIDADE

A análise de dependabilidade é uma atividade essencial que visa fornecer meios para que seja possível promover melhoria da qualidade dos serviços prestados, além de possibilitar o planejamento de melhores infraestruturas de diversos sistemas. Devido à expansão dos serviços oferecidos pela Internet, a dependabilidade tem se tornado um atributo de grande interesse no desenvolvimento de *software* e *hardware*, na implantação e operação dos serviços [MTMJ10].

Geralmente, a avaliação de dependabilidade tem por objetivo melhorar os níveis disponibilidade e confiabilidade dos serviços, além de fornecer informações sobre meios para minimizar custos. Os conceitos de dependabilidade foram inicialmente definidos com a publicação do livro *Dependability Basic Concepts and Terminology* [LAK92] e, desde então, têm sido aplicados e utilizados largamente. Os sistemas que são analisados sob as métricas de dependabilidade são denominados dependáveis [ALR01].

Para a avaliação de dependabilidade, uma abordagem bastante útil é a combinação dos resultados de modelos de diferentes níveis que possuem relação de hierarquia, isso evita o problema de *largeness*, que pode incidir em análises que abrangem muitos componentes num mesmo modelo.

Dependabilidade de um sistema pode ser definida como a habilidade que ele tem de entregar um determinado serviço no qual se possa justificadamente confiar [KZ03, LAK92]. Avizienis *et al.* [ALR01], além da definição anterior, dá uma segunda alternativa para a dependabilidade, que pode ser entendida como a capacidade de evitar falhas de serviço que são mais frequentes e mais severas do que é aceitável. Outro conceito dado por Parhami [Par88], afirma que o fabricante tem que garantir que o sistema irá executar ações especificadas ou apresentar resultados específicos de maneira confiável e em tempo oportuno.

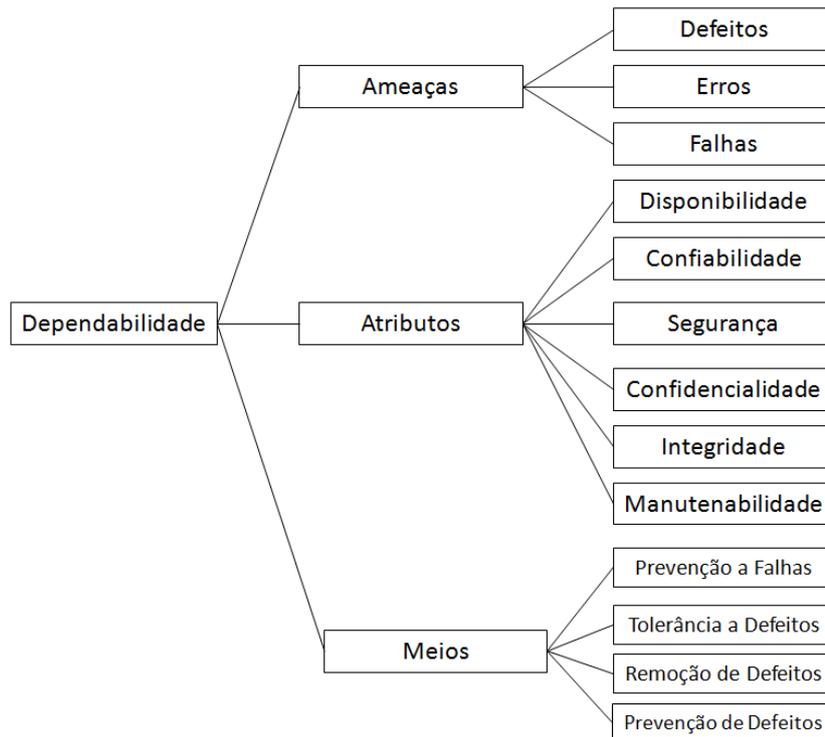
Conforme citado nos trabalhos de Avizienis *et al.* [ALR01] e Laprie *et al.* [LAK92], uma exposição sistemática dos conceitos de dependabilidade pode ser observada na Figura 2.11 que apresenta a árvore da dependabilidade e consiste em três partes:

- As ameaças: compreendem aos defeitos, erros e falhas. A falha do sistema representa o evento que ocorre quando a entrega do serviço não acontece corretamente.
- Os atributos: possibilitam a obtenção de medidas quantitativas, que muitas vezes são cruciais para a análise dos serviços oferecidos.
- Os meios: são os meios pelos quais a dependabilidade é atingida.

Um serviço é confiável quando se implementa a funcionalidade especificada do sistema. Uma falha de sistema ocorre quando o sistema não consegue fornecer sua funcionalidade especificada. Um defeito pode ser definido como a falha de um componente do sistema, um subsistema do sistema, ou outro sistema que interage com o sistema considerado. Assim, cada defeito é uma falha de algum ponto de vista. Um defeito pode causar outros defeitos, a falha do sistema, ou nenhum. Um sistema com defeitos que proporciona a sua funcionalidade especificada é dito ser tolerante a falhas, isto é, o sistema não falha, mesmo quando há componentes defeituosos [PKMK11].

Os atributos de dependabilidade são: disponibilidade, confiabilidade, segurança, confidencialidade, integridade e manutenibilidade [LAK92]. Conforme o interesse do sistema e sua complexidade, ênfase pode ser dada a diferentes atributos, ou seja, alguns podem ser mais bem utilizados que outros. Os atributos de interesse deste trabalho são disponibilidade, confiabilidade e manutenibilidade (descritos logo mais).

A análise de dependabilidade compreende o cálculo de índices como Tempo Médio para Falha (*Mean Time To Failure - MTTF*), Tempo Médio para Reparo (*Mean Time To Repair - MTTR*), Tempo Médio entre Falhas (*Mean Time Between Failures - MTBF*) e Tempo Médio de Parada (*Mean Downtime - MDT*) [CAK04], descritos a seguir:



**Figura 2.11:** Árvore de Dependabilidade  
Adaptado de [ALR01]

- MTTF - é o tempo médio para a ocorrência de falhas no sistema. O MTTF é representado pela Equação (2.15).

$$MTTF = \int_0^{\infty} R(t)dt \quad (2.15)$$

- MTTR - é o tempo médio em que o sistema está indisponível devido a atividades de manutenção. Os MTTRs estão intimamente relacionados com a política de manutenção adotada pela organização, sendo representado pela Equação (2.16).

$$MTTR = MTTF \times \frac{UA}{A} \quad (2.16)$$

Onde  $UA$  representa o tempo de inatividade do sistema (Equação 2.17) e  $A$  representa a disponibilidade do sistema (Equação 2.19).

$$UA = 1 - A \quad (2.17)$$

- MTBF - é o tempo médio entre as falhas do sistema. Os MTBFs são geralmente fornecidos pelo fabricante de *hardware*, sendo representado pela Equação (2.18).

$$MTBF = MTTR + MTTF \quad (2.18)$$

- 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.

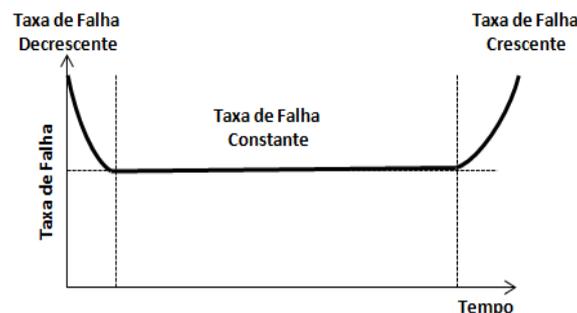
Estas medidas são de grande importância quando se deseja quantificar o efeito das falhas e dos processos de reparo e quando se precisa obter tanto o tempo de atividade como o de inatividade.

A Figura 2.12 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 [Ebe97].

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*, em que 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 a 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 certo período. Ao mesmo tempo, deve-se tomar cuidado para que o equipamento seja substituído antes de entrar na fase final.



**Figura 2.12:** Curva da Banheira

O tempo de atividade (*Uptime*) é o período de tempo em que o sistema está operacional e o tempo de inatividade (*Downtime*) corresponde ao período de tempo em que o sistema não está operacional, que pode ser devido à ocorrência de um evento de falha ou

atividade de reparo. O período de tempo de observação do sistema se refere à soma de ambos *Uptime* e *Downtime*.

Em Trivedi *et al.* [TKRM09b], os autores definem disponibilidade como a capacidade de o sistema executar a sua função prevista num instante de tempo específico ou ao longo de um período de tempo determinado. Ou seja, 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.

A disponibilidade é geralmente expressa sob a forma de uma razão entre as unidades de tempo, quando o serviço estava disponível e o período de serviço contratado. Ela pode ser obtida por análise ou simulação. A Equação 2.19 [MLC96] representa a disponibilidade estacionária ( $A$ ) de um sistema expressando a relação entre MTTF e MTTR:

$$A = \frac{MTTF}{MTTF + MTTR} \quad (2.19)$$

A disponibilidade também pode ser expressa em termos de número de noves conforme a Equação (2.20). 100 representa o nível de disponibilidade máxima que o sistema pode atingir.

$$N = 2 - \log(100 - A) \quad (2.20)$$

A confiabilidade de um item ou sistema é a probabilidade de que um sistema desempenhará satisfatoriamente (ou adequadamente) o seu propósito específico, por um dado período, até a ocorrência da primeira falha [MTMJ10, XPD04]. Ou seja, é a probabilidade de que o sistema não falhará até o tempo especificado [KZ03, RH04a].

Para descrever a confiabilidade de um dado sistema, é necessário conhecer a sua configuração, o estado em que ele é definido como operacional e as regras de operação [KZ03]. Matematicamente, a função de confiabilidade  $R(t)$  é a probabilidade de que o sistema irá funcionar corretamente, sem falha, no intervalo de tempo de 0 a  $t$  [MTMJ10], conforme apresentado na Equação 2.21.

$$R(t) = P(T > t), t \geq 0 \quad (2.21)$$

Onde  $T$  é uma variável aleatória, representando o tempo de falha ou tempo para falha.

A probabilidade de falha, ou inverso da confiabilidade é, então, representada pela Equação 2.22, que é conhecida como a função de distribuição de  $T$ .

$$F(t) = 1 - R(t) = P(T \leq t) \quad (2.22)$$

Quando um sistema deixa de funcionar, normalmente, aplicam-se operações de reparo para corrigir a falha. O sistema, então, recupera o seu estado operacional em virtude de ajustes feitos ou mesmo de substituições de componentes [XPD04]. Partindo desta

perspectiva, manutenibilidade é a capacidade de o sistema sofrer modificações e reparos [TKRM09b]. Ou seja, é a probabilidade de que um sistema seja reparado após a ocorrência de um evento de falha em um determinado período de tempo.

A manutenibilidade é descrita pela Equação 2.23, em que  $T$  denota o tempo de reparo ou o tempo total de inatividade. Essa equação representa a manutenibilidade, visto que o tempo de reparo  $T$  tem uma função de densidade  $g(t)$ .

$$V(t) = P\{T \leq t\} = \int_0^t g(t)dt \quad (2.23)$$

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. Neste sentido, técnicas para prevenção a falhas, tolerância a faltas, remoção de faltas e prevenção a faltas devem ser utilizadas para atingir a dependabilidade.

Higgins *et al.* [HMS02] afirmam que o uso de políticas de manutenção tem um impacto fundamental sobre a dependabilidade de sistemas por evitar que mecanismos destrutivos ocorram, aumentando, assim, o grau de confiabilidade e melhorando o desempenho do sistema. O conceito de política de manutenção é descrito na próxima seção.

## 2.4 POLÍTICA DE MANUTENÇÃO

Indisponibilidade, baixa confiabilidade e baixo desempenho têm sido objetos de atenção das prestadoras de serviços, gestores de infraestrutura, projetistas de aplicações e usuários, os quais exigem soluções da comunidade científica.

O conceito de falha implica uma quebra, ruptura, descontinuidade entre o esperado e o verificado. Assim, é um conceito dinâmico, pois tem subjacente uma ação. Um defeito, por sua vez, é uma característica física ou funcional, essencialmente estática. Contudo, não o é completamente, pois pode também, originar-se apenas no decorrer do funcionamento do próprio sistema.

Para o correto funcionamento de um sistema, é necessário se preocupar com a sua concepção e manutenção ao longo de sua vida operacional. Assim, 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, estratégias de manutenção podem ser utilizadas para 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 [Sou09]. A esse processo dá-se o nome de política de manutenção.

Uma política de manutenção compreende todas as ações que devem alterar um estado

do sistema, a fim de mantê-lo em um modo operacional ou de devolvê-lo a uma condição operacional se ele falhou [BM03]. Ela pode ser dividida em dois grupos básicos: Manutenção Preventiva (*Preventive Maintenance* - PM) e Manutenção Corretiva (*Corrective Maintenance* - CM).

A manutenção preventiva (PM) compreende as atividades periódicas para aumentar a vida útil do sistema e sua confiabilidade. Além disso, a manutenção pode exigir o desligamento de um sistema em modo operacional para a realização de substituição planejada de componentes, testes e revisões. A PM é crucial para alcançar o desempenho máximo dos equipamentos, oferecendo a oportunidade de detectar e corrigir eventuais problemas, antes que eles se tornem significativos e caros, minimizando, assim, o risco de paradas não planejadas [WP06].

Por outro lado, manutenção corretiva (CM) é executada para restaurar um sistema com falha para o estado operacional. As ações envolvem, basicamente, reparos ou substituições (tanto por itens novos ou usados) de todas as peças defeituosas e componentes essenciais para o bom funcionamento de todo o sistema [WP06].

As ações de manutenção realizadas variam, além do tipo de manutenção, com a política de manutenção adotada. Existem vários tipos de políticas de manutenção, 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 [WP06].

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 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 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 (2.24) 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 (2.25) em que  $C_U$  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.

$$R(t) = P\{T \geq t\} = \int_t^{\infty} f(t)dt \quad (2.24)$$

$$C(T) = \frac{C_U[1 - R(T)] + C_P R(T)}{\int_0^T R(t)dt} \quad (2.25)$$

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 (2.26).

$$C(T) = \frac{C_U}{MTBF} + \frac{C_P}{T} \quad (2.26)$$

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.

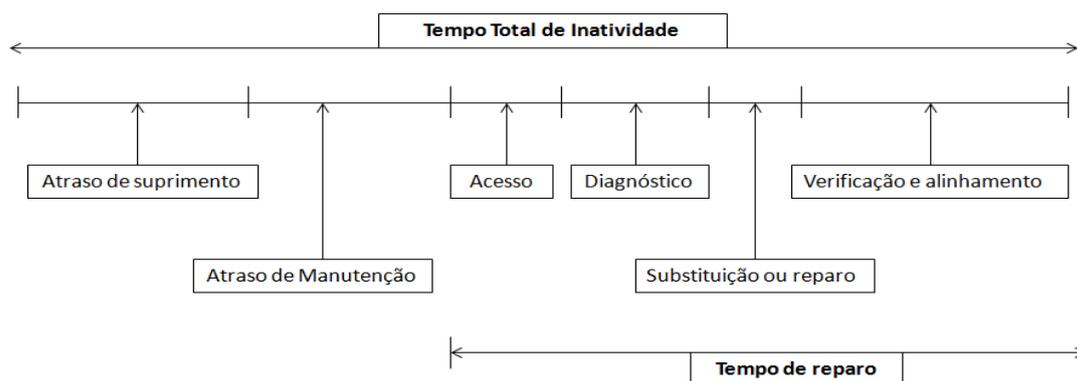
Na política de manutenção preventiva sequencial, a manutenção de um componente é realizada em intervalos de tempo desiguais. Geralmente, esse intervalo de tempo vai se tornando menor com o tempo devido ao fato de que o componente vai precisar de manutenções mais frequentes com o passar do tempo.

