Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
Departamento de Engenharia de Computação e Automação Industrial

Algoritmo de Detecção de Falhas para Sistemas Telefônicos Utilizando a Teoria do Perigo

Autor: José Carlos Lima Pinto
Orientador: Prof. Dr. Fernando José Von Zuben

Tese de Mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos pré-requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica

Banca Examinadora:
Prof. Dr. Fernando José Von Zuben (DCA / FEEC / Unicamp)
Dr. Daniel Moutinho Pataca (GPA / DTVD / CPqD)
Prof. Dr. Marco Aurélio Amaral Henrique (DCA / FEEC / Unicamp)
Prof. Dr. Paulo Cardieri (DECOM / FEEC / Unicamp)

Área de Concentração: Engenharia de Computação

Campinas – SP – Brasil

Setembro de 2006

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

Pinto, José Carlos Lima
P658a Algoritmo de detecção de falhas para sistemas telefônicos utilizando a teoria do perigo / José Carlos Lima Pinto. -- Campinas, SP: [s.n.], 2006.

Orientador: Fernando José Von Zuben
Dissertação (Mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Localização de falhas (Engenharia). 2. Sistema telefônicos digitais. 3. Inteligência artificial. 4. Sistema imune. I. Von Zuben, Fernando José. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Título em Inglês: Fault detection algorithm for telephone systems using the danger theory
Palavras-chave em Inglês: Artificial immune systems, Danger theory, Fault detection, Telephone systems
Área de concentração: Engenharia de Computação
Titulação: Mestre em Engenharia Elétrica
Banca examinadora: Daniel Moutinho Pataca, Marco Aurélio Henrique, Paulo Cardieri
Data da defesa: 27/09/2006
Programa de Pós-Graduação: Engenharia Elétrica

BANCA EXAMINADORA

Prof. Dr. Fernando José Von Zuben

(DCA / FEEC / Unicamp)

Dr. Daniel Moutinho Pataca

(GPA / DTVD / CPqD)

Prof. Dr. Marco Aurélio Amaral Henriques

(DCA / FEEC / Unicamp)

Prof. Dr. Paulo Cardieri

(DECOM / FEEC / Unicamp)

Agradecimentos

Aos amigos e colegas de trabalho pelo incentivo nesses anos de pós-graduação.

Aos membros da banca examinadora pelas valiosas críticas e sugestões.

Ao Prof. Dr. Fernando José Von Zuben pela orientação preciosa e pelo exemplo de dedicação ao ensino e à pesquisa.

À minha esposa e à minha filha pelo amor e compreensão em todos os momentos da minha vida.

A Deus.

À Carine e à Victória.

Trabalhos de Autoria Relacionados ao Tema da Tese

Publicação

PINTO, J. C. L. & VON ZUBEN, F. J. (2005). "Fault Detection Algorithm for Telephone Systems Based on the Danger Theory". In: Christian Jacob; Marcin Pilat; Peter Bentley; Jonathan Timmis. (Eds.). Proceedings of the International Conference on Artificial Immune Systems, Lecture Notes in Computer Science, v. 3627, p. 418-431, Berlin: Springer-Verlag.

Trabalhos Técnicos (CPqD)

PINTO, J. C. L. & ANTONINI, J. O. C. (2002). "Análise da Descrição do Serviço TPP".

URSINI, E. L., LAVELHA, A. C., PINTO, J. C. L. (2002). "Caracterização do Tráfego do Serviço Terminal Fixo Pré-Pago".

PINTO, J. C. L. & BASTREGHI, J. (2002). "Otimização e Planejamento da Evolução da Plataforma de Rede Inteligente da Angola Telecom".

BRATFISCH, M. V. C. & PINTO, J. C. L. (2001). "Análise dos Dados dos Sistemas de Mediação, Billing e Interconexão Vésper".

MANDEL, M. C. R. B., BALDIN, E. Y., BASTREGHI, J., PINTO, J. C. L., FIGUEIREDO, M. S. (2001). "Implantação de Ambientes Multifornecedores de Rede Inteligente".