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.

Quando um item falha, ele entra no processo de reparo. O processo de reparo em si pode ser decomposto em um número de diferentes subtarefas e tempos de atraso, como apresentado na Figura 2.13.



**Figura 2.13:** Tempo de Inatividade para Manutenção  
Adaptado de [Ebe97]

O atraso de suprimento consiste no tempo de atraso total na obtenção de peças ou componentes necessários, a fim de completar o processo de reparo. Este tempo pode ser constituído de prazos administrativos, de produção ou prazos de entrega de aquisição, reparo de falhas dos subcomponentes e tempos de transporte. Em geral, este tempo é influenciado pela complexidade da seleção de peças e componentes de reposição disponíveis na instalação [Ebe97].

O tempo de atraso de suprimento não ocorre necessariamente no início do ciclo de reparo, ele pode também ocorrer após o diagnóstico da subtarefa ter identificado o componente falho para ser substituído. No entanto, será vantajoso manter este tempo separado das outras categorias. Obviamente, o tempo de atraso de suprimento será zero se a parte necessária para substituição estiver imediatamente disponível [Ebe97].

O tempo de atraso de manutenção é o tempo gasto esperando por recursos ou instalações de manutenção. Pode também incluir tempo administrativo (notificação) e tempo de viagem. Recursos podem ser pessoal, equipamentos de teste, equipamento de suporte, ferramentas, manuais ou outros dados técnicos. Instalações podem ser os ambientes, como uma baía de serviço em uma oficina de reparo de automóvel ou um posto de teste fixo [Ebe97].

O tempo de atraso de manutenção é influenciado pelo número de canais de reparo atribuídos paralelamente. Um canal de reparo é definido como todos os recursos de manutenção e infraestruturas necessárias para iniciar e concluir o processo de reparo (com

exceção dos componentes e peças de reposição). Se um canal de reparo está imediatamente disponível em caso de falha do item, o tempo de atraso de manutenção é zero [Ebe97].

Os tempos de atraso de suprimento e tempos de atraso de manutenção são influenciados por parâmetros externos (tais como os níveis de recursos), que não fazem parte do próprio sistema, eles não são considerados parte do tempo de reparo inerente do item. O tempo de reparo inerente do item é definido como sendo a soma das durações das seguintes subtarefas: acesso, diagnóstico, reparo ou substituição, validação e alinhamento [Ebe97].

O tempo de acesso é a quantidade de tempo necessária para obter acesso ao componente com falha. Tempo de diagnóstico é a quantidade de tempo necessária para determinar a causa da falha. É também referido como tempo de isolamento de falha. O tempo de reparo ou tempo de substituição inclui o tempo para completar o processo de restauração, uma vez que o problema foi identificado e o acesso ao componente com falha foi obtido [Ebe97].

Qualquer atraso na espera de peças adicionais, pessoal, equipamentos de teste e assim por diante, é fonte de atraso no tempo de manutenção e não é considerado parte da tarefa, substituição ou reparo. Após a restauração, algumas falhas podem exigir validação, restabelecimento ou a verificação de alinhamento para garantir que a condição operacional da unidade foi devolvida. Se esta verificação é necessária, é considerada parte do tempo de reparo [Ebe97].

Com a finalidade de especificar a manutenibilidade de um item, a condição operacional aceitável da unidade e as condições sob as quais a manutenção está para ser realizada deve ser definida. Como confiabilidade, manutenibilidade é definida para ser uma probabilidade e é caracterizada pela especificação de uma distribuição de probabilidade de tempo de reparo. Há várias medidas de manutenção. A mais popular é o tempo médio para reparo (MTTR). Outras possíveis medidas incluem a moda, ou o tempo de reparo mais provável, o tempo em que uma determinada percentagem das falhas deve ser reparada, o tempo de reparo médio mais o tempo médio de inatividade para manutenção preventiva, entre outras [Ebe97].

É importante ressaltar que este trabalho se concentra em ações de manutenção programadas (manutenção preventiva) e não programadas (manutenção corretiva) destinadas a aumentar o tempo operacional ou restaurar um sistema de um estado de falha em um estado de funcionamento, levando em consideração que os itens falhos serão recuperados ou substituídos e haverá um tempo administrativo e um tempo de viagem associados à chamada da equipe de manutenção corretiva. O comportamento do sistema, após a ação de reparo, dependerá do tipo de reparação realizada. A próxima seção aborda conceitos sobre SLA.

## 2.5 ACORDO DE NÍVEL DE SERVIÇO

Acordo de Nível de Serviço (SLA) é um documento que define um conjunto de compromissos entre a empresa que fornece serviços e o cliente [Alp06]. Este documento descreve os produtos (serviços a serem contratados) e as taxas a serem alcançadas para o cumprimento de todos os compromissos assumidos [SMJ00]. A empresa que fornece o serviço deve oferecer níveis de serviços de acordo com o desempenho e as expectativas de custo de cada SLA [Alp06].

Adicionalmente, os contratos SLAs devem apresentar informações sobre o desempenho, gerenciamento de problemas, responsabilidade das partes, planos alternativos, relatórios de monitoramento, segurança e penalidades de acordo com os níveis de serviços oferecidos aos clientes [MS04]. Além disso, devem definir uma classe operacional que atenda às expectativas de disponibilidade, nas quais as multas são impostas quando os níveis de disponibilidade não são alcançados.

Indicadores automatizados para coleta e monitoramento dos itens do contrato devem incluir meios de segurança e auditoria que agreguem confiabilidade ao indicador e servem para o cumprimento de um SLA. Clientes e fornecedores devem implantar mecanismos de monitoramento dos indicadores para total conhecimento dos níveis de serviços.

O tempo de inatividade é uma métrica utilizada para a imposição de multas. Contudo, os contratos SLAs devem garantir o desempenho, além de analisar os custos como resultado da indisponibilidade inesperada de um serviço e a probabilidade que ela ocorra [Alp06].

Cada nível de serviço no SLA pode ser diferenciado para definição de parâmetros de QoS, porém os SLAs, por si sós, não surtem qualquer efeito, pois seu valor reside na maneira como são gerenciados. É essencial melhorar a habilidade da empresa prestadora de serviço para fazer cumprir o contrato com o cliente, de forma a minimizar a aplicação de penalidades (multas) por não aderência [NM05]. Os itens a seguir mostram métricas de penalização e desempenho:

- Disponibilidade - O prestador de serviços é responsável por garantir a disponibilidade acordada em contrato.
- Tempo de Inatividade - A duração da interrupção deve ser mantida de acordo com o nível de serviço oferecido por contrato. O período de sistema no modo de falha deve ser menor do que o período de suspensão acordado no contrato SLA.
- Multa - Considerando que o período de inatividade é maior do que o acordado no contrato, o prestador de serviços deve pagar uma multa. A multa baseia-se nas horas de interrupções do sistema e o custo de tais horas é estabelecido de acordo com o nível de serviço oferecido.

Os principais desafios do gerenciamento são: prover um serviço confiável e aderente ao SLA; monitorar completamente o SLA, enquanto o serviço está sendo provido; reagir

rapidamente, caso o desempenho esteja insatisfatório; antecipar tanto quanto possível qualquer queda no desempenho, focando na previsão de desempenho do serviço; assegurar faturamento acurado com base no conteúdo e QoS [DP01].

Para determinar o custo de um serviço, devem-se levar em consideração as violações de SLA ou multas por quebra de contrato. Essas multas, geralmente, são baseadas na qualidade dos serviços oferecidos e quando aplicadas representam valores elevados [LCE10].

## 2.6 CONSIDERAÇÕES FINAIS

Este capítulo apresentou uma introdução sobre redes convergentes, destacando seu uso em instituições de ensino, apresentando suas aplicações e vantagens. Em seguida, foram apresentadas técnicas para modelagem de sistemas, tais como cadeias de Markov, redes de Petri e RBD. As redes de Petri podem ser caracterizadas por suas propriedades comportamentais e propriedades estruturais. A partir das redes de Petri, surgiram várias extensões, que acrescentaram tempo ao seu formalismo. As redes de Petri estocásticas (SPNs), que são de particular interesse deste trabalho, possuem transições com tempos exponencialmente distribuídos e transições imediatas. Posteriormente, foram apresentadas as definições de dependabilidade e atributos de dependabilidade como disponibilidade, confiabilidade e manutenibilidade. Definições sobre política de manutenção, compreendendo manutenção preventiva e manutenção corretiva foram dadas e, por fim, conceitos sobre SLAs foram discutidos.

## CAPÍTULO 3

# METODOLOGIA E MODELOS

Este capítulo apresenta a metodologia para realização deste trabalho, uma breve descrição da rede utilizada para a realização do estudo de caso, os modelos RBDs, que possuem diferentes níveis de abstração, o modelo CTMC e o modelo SPN, para avaliação de políticas de manutenção.

### 3.1 INTRODUÇÃO

O uso de redes convergentes em instituições de ensino reúne uma série de aplicações capazes de auxiliar e impulsionar diversas funções tecnológicas, contribuindo largamente para as demais áreas, principalmente, do ensino à distância. Adicionalmente, integrar comunicações multimídia em uma única rede pode simplificar o gerenciamento de TI e resultar em mais economia, apresentando custos de manutenção menores e menos modificações a fazer. Neste sentido, uma rede convergente pode ajudar a reduzir as despesas de capital, com a instalação de apenas uma única infraestrutura de rede, em vez de duas, e reduzir as despesas operacionais, porque a equipe de manutenção oferece suporte a uma única rede.

Também devem ser consideradas as vantagens de terceirizar a administração da sua rede convergente para um provedor de serviços gerenciados, visando economizar em custos de mão de obra e despesas de capital. Essa opção pode ser bastante econômica para empresas e instituições menores, com recursos limitados de mão de obra de TI e telefonia.

No entanto, a incorporação dos diversos tipos de tráfego de rede em tempo real acrescenta um nível de complexidade que traz exigências de um controle eficaz e de soluções de gerenciamento para lidar com a complexidade introduzida por sua heterogeneidade. Além disso, se a rede ficar indisponível, as operações podem ser interrompidas completamente. A ocorrência de eventos de falhas e atividades de reparo nessas redes pode causar degradação nos atributos de dependabilidade, capazes de levar o sistema a um estado não operante.

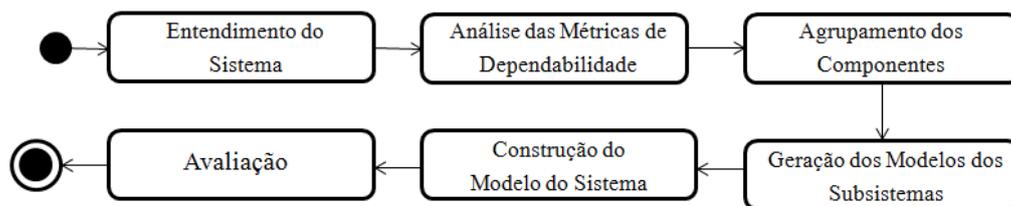
Neste sentido, o desenvolvimento de técnicas, estratégias e modelos que proporcionem meios para avaliação de dependabilidade das redes convergentes são de fundamental importância para empresas que fornecem os serviços, uma vez que os recursos computacionais devem ser usados de forma a atender os níveis de serviços estabelecidos e que se faça um uso eficiente dos recursos financeiros aplicados na infraestrutura computacional.

Como mencionado anteriormente, é preciso ter em mente que não existe um meio universal com o qual possamos avaliar todos os atributos de dependabilidade das diversas classes de sistemas computacionais [Jai91]. Além disso, qualquer erro durante aplicação

ou criação desses procedimentos ou modelos pode ocasionar falhas, pois, cada aplicação tem diversas características, restrições e comportamentos específicos.

### 3.2 METODOLOGIA

A metodologia utilizada para avaliar o impacto da política de manutenção na dependabilidade da infraestrutura de redes convergentes compreende a realização de seis etapas, são elas: entendimento do sistema, análise das métricas de dependabilidade, agrupamento dos componentes, geração dos modelos dos subsistemas, construção do modelo do sistema e avaliação. Estas etapas são mostradas na Figura 3.1 e servem para representar um grupo de atividades gerais que visam descrever um conjunto de resultados obtidos por cada fase particular. Embora os passos apresentados sejam genéricos, neste trabalho, o presente processo é realizado através da adoção de uma estratégia de modelagem híbrida, que apresenta modelos combinatoriais e modelos baseados em estados para representar as características de dependabilidade do sistema.



**Figura 3.1:** Metodologia

A primeira fase se refere à compreensão do sistema, seus componentes, suas interfaces e interações. Nessa fase, é importante a identificação do problema a ser analisado. A segunda etapa visa definir os requisitos de dependabilidade a serem alcançados, que serão considerados no processo de avaliação e nos contratos SLAs, que podem ser determinados sem ter as infraestruturas das redes implementadas. Em seguida é feito o agrupamento de componentes, de forma a gerar os subsistemas, que são aprovados para mitigar a complexidade do sistema e facilitar a avaliação do modelo final.

Os modelos gerados para a realização deste trabalho podem ser expressos em diferentes níveis de abstração. Os modelos RBDs permitem a representação de componentes dependentes em baixo nível de abstração. Deve-se também enfatizar que RBD pode ser avaliado através de simulação, sempre que distribuições mais complexas do que exponencial são adotadas. No modelo CTMC as transições podem ocorrer em qualquer instante de tempo, além de que o comportamento probabilístico é totalmente determinado pelas probabilidades de transição e a distribuição de probabilidades iniciais.

Após a construção dos modelos de subsistemas, a próxima etapa corresponde à construção do modelo do sistema, que considera eventos de falhas e atividades de reparo. Os métodos baseados em estados podem considerar dependências, no entanto, demanda a representação explícita e armazenamento dos estados.

Após as fases que compreendem a geração dos modelos, as avaliações são realizadas. As avaliações dos modelos RBDs visam encontrar parâmetros que servem de entrada para o modelo SPN. A avaliação do modelo CTMC objetiva encontrar a disponibilidade dos computadores de um *site*, considerando taxas de falhas e de reparo. A etapa de avaliação do modelo SPN objetiva analisar o efeito dos eventos de falhas e das atividades de reparo na dependabilidade do sistema. Nessa etapa, consideram-se os requisitos da infraestrutura disponíveis, a estrutura do modelo e as métricas de dependabilidade que se desejam avaliar.

O impacto da política de manutenção é calculado com base nos componentes e recursos do sistema, considerando métricas de dependabilidade, calculadas através dos modelos. Embora as atividades do método proposto sejam apresentadas através de uma estratégia de modelagem hierárquica, na prática, as atividades são realizadas de forma interativa.