MANDEL, M. C. R. B., LAVELHA, A. C., CASSANO, C., URSINI, E. L., BENJOVENGO, E. P., BASTREGHI, J., PINTO, J. C. L., RIOS, Martin, J. M. (2001). "Considerações sobre o Projeto de Implantação do Serviço Móvel Pessoal (SMP) da Telemar".

MANDEL, M. C. R. B., MARTINS, J. E., PINTO, J. C. L., ÁVILA, I. M. A. (1998). "Teste do Protocolo INAP-BR e do Padrão de Bilhetagem AMA-BR nas Centrais de Comutação da ERICSSON".

ÍNDICE

Resumo

Essa dissertação apresenta um algoritmo de detecção de falhas composto de múltiplos módulos interconectados e operando de acordo com o paradigma suportado pela Teoria do Perigo em imunologia. Esse algoritmo busca atingir características significativas que um sistema de detecção de falhas deve expressar ao monitorar um sistema telefônico. Essas características seriam basicamente a adaptabilidade, devido à forte variação que esse sistema pode ter em seus parâmetros ao longo do tempo, e a diminuição no número de falsos positivos que podem ser gerados ao se classificar como falha toda anormalidade encontrada. Cenários simulados foram concebidos para validar a proposta, sendo que os resultados obtidos foram analisados e comparados com propostas alternativas.

Palavras-Chave: sistemas imunológicos artificiais, Teoria do Perigo, detecção de falhas, sistemas telefônicos.

Abstract

This thesis presents a fault detection algorithm composed of multiple interconnected modules, and operating according to the paradigm supported by the Danger Theory in immunology. This algorithm attempts to achieve significant features that a fault detection system is supposed to express when monitoring a telephone system. These features would basically be adaptability, due to the strong variation that operational conditions may exhibit over time, and the decrease in the number of false positives, which can be generated when any abnormal behavior is erroneously classified as being a fault. Simulated scenarios have been conceived to validate the proposal, and the obtained results are then analyzed and compared with alternative proposals.

Keywords: artificial immune systems, Danger Theory, fault detection, telephone systems.

CAPÍTULO 1

Introdução

Este capítulo visa introduzir o trabalho apresentado com suas motivações, o seu enfoque e a estruturação do mesmo.

1.1 Motivação

O Sistema Imunológico Humano (SIH) com sua complexidade, sua composição em diversas camadas e elementos que se comunicam, alcançando de forma tão admirável a defesa do organismo, é objeto de muitas pesquisas nas áreas biológicas. Vários modelos são buscados na tentativa de explicar os diversos fenômenos que esse sistema engloba.

De igual forma, os pesquisadores em computação vêm, atualmente, buscar no SIH inspiração para a concepção de ferramentas computacionais. Os diversos modelos matemáticos propostos, cada um com suas peculiaridades, são capazes de explicar e enfatizar propriedades do SIH que se tornam úteis na elaboração de ferramentas de processamento de informação.

Como propriedades de expressivo interesse, podem ser ressaltadas:

- **Singularidade:** o sistema imunológico de cada indivíduo é único.
- **Detecção Imperfeita e Maturação de Afinidade:** por não exigir uma identificação precisa de cada patógeno, o sistema imunológico se torna mais flexível e aumenta sua abrangência de detecção. Além disso, na presença de novos patógenos, um mecanismo de hipermutação permite incrementar o grau de afinidade na identificação.
- **Aprendizado e Memória:** os SIHs são capazes de aprender e memorizar as estruturas dos patógenos.

- **Auto-regulação:** o SIH participa do processo de homeostase do organismo, respondendo a um fluxo de matéria e energia de modo que os estados internos do organismo não extravasem seus intervalos de viabilidade (“pontos de operação"), promovendo assim a manutenção da integridade do organismo.

Fundamentadas nessas propriedades, várias ferramentas podem ser implementadas. No entanto, pelo próprio foco principal da metáfora biológica utilizada, os sistemas de defesa e detecção de falhas conduzem a analogias e aplicações mais diretas.

Esse trabalho busca, baseado na proposta apresentada por Polly Matzinger (1994, 1998, 2001, 2002), conhecida como Teoria do Perigo, elaborar um sistema de detecção de falhas na operação de uma rede telefônica.

A teoria biológica mencionada (Teoria do Perigo) apresenta uma diferença principal em relação às teorias anteriores. Ela busca a defesa do organismo baseando-se na detecção de perigo e não na detecção de elementos que não pertencem ao corpo.

Embora seja aparentemente uma simples mudança de dicotomia, entre ⟨perigo⟩ x ⟨sem perigo⟩ e ⟨pertencente ao corpo⟩ x ⟨não pertencente ao corpo⟩, de fato esse novo paradigma apresenta ganhos significativos na explicação de fenômenos não abarcados por outros modelos. Conseqüentemente, pela utilização dessa teoria biológica como inspiração na formulação de algoritmos computacionais, novas propriedades podem ser alcançadas.

Com essa visão é que se busca nesse trabalho elaborar um algoritmo que explore alguns dos principais ganhos obtidos pela utilização dessa nova interpretação biológica.

1.2 Objetivos e Contribuições

O algoritmo elaborado tem como foco a detecção de falhas em uma rede de telefonia. Embora seja essa a aplicação, o algoritmo de detecção na verdade poderia ser aplicado na detecção de qualquer anomalia em qualquer tipo de sistema, sendo necessário para tanto

uma redefinição dos atributos ou condições operacionais a serem monitorados e uma nova sintonia de parâmetros do algoritmo.

A rede de telefonia é utilizada como objeto de detecção de falhas por fornecer um ambiente dinâmico, ou seja, onde o conceito de normal ou anormal, de padrão ou fora do padrão, observando-se suas condições de operação, pode sofrer alterações ao longo do tempo. Além disso, a detecção de falhas é uma das atividades mais requisitadas para a qualidade de serviço na operação de redes de telefonia.

A abordagem aqui explorada (baseada na Teoria do Perigo) visa fornecer um tratamento adaptativo para um ambiente variável. Baseando-se ainda nos conceitos de engenharia imunológica a serem apresentados no Capítulo 3, algumas rotinas complementares foram incorporadas ao algoritmo de detecção de falhas, visando especificamente ganho de desempenho, vinculado ao número de falsos positivos, e diminuição da intervenção humana na detecção de falhas.

Os resultados alcançados por esta dissertação indicam que modelos bio-inspirados, livres dos rigores matemáticos presentes em outras metodologias de modelagem, podem alcançar resultados bastante satisfatórios na solução de problemas de reconhecimento de padrões e aprendizado de máquina. Este trabalho mostra também como são possíveis ganhos significativos na utilização de modelos híbridos, onde somente as propriedades de interesse, pertencentes a um modelo biológico, são implementadas computacionalmente, sendo essas complementadas, ou tendo suas propriedades ampliadas, por outras rotinas computacionais não necessariamente bio-inspiradas.

1.3 Estruturação do Trabalho

Este trabalho está estruturado em sete capítulos, como descrito a seguir.

Capítulo 1: Introdução

Este capítulo visa introduzir o trabalho apresentado com suas motivações, o seu enfoque e a estruturação do mesmo.

Capítulo 2: O Sistema Imunológico e a Teoria do Perigo

Este capítulo tem por objetivo apresentar conceitos básicos do sistema imunológico, enfatizando seus vários modelos, entre eles a Teoria do Perigo, modelo esse que servirá de inspiração na elaboração do sistema de detecção de falhas.

Capítulo 3: Engenharia Imunológica e Detecção de Falhas

Este capítulo tem por objetivo apresentar o conceito de Engenharia Imunológica e introduzir aspectos relevantes de um sistema de detecção de falhas. Também é proposta deste capítulo relacionar a engenharia imunológica à detecção de falhas, apresentando as principais fontes de inspiração que um sistema imunológico pode oferecer na elaboração de uma ferramenta de detecção e, por fim, apresentar alguns trabalhos relacionados ao tema detecção de falhas, que também tomam por base a Teoria do Perigo.