### 3.3 DESCRIÇÃO DA REDE

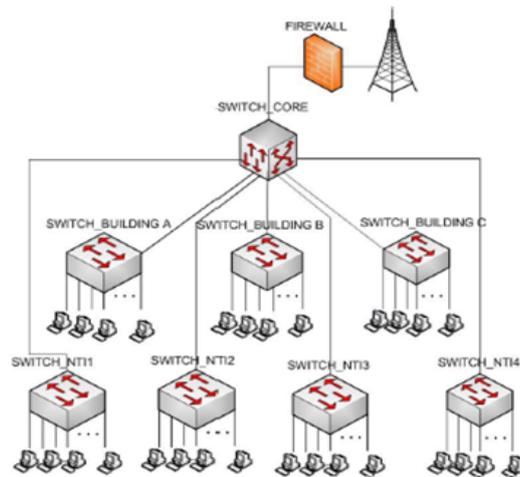
Os modelos concebidos neste trabalho representam uma infraestrutura de rede convergente de uma instituição de ensino superior, localizada no nordeste brasileiro, a qual serviu como referência para o estudo. O campus está situado em uma cidade do agreste e, juntamente com seus três polos, oferecem 16 cursos em diferentes áreas. O campus sede (*Site1*) é sediado em uma cidade que possui influência sobre toda a porção central do estado. São 37 municípios diretamente envolvidos, contando com uma população de mais de 3.120.494 habitantes (IBGE, 2010) [IBG10]. A Tabela 3.1 apresenta a área territorial onde fica situado o campus sede e seus outros três polos, bem como a população de cada um deles [IBG10].

**Tabela 3.1:** Informações Estatísticas dos Polos

Cidade / Campus	Área Territorial ( $Km^2$ )	População
Cidade1/ <i>Site1</i>	356, 179	214.006
Cidade2/ <i>Site2</i>	452, 703	70.368
Cidade3/ <i>Site3</i>	689, 156	60.378
Cidade4/ <i>Site4</i>	343, 356	25.407

A presença dos campi no interior possibilita acesso ao ensino superior público e gratuito para uma enorme parcela de estudantes, com baixa ou mesmo nula capacidade de deslocamento ou transferência para capital do estado. Devido ao grande alcance populacional e sua importância estratégica para alavancar o crescimento da região através de atividades de ensino, pesquisa e extensão, é necessária a continuidade de atividades que estão diretamente relacionadas ao correto funcionamento das redes convergentes. As atividades incluem a prática de multimídia (VoIP, videoconferência, vídeo-aula), serviços de *chats* e fóruns em ambientes de cursos de educação à distância etc., que objetivam estimular o aluno para a carreira acadêmica e profissional.

A estrutura de rede do *Site1* é como apresentada na Figura 3.2. Nessa estrutura, os computadores são interligados através de uma topologia que consiste de um nó central (CORE), ao qual os nós folhas estão interligados. Dessa forma, estes componentes formam um grafo com a topologia estrela estendida. Uma topologia em rede estendida tem uma topologia estrela central, em que cada um dos nós finais da topologia central atua como centro de sua própria topologia estrela [KR10].



**Figura 3.2:** Topologia do *Site1*

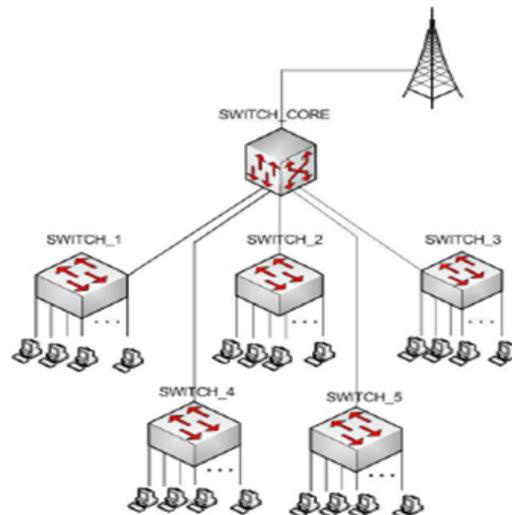
A estrutura da rede dos outros três polos (*Site2*, *Site3* e *Site4*) é como mostrada na Figura 3.3. De maneira semelhante à estrutura da sede (*Site1*), os computadores são interligados através de uma topologia que consiste de um nó central (*CORE*), para o qual os nós folhas estão interligados e estes formam a sua própria topologia estrela. Os componentes que formam cada *site* são: computadores, roteadores, *switches* e *links*.

### 3.4 MODELOS

Esta seção apresenta os modelos RBDs, o modelo CTMC e o modelo SPN concebidos. Os componentes considerados, que constituem a infraestrutura convergente representada, incluem roteadores, *switches*, *links* e computadores. Inicialmente, foram gerados modelos RBDs dos componentes do sistema, além do modelo CTMC e em seguida foi gerado o modelo SPN, para avaliação da política de manutenção.

O MTTF obtido com a avaliação do primeiro modelo RBD é utilizado posteriormente, no modelo RBD de segundo nível. O MTTF obtido com a avaliação do modelo RBD de segundo nível é utilizado como parâmetro de entrada para o terceiro modelo. Por fim, o resultado do modelo RBD de terceiro nível é utilizado para representar o MTTF do sistema no modelo SPN, representado por uma distribuição Erlang.

O primeiro modelo construído representa os componentes de um computador. O segundo modelo RBD serve para representar a quantidade total de computadores para



**Figura 3.3:** Topologia dos *Sites 2, 3 e 4*

cada *site*. O modelo CTMC apresenta as taxas de falhas e reparo dos computadores de um *site*, que permite verificar a disponibilidade do sistema. O modelo RBD de terceiro nível representa todos os componentes de um *site* e por fim, o modelo SPN serve para avaliar a política de manutenção, considerando diferentes abordagens. As seguintes subseções apresentam detalhadamente os modelos concebidos.

### 3.4.1 Modelo RBD de Primeiro Nível

O primeiro modelo concebido representa os componentes que são integrados para formar um computador, são eles: processador (CPU), disco, memória, fonte de energia, placa de rede, *cooler* e sistema operacional (SO). O MTTF dos componentes processador e disco foram utilizados com base em um trabalho de Sousa *et al.* [SMLM12] e podem ser observados na Tabela 3.2.

**Tabela 3.2:** MTTF Processador e Disco

Componente	MTTF (h)
Processador	17.520
Disco	43.800

Os MTTFs dos componentes *cooler* e fonte de energia foram obtidos a partir do manual dos equipamentos [Int12, The12]. Os MTTFs da memória e placa de rede, obtidos de um trabalho de Kim *et al.* [KMT09], estão relacionado aos valores fornecidos pelos fabricantes de *hardware*. O MTTF do sistema operacional (SO), também obtido do trabalho de Kim *et al.* [KMT09], é fornecido com base em estimativas para falhas de

*software*.

Os valores dos MTTFs dos componentes de *hardware* (*cooler*, fonte de energia, memória e placa de rede) sofreram uma redução de 5 vezes do valor fornecido [Smi05], devido o valor ser referente ao obtido pelo manual do fabricante, essa redução deve ser aplicada. O resultado da redução pode ser observado na Tabela 3.3. Os valores dos MTTFs dos componentes citados nas Tabelas 3.2 e 3.3 são utilizados como parâmetros de entrada no modelo RBD de primeiro nível.

**Tabela 3.3:** MTTF dos Componentes de um PC

Componente	MTTF (h)
<i>Cooler</i>	10.000
Fonte de Energia	20.000
Memória	96.000
Placa de Rede	24.000
SO	1.440

Como pode ser observado na Tabela 3.3, o MTTF do SO é igual a 1.440 horas, valor esse, muito baixo em relação ao MTTF dos outros componentes.

O modelo operacional para esse sistema deve ter todos os componentes em funcionamento, o mesmo pode ser dividido em duas partes: a parte de *hardware* e a parte de *software*. Observe a Figura 3.4. Caso ocorra uma falha em algum dos componentes, o sistema falha.

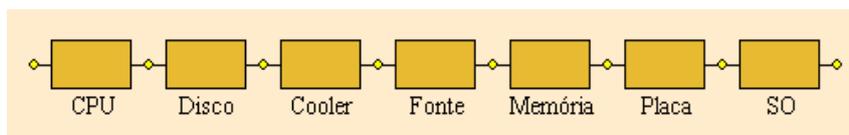


**Figura 3.4:** Modelo Operacional - Componentes de um Computador

A Figura 3.5 mostra o modelo RBD gerado para representar todos os  $n$  componentes do sistema, o qual é representado por uma estrutura em série com blocos que apresentam distribuição de falhas do tipo exponencial. A representação através de blocos em série indica que se um componente falhar, todo o sistema irá falhar.

Para avaliação do modelo, os valores são fornecidos como taxa de falhas ( $1/\text{MTTF}$ ) e o MTTF resultante representa o MTTF de um computador. O cálculo do MTTF é realizado através da Equação 2.15. É importante ressaltar que os modelos são parametrizáveis.

O resultado da avaliação do modelo RBD de primeiro nível apresenta um MTTF de 1.024,13 horas. Como esperado, a ocorrência de falhas do sistema está diretamente

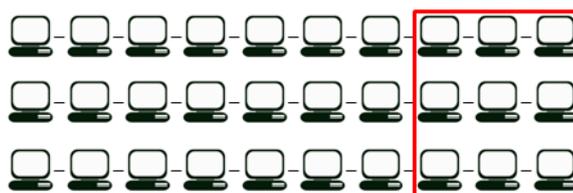


**Figura 3.5:** Modelo RBD de Primeiro Nível

realcionada a taxa de falhas do componente SO, o qual apresenta a mais baixa taxa de falhas. Esse resultado será utilizado no modelo RBD de segundo nível.

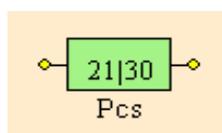
### 3.4.2 Modelo RBD de Segundo Nível

O modelo RBD de segundo nível foi concebido a fim de obter o MTTF de 30 computadores, considerando que pelo menos 70% deles deve funcionar para que o sistema não falhe. Neste caso, o modelo operacional deve possuir no mínimo 21 computadores em funcionamento. Observe a Figura 3.6, se 9 computadores, representados na Figura em um retângulo vermelho, falharem simultaneamente, o reparo deverá ser ativado e o sistema entrará em manutenção.



**Figura 3.6:** Modelo Operacional - 30 Computadores (21/9)

Esse sistema é representado por uma estrutura do tipo  $k$ -out-of- $n$  ( $k$  de  $n$ ). Neste sentido, foi gerado um modelo com um bloco  $k$  de  $n$ , com o valor de  $k$  igual a 21 e o valor de  $n$  igual a 30, para especificar a quantidade mínima de computadores que devem funcionar. A Figura 3.7 apresenta o modelo RBD gerado.

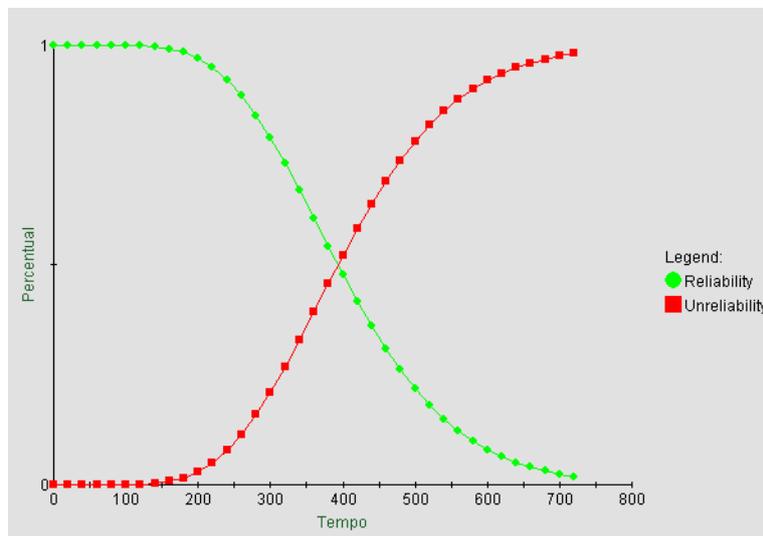


**Figura 3.7:** Modelo RBD de Segundo Nível

Neste modelo, o valor fornecido como parâmetro de entrada, também corresponde a taxa de falhas, que foi obtida com a análise do modelo RBD de primeiro nível (1/1.024,13). O cálculo do MTTF é realizado através da Equação 2.15.

O resultado da avaliação do segundo modelo apresenta um tempo médio para falhas de 406,83 horas. Esse valor será utilizado como parâmetro de entrada para o modelo RBD de terceiro nível, o qual visa representar todos os componentes de um *site*.

Além do modelo RBD concebido, foi verificada a confiabilidade para esse sistema, cujo resultado pode ser observado na Figura 3.8, que mostra a confiabilidade e a inconfiabilidade dos computadores. Os respectivos cálculos são obtidos através das Equações 2.21 e 2.22.



**Figura 3.8:** Confiabilidade X Inconfiabilidade dos Computadores

O gráfico gerado indica que o sistema apresenta um índice de confiabilidade acima de 70% antes de atingir 300 horas de funcionamento.

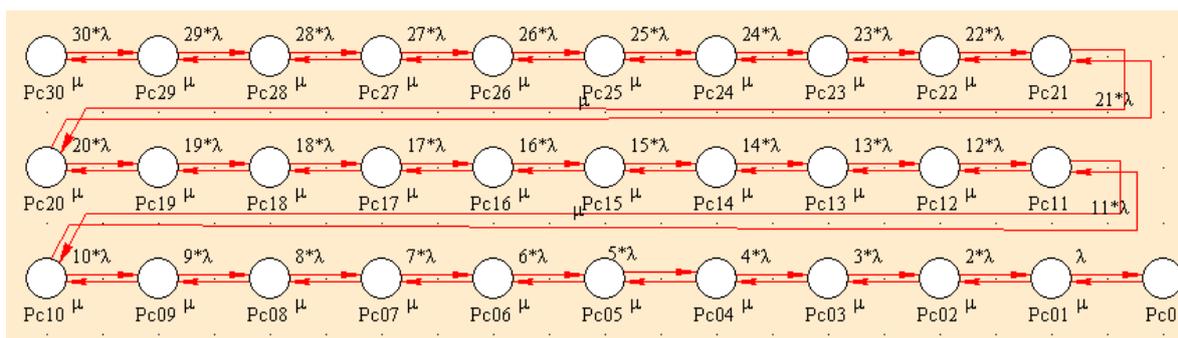
### 3.4.2.1 Modelo CTMC

A fim de verificar a disponibilidade dos computadores, foi gerado um modelo CTMC equivalente ao modelo RBD de segundo nível. O cálculo da disponibilidade é realizado através da Equação 2.19. A Figura 3.9 apresenta o modelo CTMC concebido, no qual é representada a taxa de falha ( $\lambda = 1/1.024,13h^{-1}$ ) e taxa de reparo ( $\mu = 1/48h^{-1}$ ) dos computadores. É considerado que o sistema está em estado *UP* se no mínimo 21 computadores estiverem em funcionamento, caso contrário, é considerado que o sistema está indisponível.

O resultado da análise do modelo CTMC apresenta uma disponibilidade de 54%.

### 3.4.3 Modelo RBD de Terceiro Nível

O modelo RBD de terceiro nível visa representar todos os componentes de um *site*, a fim de obter o MTTF do sistema, considerando uma rede local que possui um roteador, um *switch*, um *link* para transmissão de dados e 30 computadores. O modelo operacional para representação desse sistema indica que todos os componentes devem funcionar, observe a Figura 3.10. Entretanto, vale ressaltar que o MTTF utilizado para representar a taxa de falha dos computadores no modelo RBD de terceiro nível foi obtido com a análise



**Figura 3.9:** Modelo CTMC

do modelo RBD de segundo nível, o qual considera que pelo menos 21 computadores estão em funcionamento, ou seja, corresponde a falha de 70% da quantidade total de computadores.



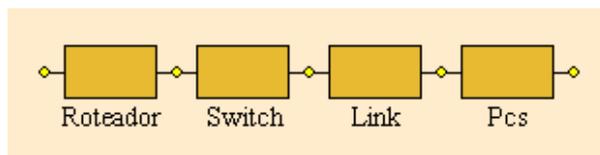
**Figura 3.10:** Modelo Operacional - Componentes de um *Site*

O MTTF do roteador e do *switch* foram obtidos do manual do equipamento [Cis12a, Cis12b]. Os valores dos MTTFs desses componentes sofreram uma redução de 5 vezes do valor fornecido pelo fabricante [Smi05], devido o valor ser referente ao obtido pelo manual do fabricante, essa redução deve ser aplicada. O MTTF do *link* foi obtido de um trabalho de Matos *et. al.* [MGC<sup>+</sup>11]. O MTTF dos computadores foi obtido com o resultado da avaliação do modelo RBD de segundo nível. Os MTTFs utilizados como parâmetros de entrada neste terceiro modelo RBD, são apresentados na Tabela 3.4.

**Tabela 3.4:** MTTF dos Componentes de um *Site*

Componente	MTTF
Roteador	89.600
<i>Switch</i>	58.985
<i>Link</i>	11.988
Computadores	406,83

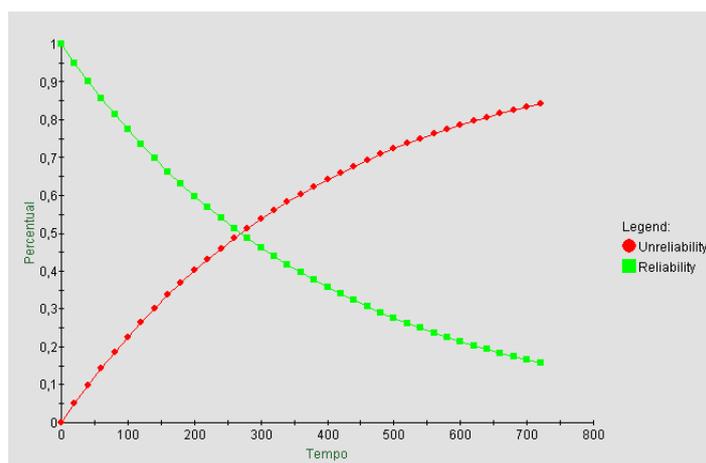
O modelo RBD gerado consiste de uma estrutura em série que representa todos os componentes de um *site*. A estrutura é assim representada, pois, na ocorrência de falha em um dos componentes, todo o sistema falha. O modelo RBD de terceiro nível é apresentado na Figura 3.11.



**Figura 3.11:** Modelo RBD de Terceiro Nível

O resultado da avaliação desse modelo apresenta um tempo médio para falha de 389,17 horas. Esse valor está diretamente relacionado ao MTTF dos computadores, que apresentam a menor taxa de falhas.

Além do modelo concebido, foi verificada a confiabilidade desse sistema, a Figura 3.12 mostra a confiabilidade e a inconfiabilidade desse sistema. Os respectivos cálculos são obtidos através das Equações 2.21 e 2.22. Como pode ser observado, o gráfico indica que o sistema apresenta um índice de confiabilidade acima de 70% antes de atingir 200 horas de funcionamento.



**Figura 3.12:** Confiabilidade X Inconfiabilidade do Sistema

#### 3.4.4 Modelo SPN

O último modelo proposto corresponde ao modelo SPN, o qual representa a política de manutenção e descreve os procedimentos a serem adotados para minimizar as ocorrências de eventos de falhas na infraestrutura. O modelo é composto por quatro sub-redes que representam os *sites* de 4 campi de uma instituição de ensino (Ver Figura 3.20) e mais uma sub-rede que indica o estado do sistema (falhado ou operante (Ver Figura 3.16 e 3.17)). Para modelar o sistema e representar a política de manutenção, adotou-se as SPNs.

A política de manutenção descreve os procedimentos a serem realizados para minimizar as ocorrências de eventos de falhas, permitindo, ainda, ações de reparo. Duas

abordagens para manutenção são utilizadas: manutenção preventiva e manutenção corretiva. A manutenção preventiva é uma manutenção planejada, na qual são realizadas ações de controle e monitoramento dos equipamentos, o que previne a ocorrência corretiva. Esse tipo de manutenção objetiva a verificação periódica do funcionamento dos equipamentos, antecipando eventuais problemas, que possam causar gastos maiores com a manutenção corretiva. A manutenção preventiva pode ser realizada quando o sistema atingir níveis críticos especificados e os intervalos entre as manutenções são fixados. O número de manutenções preventivas que serão realizadas durante o período de um ano pode variar de acordo com a classe de SLA. Já 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 política de manutenção estudada considera dois tipos de equipes de manutenção: as equipes locais (*Local Teams - LTs*) e uma equipe especialista (*Specialist Team - ST*). Como são quatro *sites*, são também quatro equipes locais, cada uma em seu determinado *site*. As LTs realizarão as manutenções preventivas, pois todas elas têm competências para realizar várias atividades de manutenção em diferentes equipamentos, porém, não são especializadas em nenhum equipamento. Além disso, são essas equipes que realizam o registro do histórico de eventos de falha, reparos e substituições dos equipamentos. A equipe local é formada por apenas uma pessoa.

A equipe especialista (ST) é chamada quando os procedimentos são insuficientes para os diagnósticos e reparos. Quando a ST visita um *site*, ela realiza análises periódicas mais complexas ou críticas. Além disso, pode realizar ou não a substituição do equipamento. É considerado que, quando essa equipe é chamada para recuperar o sistema de um estado falho, haverá um tempo de deslocamento associado à sua chegada ao *site* destino, pois foi considerado que a ST se encontra em um local diferente dos *sites* representados e haverá, então, uma determinada distância a ser percorrida até chegar ao *site*.

A ST é formada por dois especialistas. Quando essa equipe é chamada, por motivo de falha em um dos *sites*, somente uma pessoa da equipe viaja. Isso significa dizer que, os quatro *sites* compartilham a mesma equipe especialista. Assim, se acontecerem falhas simultâneas em mais de dois *sites* e a equipe especialista for ativada, haverá uma espera de recursos para os demais *sites* que falharam, ou seja, o terceiro ou quarto *site* que tiver ativado a ST terá que esperar até que o especialista termine o seu trabalho e fique disponível novamente.

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. A equipe local (LT) realiza os ajustes preventivos e pode realizar a manutenção corretiva, desde que esteja apta para isso. Além disso, quando a ST é ativada, a LT inicia as atividades de manutenção enquanto a ST não chega. O tempo médio de reparo da LT, quando esta realiza a manutenção preventiva ( $TRepPMLT\_Site$ ), é composto pelo tempo médio da parada do sistema (programada) e pelo tempo médio de reparo. O tempo médio de reparo da LT, quando esta realiza a manutenção corretiva ( $TRepLT\_Site$ ), é composto pelo tempo administrativo (que inclui a percepção da falha

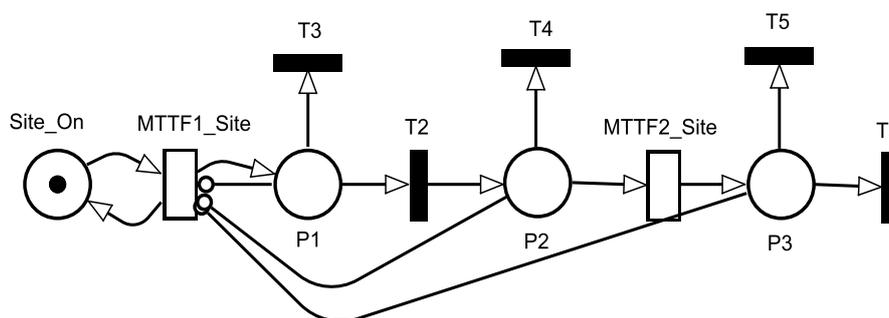
e ativação da equipe) e pelo tempo médio de reparo.

O tempo médio de reparo da ST ( $TRepSLT\_Site$ ) é composto pelo tempo administrativo, pelo tempo médio de chegada da equipe ao local (deslocamento) e pelo tempo médio de reparo. Quando a manutenção é realizada pela ST, o  $TRepSLT\_Site$  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 SPN foi concebido conforme os procedimentos da política de manutenção estudada. 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  $ST$  representa o grau de concorrência relacionado ao processo de reparo pela equipe especialista.

O MTTF dos componentes do sistema é igual a 389,17 horas, esse valor foi obtido com a avaliação do modelo RBD de terceiro nível. No modelo SPN, o valor é distribuído para as variáveis  $MTTF1\_Site$  e  $MTTF2\_Site$ , em que cada uma delas recebe o valor 194,58 associado ao *delay*. A Figura 3.13 apresenta como fica representada no modelo a parte que contém a distribuição Erlang e as transições imediatas ( $T3, T4$  e  $T5$ ), as quais habilitam a manutenção preventiva.

A marca atribuída ao lugar  $Site\_On$  indica o *site* em estado operacional. Os estados entre as transições exponenciais  $MTTF1\_Site$  e  $MTTF2\_Site$  servem para representar momentos críticos, que pode ser pela ocorrência de erros ou defeitos, mas que não ocasionou a parada do sistema. A política de manutenção habilita as manutenções preventivas por meio da função de habilitação  $\{\#RepPMLT\_Site > 0\}$  nas transições imediatas  $T3, T4, T5$  (Ver Figura 3.13). A Tabela 3.5 apresenta as funções de habilitação para a política de manutenção preventiva em cada *site*. Caso a falha não seja recuperada, após sua ocorrência, uma marca será atribuída ao lugar  $Site\_Off$ , indicando indisponibilidade da rede e necessidade de reparo corretivo (Ver Figura 3.15).

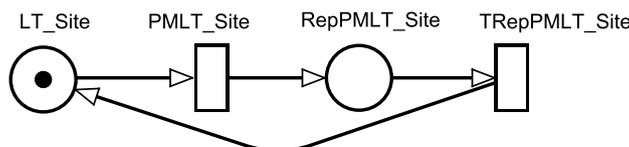


**Figura 3.13:** SPN Tempos para Falha de um *Site*.

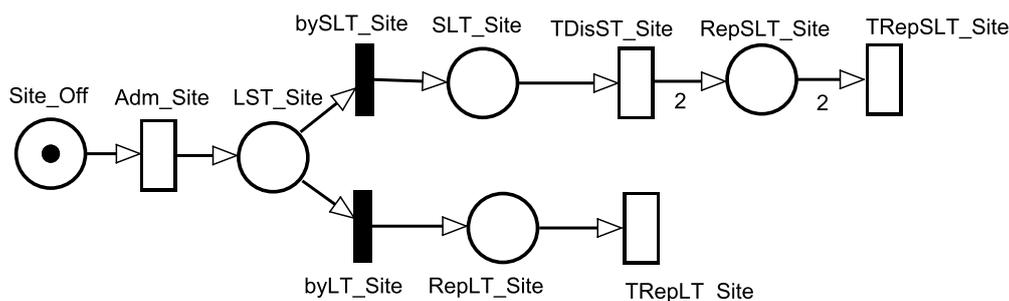
A Figura 3.14 apresenta a sub-rede para o processo de manutenção preventiva, que é realizada pela equipe local. A marcação inicial atribuída ao lugar  $LT\_Site$  representa a disponibilidade da equipe de manutenção preventiva.

**Tabela 3.5:** Função de Habilitação para a Manutenção Preventiva

<i>Site</i>	Transições	Função de Habilitação
<i>Site1</i>	$T3, T4, T5$	$\#RepPMLT\_S1 > 0$
<i>Site2</i>	$T23, T25, T26$	$\#RepPMLT\_S2 > 0$
<i>Site3</i>	$T38, T40, T41$	$\#RepPMLT\_S3 > 0$
<i>Site4</i>	$T53, T55, T56$	$\#RepPMLT\_S4 > 0$

**Figura 3.14:** SPN Manutenção Preventiva pela Equipe Local

A ocorrência de erros e defeitos no sistema é considerada como um momento crítico, que está representado pelos estados que se encontram entre as transições  $MTTF1\_Site$  e  $MTTF2\_Site$ . Caso esses defeitos sejam percebidos, a equipe local é ativada para realizar a manutenção preventiva, então, ajustes podem ser realizados para tentar recuperar os erros e defeitos. Se a LT conseguir recuperá-los, a marcação voltará ao seu estado inicial (uma marca no lugar  $Site\_On$ ). Porém, se os ajustes realizados forem insuficientes para recuperá-los, o site passará para um estado de falha. Neste caso, uma marca será depositada no lugar  $Site\_Off$ , indicando a necessidade de realização de manutenções corretivas. A Figura 3.15 apresenta a parte do modelo que contém o lugar  $Site\_Off$  e os próximos estados que representam como pode ser realizada a manutenção corretiva.

**Figura 3.15:** SPN Manutenção Corretiva

Quando uma marca é depositada no lugar  $Site\_Off$  (parada do  $site$ ), haverá um tempo administrativo que está representado pela transição exponencial  $Adm\_Site$ . Esse tempo administrativo se refere a percepção da parada do sistema, verificação da causa e o tempo para ativar a equipe que será responsável para realizar a manutenção corretiva. É importante ressaltar que a equipe local pode realizar a manutenção corretiva, mas se for percebido que a causa da falha só pode ser recuperada pela equipe especialista (como substituição de componentes), a ST será notificada e convidada a ir ao  $site$ .

O lugar  $LST\_Site$  (Figura 3.15) representa a possibilidade de a manutenção ser realizada por ambas as equipes. O disparo da transição imediata  $byLT\_Site$  ativa o reparo pela equipe local (LT). O estado  $RepLT\_Site$  indica que a LT está no *site* realizando o reparo e o disparo da transição  $TRepLT\_Site$  deposita uma marca no lugar  $Site\_On$ , indicando que, após o tempo gasto pela equipe com as atividades de reparo, o *site* voltará ao seu estado de funcionamento normal e a equipe é liberada.

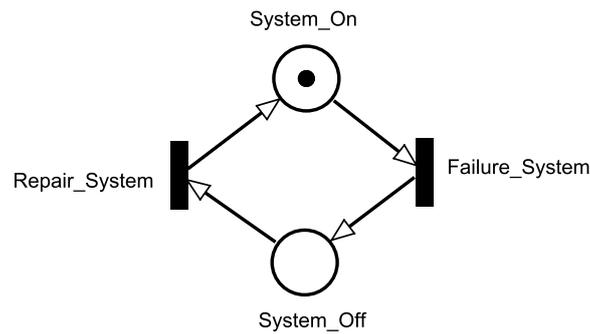
O disparo da transição imediata  $bySLT\_Site$  (Figura 3.15) ativa o reparo pela equipe especialista (ST). Neste caso, a equipe local (LT) ajuda a equipe especialista nas ações de reparo. Assim que a transição  $bySLT\_Site$  for disparada, a LT inicia o processo de reparo ( $SLT\_Site$ ). A transição exponencial  $TDisST\_Site$  representa o tempo gasto pela ST para chegar ao *site*, após seu disparo, estarão as duas equipes realizando ações de reparo ( $RepSLT\_Site$ ). Os pesos de valor 2 nos arcos indicam o depósito e a retirada de 2 marcas do lugar. A transição  $TRepSLT\_Site$  indica que, após o tempo gasto pelas equipes com as atividades de reparo, o *site* voltará ao seu estado de funcionamento normal ( $Site\_On$ ) e as equipes serão liberadas. A Tabela 3.6 apresenta o tempo gasto pela equipe especialista para chegar aos sites, os quais foram calculados de acordo com a distância que deve ser percorrida para se chegar ao local que concentra cada instituição, considerando que a equipe especialista se encontra em uma cidade diferente das que possuem os campi.

**Tabela 3.6:** Tempo de Deslocamento da Equipe Especialista

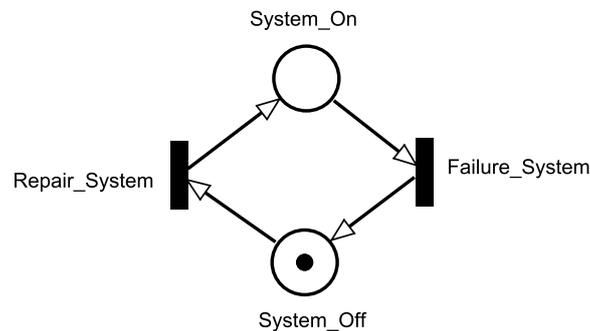
<i>Site</i>	Transição	Deslocamento (H)
<i>Site1</i>	$TDisST\_S1$	1,85
<i>Site2</i>	$TDisST\_S2$	2,15
<i>Site3</i>	$TDisST\_S3$	1,95
<i>Site4</i>	$TDisST\_S4$	1,28

Incluindo os tempos de identificação do defeito, disponibilização de placas, equipamentos, deslocamento da equipe de manutenção, substituição e conserto da parte defeituosa e retorno do serviço à normalidade, deve ser assegurado o menor tempo médio de reparo possível (MTTR) pelas equipes. Além disso, é particularmente importante o dimensionamento do estoque de unidades sobressalentes, pois sua falta poderá resultar em reparos extremamente prolongados. Neste caso, o MTTR poderá ser bem maior, além disso, em casos específicos, como a dificuldade da equipe chegar ao local (intempérie) também podem aumentar o MTTR.

A sub-rede, que indica se o sistema está ou não em funcionamento considera que para o sistema estar funcionando, todos os sites precisam estar funcionando, se qualquer um dos *sites* sofrer alguma parada, esta sub-rede indicará necessidade de reparo. A marcação inicial, indica que o sistema está funcionando ( $System\_On$ ) (Ver Figura 3.16). Na ocorrência de falha em algum dos *sites*, uma marca será depositada no lugar  $System\_Off$  (Ver Figura 3.17).



**Figura 3.16:** SPN Sistema em Estado Operacional



**Figura 3.17:** SPN Sistema em Estado Não Operacional

A transição imediata *Failure\_System* verifica a disponibilidade dos *sites* por meio de uma função de habilitação e indica a necessidade de reparo, caso algum dos *sites* esteja indisponível. Ou seja, a política de manutenção habilita as manutenções corretivas através da função de habilitação contida na transição *Failure\_System*. Após o disparo dessa transição, uma marca será depositada no lugar *System\_Off*, indicando indisponibilidade do sistema. De maneira semelhante, a política de manutenção habilita o reparo através da função de habilitação na transição imediata *Repair\_System*. A Tabela 3.7 apresenta as funções de habilitação para as transições *Failure\_System* e *Repair\_System*.

**Tabela 3.7:** Função de Habilitação Sistema

Transição	Função de Habilitação
<i>Failure_System</i>	$((\#Site1\_On = 0) \text{ OR } (\#Site2\_On = 0) \text{ OR } (\#Site3\_On = 0) \text{ OR } (\#Site4\_On = 0))$
<i>Repair_System</i>	$\text{NOT } ((\#Site1\_On = 0) \text{ OR } (\#Site2\_On = 0) \text{ OR } (\#Site3\_On = 0) \text{ OR } (\#Site4\_On = 0))$

A função de habilitação *Repair\_System* verifica se os *sites* voltaram a ficar disponível após a ação de reparo, se sim, uma marca será depositada no lugar *System\_On*, indicando que o sistema está disponível novamente.

Como mencionado anteriormente, a equipe especialista é composta por duas pessoas que têm habilidades para consertar os equipamentos do *site*. A Tabela 3.8 apresenta o número de pessoas em cada equipe e suas respectivas definições no modelo. A Figura 3.18 apresenta o lugar que representa o número de pessoas na ST. A marcação inicial deste lugar é 2, mas, à medida que esta equipe for sendo chamada, a marcação pode variar entre 0, 1 e 2 marcas.

**Tabela 3.8:** Número de Pessoas por Equipes

Tipo da Equipe	Número Pessoas
Equipe Local (LT)	$NLT = 1$
Esquipe Especialista (ST)	$NST = 2$



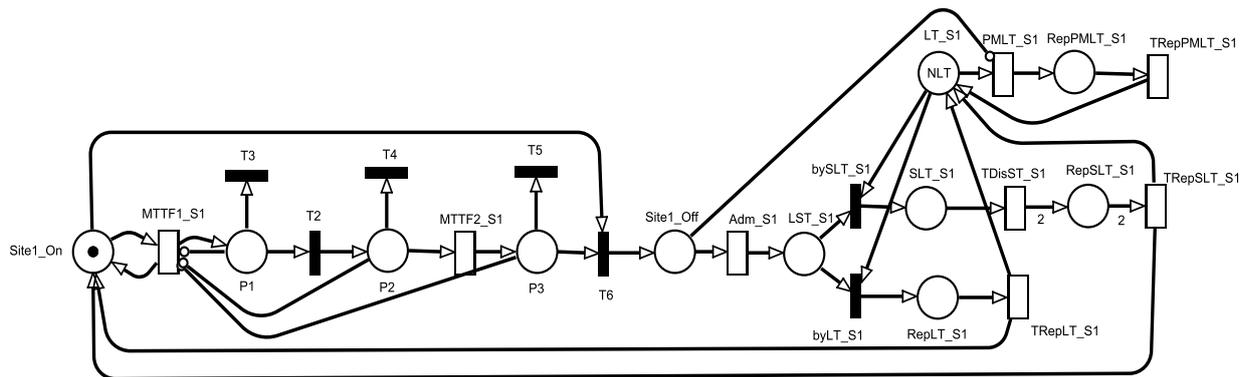
**Figura 3.18:** SPN Equipe Especialista

As SPNs que representam os *sites* dos 4 campi da instituição possuem a mesma estrutura e são denominadas como *Site1*, *Site2*, *Site3* e *Site4*. Os nomes atribuídos aos lugares e transições de cada rede são diferenciado pelo *site* ao qual pertencem ( $S1$ ,  $S2$ ,  $S3$  e  $S4$ ). Também são atribuídos valores associados as transições que podem ser diferentes para cada *site*. A Figura 3.19 apresenta a estrutura do modelo para o *Site1*. O arco direcionado, que liga o lugar *Site1\_On* à transição imediata  $T6$ , indica que, após a ocorrência de falhas, a marca será retirada do lugar *Site1\_On*. Com o disparo da transição  $T6$  a marca será depositada no lugar *Site1\_Off*. O arco inibidor que liga o estado *Site1\_Off* à transição estocástica temporizada  $PMLT\_S1$  serve para inibir o disparo desta transição, se tiver uma marca no lugar *Site1\_Off*. Assim, após a ocorrência de falha, somente a manutenção corretiva pode ser ativada. O arco direcionado que sai da transição  $TRepLT\_S1$  para os estados *Site1\_On* e  $LT\_S1$  deposita uma marca em cada um desses estados, indicando que o *Site1* foi recuperado e está disponível novamente, além de liberar a equipe que fez o reparo. O mesmo pode ser explicado para os outros *sites*.

A Figura 3.20 mostra como fica o modelo SPN composto pelos quatro sites. O arco direcionado que sai da transição  $TRepSLT\_S1$  para os estados *Site1\_On*,  $LT\_S1$  e  $ST$  deposita uma marca em cada um desses estados, indicando que o *Site1* foi recuperado e está disponível novamente, além de liberar as equipes que realizaram o reparo. Do mesmo modo é feito para os outros *sites*.

### 3.5 CONSIDERAÇÕES FINAIS

Neste capítulo, foi apresentada a metodologia para realização deste trabalho, que compreende a realização de seis etapas, que vão desde o entendimento do sistema até a etapa



**Figura 3.19:** SPN Site1

de avaliação dos resultados. Em seguida, foi definida a rede utilizada como base para realização do estudo de caso deste trabalho, a qual é composta por quatro *sites* que são separados geograficamente e interligados por uma rede convergente WAN. Posteriormente, foram apresentados os modelos RBDs, com três níveis de abstração distintos e o modelo CTMC. O modelo RBD de primeiro nível serve para obter o MTTF de um computador, considerando os componentes essenciais para seu funcionamento. O modelo RBD de segundo nível foi concebido para obter a taxa de falhas dos computadores de um *site*, considerando que pelo menos 70% do total tem que estar em funcionamento. Em adição foi concebido um modelo CTMC para calcular a disponibilidade desses computadores. Já o terceiro e último modelo RBD foi concebido para obter a taxa de falha de toda a rede de um *site*, considerando como componentes os computadores, *links*, *switch* e roteador. Também foi mostrado e conceituado o modelo SPN concebido, o qual permite avaliar uma política de manutenção, que adota abordagens preventivas e corretivas. Quando a manutenção preventiva é ativada, somente a equipe local será responsável por sua realização. Entretanto, a manutenção corretiva pode ser realizada por ambas equipes locais e especialistas. O valor obtido com a análise do modelo RBD de terceiro nível serve para representar as falhas (MTTFs) do sistema no modelo SPN. Isso representa a relação hierárquica entre os modelos concebidos.

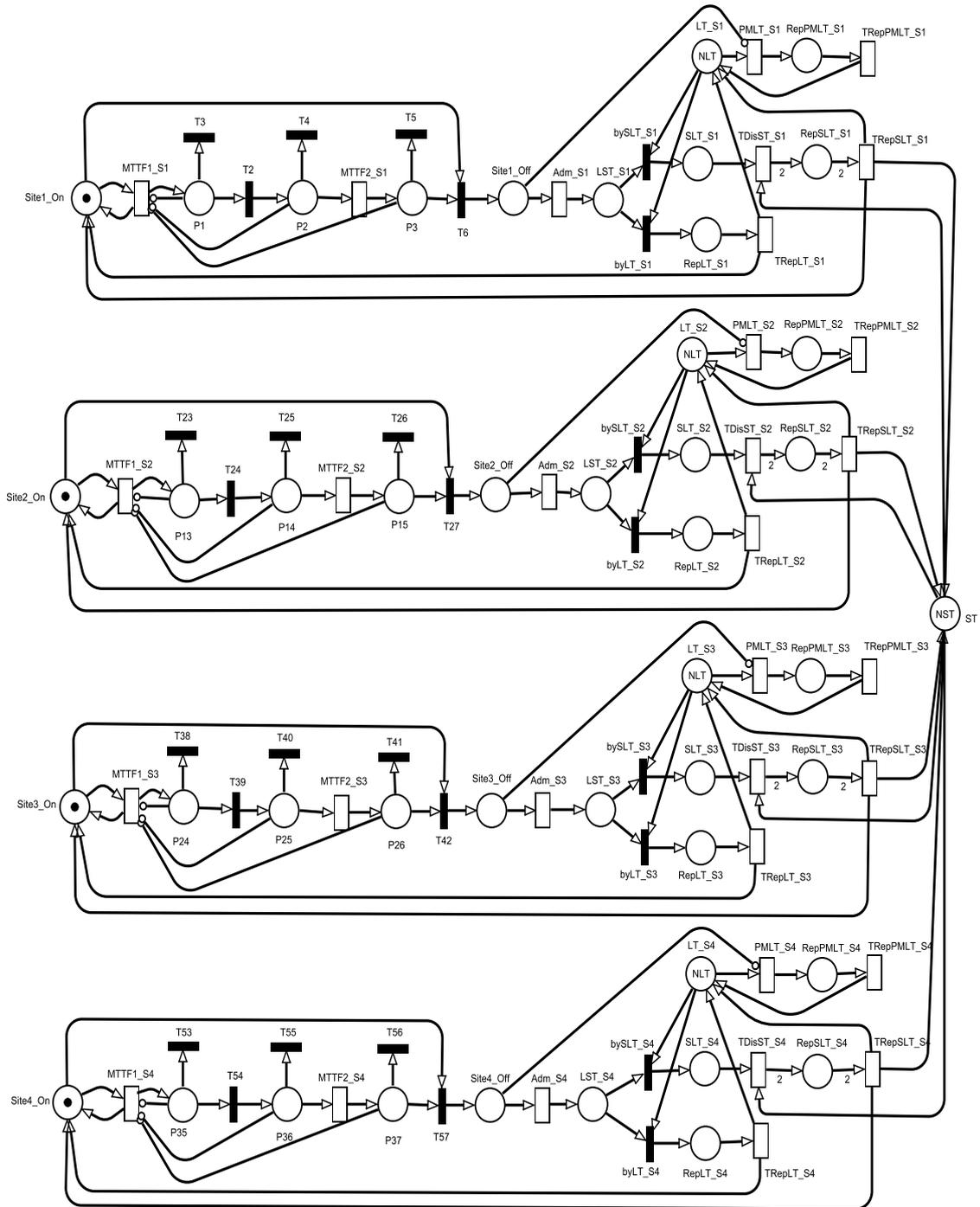


Figura 3.20: Modelo SPN Completo

## CAPÍTULO 4

# ESTUDO DE CASO

Este capítulo apresenta inicialmente os cálculos dos custos para a política de manutenção e suas equações. Em seguida apresenta informações de um contrato SLA e os cenários criados para o acompanhamento dos parâmetros do contrato. O objetivo é investigar o impacto de uma política de manutenção sobre a disponibilidade do sistema. As análises consideram variações no período entre as realizações de manutenções preventivas e diferentes tempos de reparo. Os resultados das avaliações permitem o gerenciamento de níveis de serviços.

### 4.1 DEFINIÇÃO DOS CUSTOS

Primeiramente foram definidos os custos de manutenção, em seguida são apresentadas as equações (Seção 4.2) que permitem a realização dos cálculos para contabilizar o custo anual de cada tipo de manutenção, os quais são feitos considerando os custos de mão de obra das equipes e os custos de reparo dos componentes. As três subseções seguintes apresentam valores utilizados para calcular os custos de manutenção.

#### 4.1.1 Custo de Mão de Obra das Equipes

O custo de mão de obra das equipes representa o valor gasto com as equipes de manutenção. No contrato, é estipulado um valor fixo anual sem contabilizar as horas de trabalho da equipe, diminuindo o custo consideravelmente, em relação ao valor normal cobrado sem contrato. O custo de mão de obra de um determinado tipo de equipe considera a qualificação da equipe, o tipo de manutenção que ela está realizando e o total de manutenções realizadas, ou seja, o valor é pago por visita ao *site*.

A equipe local é uma equipe técnica, portanto, o custo dessa equipe é menor, quando comparado ao custo da equipe especialista. Já a equipe especialista, tem conhecimentos específicos para realizar o reparo corretivo nos componentes de um *site*. Observe na Tabela 4.1 os valores atribuídos à mão de obra das equipes (por visita). Quando o reparo corretivo é feito pelas duas equipes, é considerada a soma dos custos da equipe local e especialista.

#### 4.1.2 Custo de Reparo dos Componentes

O custo médio de reparo dos componentes de cada *site* é calculado de acordo com a quantidade de componentes que ele possui. O custo médio de reparo de um compo-

**Tabela 4.1:** Custos de Mão de Obra das Equipes

Tipo de Equipe	Tipo de Manutenção	Custo de Mão de Obra (R\$)
Local	Preventiva	115,00
Local	Corretiva	172,50
Especialista	Corretiva	258,75
Local e Especialista	Corretiva	431,25

nente representa 10% do seu valor de aquisição. O valor de aquisição e de reparo dos componentes podem ser observados na Tabela 4.2.

**Tabela 4.2:** Custos de Aquisição e de Reparo dos Componentes de um *Site*

Componentes	Custo de Aquisição (R\$)	Custo de Reparo (R\$)
Roteador	754,02	75,40
<i>Switch</i>	828,57	82,85
<i>Link</i>	317,37	31,73
Computador	949,00	94,90

O custo médio de reparo de um roteador é igual a R\$ 75,40, o reparo de um *switch* custa em média R\$ 82,85, o reparo do link custa em média R\$ 31,73 e o custo médio de reparo de um computador é R\$ 94,90. O custo de reparo dos componentes considera a quantidade de componentes que devem ser reparados e o total de reparos no período de um ano. Vale enfatizar que todos os *sites* possuem um roteador, um *switch*, um *link* e 30 computadores. Quando o reparo é preventivo é considerado o custo de reparo de todos os componentes de um *site*. Porém, quando o reparo é corretivo, é considerado o custo de reparo de um roteador, de um *switch*, de um *link* e de 10 computadores.

A análise do modelo RBD de segundo nível (Seção 3.4.2) apresenta o MTTF dos computadores de um *site*, considerando que pelo menos 70% (21 PCs) deles devem funcionar para que o sistema não falhe. Assim, se ocorrer a falha simultânea em mais de 9 computadores, o sistema irá falhar. Neste sentido, o reparo corretivo para os componentes de um *site* inclui o custo médio de reparo de um roteador, de um *switch*, de um *link* e de 10 computadores. A Tabela 4.3 apresenta o valor total para o reparo dos componentes de um *site*, quando o reparo é preventivo e quando o reparo é corretivo.

### 4.1.3 Custo de um Site

O custo médio de um *site* considera a soma dos gastos com mão de obra e reparo dos componentes. A Tabela 4.4 apresenta o custo de um *site* de acordo com o tipo de manutenção que é realizada e o tipo de equipe que realiza a manutenção.

**Tabela 4.3:** Custo Total de Reparo dos Componentes de um *Site*

<b>Componentes</b>	<b>Reparo dos Componentes (R\$)</b>
Roteador	75,40
<i>Switch</i>	82,85
<i>Link</i>	31,73
10 Computadores	949,00
30 Computadores	2.847,00
<b>Total com 10 PCs</b>	<b>1.138,98</b>
<b>Total com 30 PCs</b>	<b>3.036,98</b>

**Tabela 4.4:** Custos de um *Site*

<b>Tipo de Equipe</b>	<b>Tipo de Manutenção</b>	<b>Custo de um <i>Site</i> (R\$) (Pro Visita)</b>
Local	Preventiva	3.152,00
Local	Corretiva	1.311,50
Local e Especialista	Corretiva	1.570,25

## 4.2 CUSTOS DE MANUTENÇÃO

O custo de manutenção é dividido em três partes principais, descritas a seguir:

- O custo associado à manutenção preventiva, realizada pela equipe local.
- O custo associado à manutenção corretiva, realizada somente pela equipe local.
- O custo associado à manutenção corretiva, realizada pela equipe local e especialista.

É importante enfatizar que o custo associado às manutenções considera o custo de mão de obra da equipe e o custo de reparo dos componentes de um *site*. O custo de mão de obra de um determinado tipo de equipe considera a qualificação da equipe, o tipo de manutenção que ela está realizando e a quantidade de vezes que ela realiza a manutenção, ou seja, o valor é cobrado por visita ao *site*. O custo de reparo dos componentes considera a quantidade de componentes que devem ser reparados e a quantidade de vezes que eles serão reparados. As subseções a seguir apresentam as equações para calcular o custo anual das manutenções. A notação utilizada é referente à notação do *software* TimeNet 4.1 [Zim12].

### 4.2.1 Custo da Manutenção Preventiva

O custo associado à manutenção preventiva realizada pela equipe local contempla o custo de mão de obra da equipe local e o custo de reparo dos componentes. Quando a manutenção é preventiva, é levado em conta que todos os componentes de um *site* serão reparados. Vale ressaltar que para cada *site*, os cálculos são realizados separadamente. A Equação 4.1 estima o custo da manutenção preventiva, que é realizada pela equipe local.

$$C_{pm} = [((P\{\#LT\_Site > 0\}) \times (1/PMLT\_Site)) \times (C_{LTpm} + CRc) \times T] \quad (4.1)$$

Onde:

- $P\{\#LT\_Site > 0\}$  é a probabilidade de haver marcas lugar  $LT\_Site$ .
- $(1/PMLT\_Site)$  representa a taxa (inverso do tempo médio) da transição  $PMLT\_Site$ , a qual indica o tempo entre a realização de manutenções preventivas.
- $C_{LTpm}$  é o custo da equipe local, relativo às manutenções preventivas.
- $CRc$  representa o custo de reparo dos componentes de um *site*.
- $T$  representa o período de cálculo do custo da manutenção.

### 4.2.2 Custo da Manutenção Corretiva - Equipe Local

O custo associado à manutenção corretiva realizada somente pela equipe local contempla o custo de mão de obra da equipe local, relativo às manutenções corretivas e o custo de reparo dos componentes. Quando a manutenção é corretiva, é levado em conta que apenas uma determinada quantidade de componentes de um *site* é reparada. A Equação 3.3 estima o custo da manutenção corretiva, quando realizada pela equipe local.

$$C_{cmlt} = [((NLT - ((E\{\#RepLT\_Site\}) + (E\{\#SLT\_Site\}) + ((E\{\#RepSLT\_Site\})/2))) \times (C_{LTcm} + CRc) \times T] \quad (4.2)$$

Onde:

- $NLT$  é o número de pessoas da equipe local.
- $E\{\#RepLT\_Site\}$  é a esperança de haver marcas no lugar  $RepLT\_Site$ .
- $E\{\#SLT\_Site\}$  é a esperança de haver marcas no lugar  $SLT\_Site$ .

- $E\{\#RepSLT\_Site\}$  é a esperança de haver marcas no lugar  $RepSLT\_Site$ . O resultado é dividido por 2 devido à possibilidade de ter diferentes tipos de equipes no lugar  $RepSLT\_Site$ .
- $C\_LTcm$  é o custo da equipe local, relativo às manutenções corretivas.
- $CRc$  representa o custo de reparo dos componentes de um *site*.
- $T$  representa o período de cálculo do custo da manutenção.

#### 4.2.3 Custo da Manutenção Corretiva - Equipe Local e Especialista

O custo associado à manutenção corretiva realizada pela equipe local e especialista contempla o custo de mão de obra das duas equipes relativo às manutenções corretivas e o custo de reparo dos componentes. Quando a manutenção é corretiva, é levado em conta que apenas uma determinada quantidade de componentes de um *site* é reparada. A Equação 4.3 estima o custo da manutenção corretiva, quando realizada pela equipe local e especialista.

$$C\_cmslt = [(NST - ((E\{\#RepSLT\_Site\})/2) \times (C\_SLTcm + CRc) \times T] \quad (4.3)$$

Onde:

- $NST$  é o número de pessoas da equipe especialista.
- $E\{\#RepSLT\_Site\}$  é a esperança de ter marcas no lugar  $RepSLT\_Site$ . O resultado é dividido por 2 devido à possibilidade de ter diferentes tipos de equipes no lugar  $RepSLT\_Site$ .
- $C\_SLTcm$  é custo da equipe local e especialista, relativo às manutenções corretivas.
- $CRc$  representa o custo de reparo dos componentes de um *site*.
- $T$  representa o período de cálculo do custo da manutenção.

### 4.3 DEFINIÇÃO DOS CENÁRIOS E DOS PARÂMETROS DO CONTRATO SLA

Cada vez mais as empresas estão contratando serviços baseados em níveis de serviços (SLAs). Portanto, para administrar esses contratos é necessário que as declarações de um prestador de serviços (que deve realizar um serviço de acordo com as garantias acordadas) sejam gerenciadas pelas partes envolvidas no contrato. O gerenciamento dos níveis de serviços do contrato SLA (*Service Level Management* - SLM) é realizado através do acompanhamento dos parâmetros do contrato, para identificar o desempenho do serviço e os parâmetros que afetam seu provimento. O SLM permite um planejamento mais

eficiente do negócio e a administração de indicadores, que posteriormente poderão ser melhorados.

Os cenários concebidos para a realização do estudo de caso visa investigar o impacto de uma política de manutenção, considerando a realização de manutenções preventivas e corretivas. Inicialmente, são apresentados alguns parâmetros básicos para um contrato SLA e, posteriormente, são realizadas as avaliações no modelo SPN proposto para verificar os níveis de serviços alcançados em cada *site*. A Tabela 4.5 apresenta para cada classe de SLA a disponibilidade, o número de noves da disponibilidade e o *downtime* correspondente. São apresentadas nove classes de SLAs com níveis de serviços previamente especificados.

**Tabela 4.5:** Parâmetros de um Contrato SLA

SLA	Disponibilidade (%)	9s	<i>Downtime</i> (h)
I	90%	1	876
II	91%	1,04	788,40
III	92%	1,09	700,80
IV	93%	1,15	613,20
V	94%	1,22	525,60
VI	95%	1,30	438,00
VII	96%	1,39	350,40
VIII	97%	1,52	262,80
IX	98%	1,69	175,20

Como apresentado, o SLA de nível I garante disponibilidade de 90%, o que é o mesmo que uma paralisação máxima (*downtime*) de 876 horas, considerando o período de um ano. O SLA II garante disponibilidade de 91%, o que corresponde a 788,40 horas de inatividade. O SLA III oferece disponibilidade de 92%, com 700,80 horas associadas ao tempo de inatividade. OS SLAs IV, V, VI, VII, VIII e IX apresentam disponibilidade de 93%, 94%, 95%, 96%, 97% e 98% respectivamente.

#### 4.3.1 Avaliação dos Cenários

A fim de investigar os níveis de serviços, foram feitas análises no modelo SPN utilizando parâmetros que compreendem diferentes cenários. O objetivo das avaliações é a captura dos níveis de disponibilidade. Além disso, os tempos de inatividade do *site* também são verificados, caso o valor observado seja maior que o valor especificado no SLA, multas serão aplicadas à empresa de manutenção. Para cada classe de SLA, é apresentado um valor de multa correspondente.

O tempo médio para falhas (MTTF) de um *site* obtido com a avaliação do modelo RBD de terceiro nível é igual a 389,17 horas, considerando a taxa de falhas dos componentes já citados. O tempo médio de reparo (MTTR) para as manutenções preventivas e

corretivas são iguais e para cada classe de SLA é apresentado um MTTR correspondente.

O estudo de caso objetiva analisar o impacto da política de manutenção através de variações nos intervalos entre manutenções preventivas e no MTTR das manutenções preventivas e corretivas. Os cenários criados contemplam os intervalos entre manutenções preventivas e MTTRs apresentados na Tabela 4.6. O intervalo para a realização de manutenções preventivas para os SLAs I, II e III é igual a 6.480 horas, isso equivale à realização de manutenções preventivas a cada seis meses e os MTTRs variam entre 72, 48 e 24 horas. Para os SLAs IV, V e VI é assumido que as manutenções preventivas são realizadas a cada 2.160 horas (três meses) com variações de MTTR entre 72, 48 e 24 horas. Por fim, para os SLAs VII, VIII e IX, o intervalo entre realização de manutenções preventivas é de 720 horas (uma vez por mês). O valor das multas pode ser observado na Tabela 4.6.

**Tabela 4.6:** Parâmetros dos Cenários

SLA	Intervalos PM (h)	MTTR (h)	Multa (R\$)
I	6.480	72	125,00
II	6.480	48	150,00
III	6.480	24	180,00
IV	2.160	72	216,00
V	2.160	48	259,20
VI	2.160	24	311,04
VII	720	72	373,25
VIII	720	48	447,90
IX	720	24	537,48

É importante lembrar que o MTTR da manutenção corretiva realizada pela equipe local se refere à soma do tempo administrativo com o tempo médio de reparo de um *site*. Já o MTTR para a manutenção corretiva realizada pela equipe especialista (SLT) se refere à soma do tempo administrativo, do tempo de deslocamento e do tempo médio de reparo de um *site*. O tempo de deslocamento da equipe especialista é diferente para cada *site*. Essa é a única diferença entre os parâmetros dos *sites*. Portanto, o estudo de caso contempla as avaliações apenas para o *site* 1, cujo tempo de deslocamento é de 1,85 horas. Mas, vale ressaltar que a equipe especialista é compartilhada pelos quatro *sites* apresentados no modelo SPN.

A disponibilidade percebida pode ser menor que a oferecida no SLA, neste caso, o tempo de inatividade é utilizado para calcular o valor da multa. Outra questão que precisa ser destacada é que o modelo SPN proposto apenas considera a realização de manutenções preventivas e corretivas, porém, não é considerada a diminuição de manutenções corretivas quando manutenções preventivas são realizadas com uma maior frequência.

Os resultados obtidos através de análise estacionária do *site* 1 podem ser observados na Tabela 4.7. São apresentados para os nove cenários, a disponibilidade percebida, o

*downtime* e o valor da multa, que deve ser aplicado pelo não cumprimento do serviço oferecido no SLA.

**Tabela 4.7:** Resultados das Avaliações dos Cenários

Cenários	Disponibilidade	<i>Downtime</i> (h)	Multa (R\$)
I	84, 53%	1.355, 55	59.943, 69
II	89, 04%	959, 95	25.733, 11
III	94, 08%	518, 19	0, 00
IV	85, 18%	1.298, 01	14.7919, 80
V	89, 46%	923, 13	103.039, 46
VI	94, 28%	500, 68	19.494, 76
VII	86, 93%	1.144, 99	296.578, 41
VIII	90, 59%	824, 55	251.606, 82
IX	94, 82%	453, 59	149.630, 23

Adicionalmente, o custo anual para a manutenção preventiva, o custo anual para a manutenção corretiva realizada pela equipe local, o custo anual para a manutenção corretiva realizada pela equipe local e especialista e o custo do contrato anual são apresentados na Tabela 4.8, cuja unidade monetária é R\$.

**Tabela 4.8:** Resultados das Avaliações dos Cenários - Custo das Manutenções

Cenários	Custo Anual da Manutenção Preventiva (LT)	Custo Anual da Manutenção Corretiva (ST)	Custo Anual da Manutenção Corretiva (SLT)	Custo Anual do Contrato de Manutenção
I	3.670, 30	10.130, 77	12.129, 50	22.260, 27
II	3.880, 61	9.640, 20	11.542, 15	21.182, 35
III	4.115, 52	6.816, 27	8.161, 07	14.977, 34
IV	10.848, 64	9.703, 39	11.617, 80	21.321, 19
V	11.517, 52	9.270, 26	11.099, 22	20.369, 48
VI	12.271, 77	6.585, 35	7.884, 59	14.469, 94
VII	31.129, 47	8.563, 71	10.253, 27	18.816, 98
VIII	33.453, 12	8.282, 23	9.916, 26	18.198, 49
IX	36.146, 53	5.966, 10	7.143, 17	13.109, 27

Vale enfatizar que os resultados apresentados correspondem as avaliações do *site* 1.

A disponibilidade é obtida através da Equação 2.19. O *downtime* é calculado através da Equação 4.4.

$$Down = (1 - Disponibilidade)8760 \quad (4.4)$$

O valor da multa é calculado através da Equação 4.5.

$$M = (Down - DownSLA)Multa \quad (4.5)$$

Onde:

- *Down* corresponde ao valor de *downtime* alcançado com a avaliação do modelo SPN.
- *DownSLA* corresponde ao valor de *downtime* oferecido no contrato SLA.
- *Multa* corresponde ao valor da multa para a determinada classe de SLA.

Com a avaliação do primeiro cenário a disponibilidade percebida é de 84,53%, o mesmo que uma paralisação de 1.355,55 horas por ano. Comparando o resultado percebido com a disponibilidade oferecida no SLA de nível I, pode-se concluir que os serviços não foram cumpridos. Ou seja, o intervalo de 6.480 horas (seis meses) entre manutenções preventivas e MTTR de 72 horas, são insuficientes para fornecer uma disponibilidade de 90% com a taxa de falhas do *site* sendo igual a 389,17 horas. Nesse caso, uma multa no valor de R\$ 59.943,69 deverá ser aplicada à empresa de manutenção. O custo anual da manutenção preventiva é de R\$ 3.670,30. Esse valor relativamente baixo é explicado devido a realização de apenas duas manutenções preventivas no período de um ano. O custo anual das manutenções corretivas, realizada somente pela equipe local é de R\$ 12.129,50. O custo anual das manutenções corretivas, realizada pela equipe local e especialista é de R\$ 22.260,27. O custo anual do contrato para o cenário I corresponde a R\$ 22.260,27. Como pode ser observado, o custo da manutenção preventiva é consideravelmente menor que os custos das manutenções corretivas. Mas, é importante enfatizar que o MTTF do *site* é muito baixo, o que influencia a realização de manutenções corretivas, elevando os custos.

Com a avaliação do segundo cenário a disponibilidade percebida é de 89,04%, o mesmo que uma paralisação de 959,95 horas por ano. Comparando o resultado percebido com a disponibilidade oferecida no SLA de nível II, pode-se concluir que os serviços não foram cumpridos como esperado. Ou seja, o intervalo de 6.480 horas entre manutenções preventivas e MTTR de 48 horas, são insuficientes para fornecer uma disponibilidade de 91% com a taxa de falhas do *site* sendo igual a 389,17 horas. Nesse caso, uma multa no valor de R\$ 25.733,11 deverá ser aplicada à empresa de manutenção. O custo anual da manutenção preventiva é de R\$ 3.880,61. Os custos das manutenções corretivas podem ser observados na Tabela 4.8. O custo anual do contrato para o cenário II corresponde a R\$ 21.182,35.

Com a avaliação do terceiro cenário a disponibilidade percebida é de 94,08%, o mesmo que uma paralisação de 518,19 horas por ano. Comparando o resultado percebido com

a disponibilidade oferecida no SLA de nível III, pode-se concluir que os serviços foram cumpridos satisfatoriamente. Ou seja, o intervalo de 6.480 horas entre manutenções preventivas e MTTR de 24 horas, são suficientes para fornecer uma disponibilidade acima de 92%, mesmo com a taxa de falhas do *site* sendo igual a 389,17 horas. Nesse caso, nenhuma multa será aplicada à empresa de manutenção. O custo anual da manutenção preventiva é de R\$ 4.115,52. Os custos das manutenções corretivas e do contrato podem ser observados na Tabela 4.8.

Com a avaliação do quarto cenário a disponibilidade percebida é de 85,18%, o mesmo que uma paralisação de 1.298,01 horas por ano. Comparando o resultado percebido com a disponibilidade oferecida no SLA de nível IV, pode-se concluir que os serviços não foram cumpridos como esperado. Ou seja, o intervalo de 2.160 horas (três meses) entre manutenções preventivas e MTTR de 72 horas, são insuficientes para fornecer uma disponibilidade de 93% com a taxa de falhas apresentada. Nesse caso, uma multa no valor de R\$ 14.7919,80 deverá ser aplicada à empresa de manutenção. O custo anual da manutenção preventiva é de R\$ 10.848,64. Os custos das manutenções corretivas e do contrato podem ser observados na Tabela 4.8.

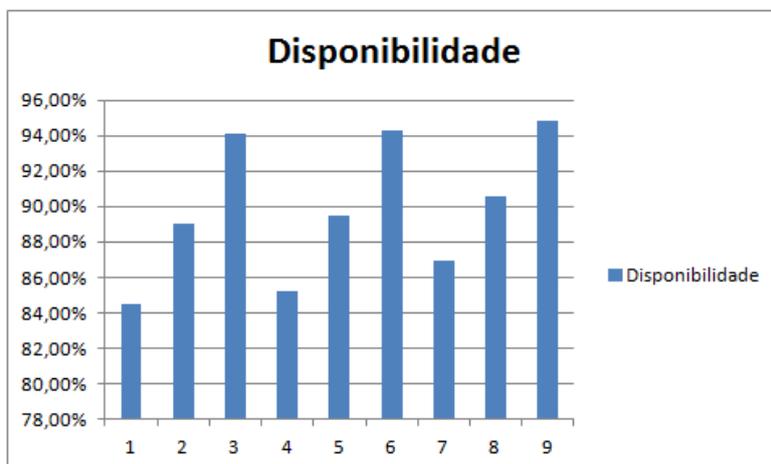
Com as avaliações dos cenários 5 e 6 as disponibilidades percebidas são de 89,46% e 94,28%, o mesmo que uma paralisação de 923,13 e 500,68 horas por ano. Comparando os resultados percebidos com a disponibilidade oferecida nos SLAs de nível IV e VI, respectivamente, pode-se concluir que os serviços não foram cumpridos como esperado. Ou seja, o intervalo de 2.160 horas (três meses) entre manutenções preventivas e MTTRs de 48 e 24 horas, são insuficientes para fornecer disponibilidades de 94% e 95% com a taxa de falhas apresentada. O valor das multas, os custos das manutenções e do contrato podem ser observados na Tabela 4.8.

Com as avaliações dos cenários 7, 8 e 9 as disponibilidades percebidas são de 86,93%, 90,59% e 94,82%, respectivamente. Comparando os resultados percebidos com a disponibilidade oferecida nos SLAs de nível VII, VIII e IX, pode-se concluir que os serviços não foram cumpridos como esperado. Ou seja, o intervalo de 720 horas (um mês) entre manutenções preventivas e MTTRs de 72, 48 e 24 horas, são insuficientes para fornecer disponibilidades de 96%, 97% e 98%, considerando que a taxa de falhas é de 389,17 horas. O valor de *downtime*, das multas, dos custos das manutenções e do contrato podem ser observados na Tabela 4.8.

### 4.3.2 Discussão de Resultados

Os resultados obtidos com as avaliações dos nove cenários (Tabela 4.7) foram utilizados para aferição dos níveis de serviços fornecidos por diferentes classes de SLAs (Tabela 4.6). A Figura 4.1 apresenta a disponibilidade alcançada em cada cenário.

Como era de se esperar, os resultados observados indicam que quanto maior o intervalo entre as manutenções preventivas, menor será a disponibilidade alcançada. Além disso, o parâmetro MTTR influencia consideravelmente sobre a disponibilidade. MTTRs de 24 horas aumentam significativamente os níveis de disponibilidade do *site*, apesar da taxa



**Figura 4.1:** Disponibilidade Alcançada com as Avaliações dos Cenários

de falhas ser constante. Os cenários III, VI e IX apresentam índices de disponibilidade maior que os outros cenários que apresentam MTTR maior. Porém, apenas no cenário III o índice de disponibilidade é satisfatório, quando comparado ao SLA correspondente. É importante enfatizar que os níveis de disponibilidade assumidos para o SLA servem para a verificação dos níveis de serviços alcançados. No entanto, diferentes combinações poderiam ser realizadas.

Quando as manutenções preventivas são realizadas com maior frequência, o custo associado também é maior, entretanto, os altos custos podem ser compensados com o aumento da disponibilidade. Em relação aos custos de manutenções corretivas, o valor correspondente pode ser justificado devido à quantidade de equipamentos que devem ser reparados e ao custo da mão de obra corretiva ser maior que o custo da mão de obra preventiva. Além disso, os custos são referentes ao período de um ano.

Uma questão que precisa ser ressaltada, é que as avaliações apresentadas são correspondentes do *site* 1, devido à maioria dos parâmetros ser igual para os outros três *sites* apresentados no modelo SPN completo (Figura 3.20). A única diferença é o tempo de deslocamento associado à distância que deve ser percorrida pela equipe especialista para chegar ao local, considerando que a ST se encontra na capital e os campi (*sites*) estão situados em cidades do interior.

#### 4.4 CONSIDERAÇÕES FINAIS

Este capítulo enfatizou uma descrição dos custos das manutenções, em seguida apresentou as equações para realização dos cálculos, as quais são divididas em três partes: o custo associado à manutenção preventiva, realizada pela equipe local; o custo associado à manutenção corretiva, realizada somente pela equipe local; e o custo associado à manutenção corretiva, realizada pela equipe local e especialista. Também foi dada uma definição dos cenários e dos parâmetros utilizados no modelo SPN, os quais são utilizados

para a realização do estudo de caso. Foram definidos nove cenários distintos, que permitem a verificação de níveis de serviços alcançados quando variações nos intervalos de manutenções preventivas e no MTTR são feitas. Por fim, foram discutidos os resultados.

# CONCLUSÕES

Nos últimos anos, houve um acentuado crescimento nos serviços de tráfego de dados (*www*, *e-mail*, *e-commerce*, *e-business* etc.), em detrimento dos serviços de tráfego de voz, popularização da Internet e aplicações multimídia. Assim, as operadoras de telecomunicações buscaram novas oportunidades de negócio e passaram a investir em uma segunda rede, capaz de ofertar diversos tipos de serviços.

As redes convergentes permitem a integração dos serviços de dados, voz e vídeo, sobre uma plataforma de rede unificada. O uso de uma rede única e integrada substitui a utilização de um grande número de redes separadas ou aplicações para cada um dos sistemas. Com base nessa vantagem, tem-se aumentado o número de usuários conscientes dos benefícios de contar com uma rede convergente para todos os serviços técnicos tradicionais.

A maioria das empresas e instituições deverão adotar, em um futuro razoavelmente próximo, as redes convergentes, visando uma redução substancial de despesas de operação e manutenção da rede. No entanto, quanto maior ou mais crítica for a rede, mais importante será a construção de uma infraestrutura confiável e de qualidade, pois a produtividade perdida decorrente do tempo de paralisação pode afetar um maior número de usuários. Adicionalmente, a incorporação dos diversos tipos de tráfego de rede em tempo real acrescenta um nível de complexidade que traz exigências de um controle eficaz e de soluções de gerenciamento para lidar com a complexidade introduzida por sua heterogeneidade.

A ocorrência de eventos de falhas e atividades de reparo nessas redes pode causar degradação nos atributos de dependabilidade, capazes de levar a rede a um estado não operante. Além disso, se a rede ficar indisponível, as operações podem ser interrompidas completamente. Neste sentido, este trabalho apresentou uma abordagem de modelagem hierárquica e heterogênea que considera as vantagens dos diagramas de blocos de confiabilidade (RBD), das cadeias de Markov de tempo contínuo e das redes de Petri estocásticas (SPN), para analisar o impacto de políticas de manutenção. Adicionalmente, disponibilidade, *downtime* e custos de manutenção são considerados. Os modelos RBDs devem ser adotados para mitigar os componentes que devem compor o modelo SPN sempre que dependências complexas não são consideradas.

Uma infraestrutura de rede convergente de instituições de ensino, que possui quatro polos separados geograficamente foi utilizada como referência para o estudo. Um total de nove cenários foram realizados para verificação do cumprimento dos serviços oferecidos por diferentes classes de SLAs. Os resultados obtidos através de análises estacionária do modelo SPN são comparados com os parâmetros fornecidos nas classes de SLAs para veri-

ficação de níveis de serviços. Pode-se concluir que a realização de manutenções preventivas em intervalos curtos e MTTR mais baixo impactam fortemente sobre a disponibilidade da rede. Os *softwares* utilizados para as análises dos modelos são o Sharpe [STP96] e TimeNET 4.1 [Zim12].

Os modelos heterogêneos propostos demonstram, através dos resultados apresentados, que podem ajudar no planejamento da infraestrutura de redes e permite avaliações e medições de atributos de dependabilidade para o gerenciamento dos níveis de serviços. A avaliação baseada em modelos possibilita a análise 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.

As principais contribuições deste trabalho são as proposições de modelos hierárquicos e heterogêneos para infraestrutura de redes convergentes, a avaliação do impacto de políticas de manutenção preventivas e corretivas e a verificação de níveis de serviços. O estudo de caso apresentado contempla variações nos intervalos entre manutenções preventivas e no tempo de reparo, assim os resultados podem ser observados e melhorias nas estratégias para checagem de falhas em componentes, um processo eficiente de substituição de componentes e meios redundantes para assegurar a continuidade dos serviços podem ser estudados.

Os resultados obtidos demonstram um aumento significativo na disponibilidade e diminuição acentuada no tempo de inatividade quando as manutenções preventivas são realizadas com maior frequência e os MTTFs são mais baixos, o que pode compensar os custos associados às manutenções. Além disso, o simples fato de a manutenção preventiva reduzir o risco de paradas não programadas, devido à ocorrência de falhas nos componentes, já a coloca como uma opção melhor do que a manutenção corretiva.

Vale ressaltar que o gerenciamento do SLA permite a administração de indicadores e a verificação de problemas, permitindo uma proatividade que trará benefícios para as partes envolvidas no contrato, além disso, um contrato de SLA bem feito atesta a qualidade do prestador de serviços e pode funcionar ainda como uma vantagem competitiva. Por fim, é importante concluir que os objetivos definidos inicialmente foram alcançados.

### 5.0.1 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 da rede convergente será adotado para representar a diminuição de manutenções corretivas quando as manutenções preventivas forem realizadas em intervalos mais curtos. Adicionalmente, será considerado o planejamento de níveis de serviços para contratos SLAs e estratégias para um processo eficiente de substituição de componentes e meios redundantes para assegurar a continuidade dos serviços.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [All05] R. Allan. A reliability-centred asset maintenance method for assessing the impact of maintenance in power distribution systems. In *Power Engineering Society General Meeting, 2005. IEEE*, page 2649 Vol. 3, june 2005.
- [Alp06] M. A. C. Alpendre. *Service level agreement: A concept to know how to use*. PhD thesis, University of Coimbra, 2006.
- [ALR01] A. Avizienis, J. Laprie, and B. Randell. Fundamental concepts of dependability, 2001.
- [ALRL04] A. Avizienis, J. C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions on*, 1(1):11 – 33, jan.-march 2004.
- [Bal01] Gianfranco Balbo. Introduction to stochastic petri nets. In Ed Brinksma, Holger Hermanns, and Joost-Pieter Katoen, editors, *Lectures on Formal Methods and Performance Analysis*, volume 2090 of *Lecture Notes in Computer Science*, pages 84–155. Springer Berlin / Heidelberg, 2001.
- [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.
- [BM02] U.N. Bhat and G.K. Miller. *Elements of applied stochastic processes*. Wiley series in probability and statistics. Wiley-Interscience, 2002.
- [BM03] W. R. Blischke and D. N. P. Murthy. *Case studies in reliability and maintenance / edited by Wallace R. Blischke, D.N. Prabhakar Murthy*. John Wiley, Hoboken, NJ :, 2003.
- [CAK04] Yung-Ruei Chang, S.V. Amari, and Sy-Yen Kuo. Computing system failure frequencies and reliability importance measures using obdd. *Computers, IEEE Transactions on*, 53(1):54 – 68, jan 2004.
- [CFGR95] G. Chiola, G. Franceschinis, R. Gaeta, and M. Ribaud. Greatspn 1.7: Graphical editor and analyzer for timed and stochastic petri nets. 24:47–68, 1995.

- [Cis12a] Cisco. *Cisco Router Guide*, 2012. Disponível em: [http://www.cisco.com/en/US/prod/collateral/routers/ps368/ps367/product\\_data\\_sheet0900aecd8057f3d2.pdf](http://www.cisco.com/en/US/prod/collateral/routers/ps368/ps367/product_data_sheet0900aecd8057f3d2.pdf). Acessado em: 15 de Julho de 2012.
- [Cis12b] Cisco. *Cisco Switch Guide*, 2012. Disponível em: [http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps5023/product\\_data\\_sheet0900aecd80371991.pdf](http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps5023/product_data_sheet0900aecd80371991.pdf). Acessado em: 15 de Julho de 2012.
- [CL08] C. Cassandras and S. Lafortune. *Introduction to Discrete Event Systems*. Springer, 2008.
- [CMBC93] G. Chiola, M.A. Marsan, G. Balbo, and G. Conte. Generalized stochastic petri nets: a definition at the net level and its implications. *Software Engineering, IEEE Transactions on*, 19(2):89–107, feb 1993.
- [CMT89] G Ciardo, J Muppala, and K Trivedi. Spnp: stochastic petri net package, 1989.
- [CSM<sup>+</sup>10] G. Callou, E. Sousa, P. Maciel, E. Tavares, C. Araujo, B. Silva, N. Rosa, M. Marwah, R. Sharma, A. Shah, T. Christian, J.P. Pires, and F.S. Magnani. Impact analysis of maintenance policies on data center power infrastructure. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 526–533, oct. 2010.
- [CV97] J. Cardoso and R. Valette. *Redes de Petri*. Florianópolis, 1997.
- [CXZ11] Gao Chao, Wang Xin, and Zhao Zhixin. Reliability analysis of a repairable system with step-by-step maintenance policy. In *Control Conference (CCC), 2011 30th Chinese*, pages 4346–4350, july 2011.
- [DAJ95] A.A. Desrochers and R.Y. Al-Jaar. *Applications of Petri Nets in Manufacturing Systems: Modeling, Control, and Performance Analysis*. IEEE Press, 1995.
- [DP01] G. Désoblin and H. Papini. Sla management: a key differentiator for service providers. 2001. Alcatel Telecommunications Review, 3 rd Quarter.
- [EAS98] J. Endrenyi, G. J. Anders, and A. M. L. Silva. Probabilistic evaluation of the effect of maintenance on reliability. an application [to power systems]. *Power Systems, IEEE Transactions on*, 13(2):576–583, may 1998.
- [Ebe97] Charles E. Ebeling. *An Introduction to Reliability and Maintainability Engineering*. University of Dayton, 1997.
- [FFS91] Gerard. Florin, Céline. Fraize, and Natkin. Stéphane. *Stochastic Petri nets: properties, applications and tools*. Number 31(4):669-697. Microelectronics Reliability, 1991.

- [Gal09] Sérgio. M. Lins. Galdino. *ISPN: Modelagem e Avaliação Estocástica Intervalar*. PhD thesis, Universidade Federal de Pernambuco, Centro de Informática, 2009.
- [GKZ<sup>+</sup>95] Reinhard German, Christian Kelling, Armin Zimmermann, Günter Hommel, Technische Universität Berlin, and Fachgebiet Prozedatenverarbeitung Und Robotik. Timenet - a toolkit for evaluating non-markovian stochastic petri nets. *Performance Evaluation*, 24:69–87, 1995.
- [GKZH95] Reinhard German, Christian Kelling, Armin Zimmermann, and Günter Hommel. Timenet: a toolkit for evaluating non-markovian stochastic petri nets. *Performance Evaluation*, 24:69 – 87, 1995.
- [GMMC11a] A.P. Guimaraes, P.R.M. Maciel, R. S. Matos, and K. Camboim. Impact analysis of availability on computer networks infrastructure. In *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, pages 171 –175, may 2011.
- [GMMC11b] A. Guimarães, P. Maciel, R. Matos, and K. Camboim. Dependability analysis in redundant communication networks using reliability importance. vol. 4:12–17, 2011.
- [HA12] Kingsley Hughes. Adrian. Making sense of mean time to failure, 2012. Disponível em: <http://www.zdnet.com/blog/hardware/making-sense-of-mean-time-to-failure-mttf/310>. Acessado em: 15 de Julho de 2012.
- [Her01] U. Herzog. *Formal Methods for Performance Evaluation. Lecture Notes in Computer Science*, 2090:1–37, 2001.
- [HMS02] L. Higgins, K. Mobley, and R. Smith. *Maintenance Engineering Handbook*. 2002.
- [IBG10] IBGE. Instituto brasileiro de geografia e estatística, 2010.
- [Int12] Intel. *Intel Cooler Guide*, 2012. Disponível em: <http://www.intel.com/support/processors/sb/CS-030272.htm?wapkw=coller+lga1366.pdf>. Acessado em: 15 de Julho de 2012.
- [Jai91] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons New York, 1991.
- [KM86] U. W. Kulisch and W. L. Miranker. The arithmetic of the digital computer: A new approach. *SIAM Review*, 28(1):pp. 1–40, 1986.
- [KMT09] Dong Seong Kim, F. Machida, and K.S. Trivedi. Availability modeling and analysis of a virtualized system. In *Dependable Computing, 2009. PRDC '09. 15th IEEE Pacific Rim International Symposium on*, pages 365 –371, nov. 2009.

- [KR10] James. F. Kurose and Keith. W. Ross. *Redes de Computadores e a Internet: Uma Abordagem Top-down*. Companion Website, 2010.
- [KZ03] Way. Kuo and Ming. J. Zuo. *Optimal Reliability Modeling: Principles and Applications*. New Jersey, 2003.
- [LAK92] J. C. Laprie, A. Avizienis, and H. Kopetz. *Dependability: Basic Concepts and Terminology*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1992.
- [LCE10] Hui Li, Giuliano Casale, and Tariq Ellahi. Sla-driven planning and optimization of enterprise applications. In *Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering, WOSP/SIPEW '10*, pages 117–128, New York, NY, USA, 2010. ACM.
- [Lil00] D. J Lilja. *Measuring Computer Performance: A Practitioner's Guide*. Cambridge University Press, 2000.
- [LZGS84] E. D. Lazowska, J. Zahorjan, G. S. Graham, and K. C. Sevcik. Quantitative system performance: computer system analysis using queueing networks models. In Upper Saddle River, editor, *Prentice-Hall, Inc.*, number 22, 76, NJ, EUA, 1984.
- [MA05] D.A. Menascé and V.A.F. Almeida. *Performance by Design: Computer Capacity Planning by Example*. 2005.
- [MAD05] D. A. Menascé, V. A. F. Almeida, and L. W. Dowdy. *Performance by Design: Computer Capacity Planning by Example*. 2005.
- [MAMC86] G. Balbo M. Ajmone Marsan and G. Conte. *Performance Models of Multiprocessor systems*. 1986.
- [MAMF95] G. Conte S. Donatelli M. Ajmone Marsan, G. Balbo and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. Wiley Series in Parallel Computing, 1995.
- [MGC<sup>+</sup>11] R. Matos, A. Guimaraes, K. Camboim, P. Maciel, and K. Trivedi. Sensitivity analysis of availability of redundancy in computer networks. *IARIA*, page 115 to 121, 2011.
- [Mis06] N. Mistry. The importance of policy-based resource control in future networks. In *Nortel Technical Journal*, 2006.
- [MLC96] P.R.M. Maciel, R.D. Lins, and P.R.F Cunha. *Introduction of the Petri Net and Applied*. Campinas, SP, 1996.
- [Mol82] M.K. Molloy. Performance analysis using stochastic petri nets. *Computers, IEEE Transactions on*, C-31(9):913 –917, sept. 1982.

- [MS04] J. McConnell and E. Siegel. *Practical service level management: delivering high-quality web-based services*. Cisco Press, 2004.
- [MTMJJK10] P. R. M. Maciel, K. Trivedi, R. Mathias JR, and D. Kim. *Dependability Modeling In: Performance and Dependability in Service Computing: Concepts, Techniques and Research Directions*. Ed. Hershey: IGI Global, Pennsylvania, USA., 2010.
- [Mur89] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, apr 1989.
- [NGB<sup>+</sup>05] Kiran Nagaraja, G. Gama, R. Bianchini, R.P. Martin, Jr. Meira, W., and T.D. Nguyen. Quantifying the performability of cluster-based services. *Parallel and Distributed Systems, IEEE Transactions on*, 16(5):456–467, may 2005.
- [NKS06] Satish Natti, Mladen Kezunovic, and Chanan Singh. Sensitivity analysis on the probabilistic maintenance model of circuit breaker. In *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, pages 1–7, june 2006.
- [NM05] A. T. Nassif and S. A. J. Martins. Convergência das redes de comunicação: Aspectos técnicos e econômicos. *Rev. Fac. Ing. - Univ. Tarapacá [online]*, vol.13, n.2, pp. 13-19, 2005. ISSN 0718-1337.
- [Oli11] Alex. Oliveira. As redes convergentes, uma aliada na tecnologia do ensino, Janeiro 2011.
- [Pal10] I. Palade. Reliability and availability engineering for voip communications systems. In *Communications (COMM), 2010 8th International Conference on*, pages 309–312, june 2010.
- [Par88] Behrooz Parhami. From defects to failures: a view of dependable computing. *SIGARCH Comput. Archit. News*, 16(4):157–168, September 1988.
- [Pet66] C.A Petri. Kommunikation mit automaten. 1966.
- [Pet77] J. L Peterson. *Petri Nets*, volume Vol. 9, No. 3,. Computing Surveys,, 1977.
- [PKMK11] Maciel Paulo, S. Trivedi Kishor, Rivalino Matias, and Dong Kim. *Dependability Modeling*, volume 1. IGI Global, Pennsylvania, 2011.
- [Rei85] Wolfgang Reisig. *Petri nets: an introduction*. Springer-Verlag New York, Inc., New York, NY, USA, 1985.
- [Rel10] ReliaSoft. Blocksim: System reliability and maintainability analysis software tool, 2010.

- [Rel12] ReliaSoft. Blocksim: System reliability and maintainability analysis software tool, 2012. Disponível em: <http://blocksim.reliasoft.com/>. Acessado em: 13 de Fevereiro de 2012.
- [RH04a] Marvin Rausand and Arnljot Høyland. *System reliability theory: models, statistical methods, and applications*. 2004.
- [RH04b] Marvin Rausand and Arnljot Høyland. *System reliability theory: models, statistical methods, and applications*. 2004.
- [San08] J. Santos. Comunicação unificada, a velha nova ordem. In *CIO gestão*, 2008.
- [Sem08] G. Semaan. Designing networks with the optimal availability. In *Optical Fiber communication/National Fiber Optic Engineers Conference, 2008. OFC/NFOEC 2008. Conference on*, pages 1 –6, feb. 2008.
- [Shi04] V.T.S. Shi. Evaluating the performability of tactical communications networks. *Vehicular Technology, IEEE Transactions on*, 53(1):253 – 260, jan. 2004.
- [SJ06] Antje Stobbe and Tobias Just. *IT, telecoms & New Media: The dawn of technological convergence*. Deutsche Bank Research, Germany, 2006.
- [Smi05] D.J. Smith. Reliability, maintainability and risk: Practical methods for engineers. In *Butterworth-Heinemann*, 2005.
- [SMJ00] R. Sturm, W. Morris, and M. Jander. *Foundations of Service Level Management*. Sams Publishing, 2000.
- [SMLM12] E. Sousa, P. Maciel, F. Lins, and M. Marinho. Maintenance policy and its impact on the performability evaluation of eft systems. 2012. International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.2, No.2, April 2012.
- [SMT<sup>+</sup>10] B. Silva, P. Maciel, E. Tavares, C. Araujo, G. Callou, E. Sousa, N. Rosa, M. Marwah, R. Sharma, A. Shah, T. Christian, and J.P. Pires. Astro: A tool for dependability evaluation of data center infrastructures. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 783 –790, oct. 2010.
- [Sou09] E. T. G. Sousa. *Avaliação do Impacto de uma Política de Manutenção na Performabilidade de Sistemas de Transferência Eletrônica de Fundos*. PhD thesis, Universidade Federal de Pernambuco, 2009.
- [ST87] Robin A. Sahner and Kishor S. Trivedi. Reliability modeling using sharpe. *Reliability, IEEE Transactions on*, R-36(2):186 –193, june 1987.

- [ST00] A. Sesmun and L.F. Turner. Using performability in the design of communication networks. *Computers and Digital Techniques, IEE Proceedings* -, 147(5):355 –363, oct 2000.
- [STP96] Robin A. Sahner, Kishor S. Trivedi, and Antonio Puliafito. *Performance and reliability analysis of computer systems: an example-based approach using the SHARPE software package*. Kluwer Academic Publishers, Norwell, MA, USA, 1996.
- [TDe12] Network. Group. TDec. Educação, Junho 2012.
- [The12] Thermaltake. *Thermaltake Guide*, 2012. Disponível em: <http://www.thermaltakeusa.com/Product.aspx?C=1246&ID=1975#Tab1>. Acessado em: 15 de Julho de 2012.
- [TKRM09a] K.S. Trivedi, Dong Seong Kim, A. Roy, and D. Medhi. Dependability and security models. In *Design of Reliable Communication Networks, 2009. DRCN 2009. 7th International Workshop on*, pages 11 –20, oct. 2009.
- [TKRM09b] K.S. Trivedi, Dong Seong Kim, A. Roy, and D. Medhi. Dependability and security models. In *Design of Reliable Communication Networks, 2009. DRCN 2009. 7th International Workshop on*, pages 11 –20, oct. 2009.
- [TM93] K.S. Trivedi and M. Malhotra. Reliability and performability techniques and tools: A survey. In *Messung, Modellierung und Bewertung von Rechen- und Kommunikationssystemen*, pages 27 – 48, 1993.
- [VPMM06] Sergiy A. Vilkomir, David L. Parnas, Veena B. Mendiratta, and Eamonn Murphy. Segregated failures model for availability evaluation of fault-tolerant systems. In *Proceedings of the 29th Australasian Computer Science Conference - Volume 48, ACSC '06*, pages 55–61, Darlinghurst, Australia, Australia, 2006. Australian Computer Society, Inc.
- [Wil04] A. Willig. *Performance Evaluation Techniques*. Summer, 2004.
- [WP06] H. Wang and H. Pham. *Reliability and Optimal Maintenance*. Springer Verlag, 2006.
- [XPD04] Min Xie, Kim-Leng Poh, and Yuan-Shun Dai. *Computing System Reliability: Models and Analysis*. Springer, 2004.
- [YPMM11] S.Y. Yerima, G.P. Parr, S. McCLean, and P.J. Morrow. Modelling and evaluation of a policy-based resource management framework for converged next generation networks. In *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*, pages 563 –570, may 2011.

- [Zim12] A. Zimmermann. Modeling and evaluation of stochastic petri nets with timenet 4.1. 2012. Ilmenau, Germany. Disponível em: <http://valuetools.org/2012/show/program-final>. Acessado em: 12 de Agosto de 2012.
- [ZJKK07] W. Zou, M. Janic, R. Kooij, and F. Kuipers. On the availability of networks. Dec. 2007. in Proc. of BroadBand Europe 2007, Antwerp.
- [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.
- [ZLYY09] Shangqin Zhong, Chen Liu, Zhengqiu Yang, and Danfen Yan. Privacy protection model for distributed service system in converged network. In *E-Business and Information System Security, 2009. EBISS '09. International Conference on*, pages 1 –5, may 2009.