Capítulo 4: Chamadas Telefônicas e a Inspiração Biológica na Detecção de Falhas

Este capítulo tem por objetivo apresentar modelos de chamadas telefônicas, tanto para redes de pacotes como para redes de circuito, bem como mostrar como a inspiração na Teoria do Perigo pode se relacionar com as chamadas telefônicas de forma a se obter resultados de alta qualidade na detecção de falhas.

Capítulo 5: O Algoritmo de Detecção de Falhas Baseado na Teoria do Perigo

Este capítulo apresenta o algoritmo de detecção de falhas baseado na Teoria do Perigo de Polly Matzinger. Apresenta separadamente cada propriedade explorada no algoritmo, mostrando seu papel no contexto global. Faz também uma comparação entre o algoritmo proposto e os trabalhos relacionados.

Capítulo 6: Resultados Obtidos

Este capítulo apresenta os resultados obtidos baseando-se em simulações efetuadas com o algoritmo. Procura também contrastar as visões da Teoria do Perigo com a visão Próprio/Não-Próprio.

Capítulo 7: Conclusões e Trabalhos Futuros

Neste capítulo, baseando-se nos resultados obtidos, são levantados os aspectos positivos do uso da Teoria do Perigo em sistemas de detecção de falhas. São também propostas perspectivas futuras que visam dar continuidade à pesquisa.

CAPÍTULO 2

O Sistema Imunológico e a Teoria do Perigo

Este capítulo tem por objetivo apresentar conceitos básicos do sistema imunológico, mostrando elementos e propriedades que servirão de inspiração na elaboração do sistema de detecção de falhas. Apresentam-se também a Teoria do Perigo e outros modelos imunológicos, fazendo comparações entre eles e buscando mostrar qual a inovação desse modelo frente aos demais.

2.1 Introdução

Por muitos anos os imunologistas aceitaram o ponto de vista de que o principal objetivo do sistema imunológico era defender o corpo contra elementos externos, ou não próprios ao mesmo. Embora essa visão tenha trazido muitos ganhos à imunologia, o fato de se ter um modelo baseado na distinção entre o que vem a ser entidades próprias ou externas ao corpo limita a interpretação de uma série de fenômenos imunológicos, deixando-os, em muitos casos, sem explicação plausível. Dentre eles podemos citar questões relacionadas a tumores, transplantes e auto-imunidade.

Polly Matzinger propõe um novo paradigma em que essa visão ⟨próprio⟩ x ⟨não-próprio⟩ é refutada e a noção de perigo se torna o fundamento para a reação do sistema imune. Essa nova visão denomina-se Teoria do Perigo (Matzinger, 1994, 1998, 2001, 2002). Com base nessa teoria, as questões imunológicas supracitadas, bem como outras ainda não compreendidas plenamente, passam a ser explicáveis e compreensíveis.

Este capítulo visa apresentar os conceitos imunológicos básicos, conforme exposto por Janeway *et al.* (2002), as hipóteses sobre as quais a Teoria do Perigo de Matzinger se

sustenta, passando então a apresentar a Teoria do Perigo e terminando com uma breve comparação entre os modelos mencionados no capítulo.

2.2 Conceitos Imunológicos Básicos

A Imunologia é uma ciência relativamente nova. Sua origem é atribuída a Edward Jenner, que descobriu em 1796 que a *vacínia* ou *cowpox* induzia proteção contra a varíola humana, uma doença freqüentemente fatal. Jenner deu ao seu processo o nome de vacinação, termo que ainda hoje descreve a inoculação de indivíduos sadios com amostras enfraquecidas ou atenuadas de agentes causadores de doenças, a fim de obter proteção natural do organismo contra a enfermidade (Janeway *et al*., 2002).

Quando Jenner introduziu a vacinação, nada sabia a respeito dos agentes infecciosos que causam doenças: foi somente mais tarde, no século XIX, que Robert Koch provou que as doenças infecciosas eram causadas por microorganismos, cada um responsável por uma determinada enfermidade ou patologia. Reconhecemos atualmente quatro grandes categorias de microorganismos causadores de doença ou patógenos: os vírus, as bactérias, os fungos e os parasitas.

As descobertas de Koch e de outros grandes microbiologistas do século XIX possibilitaram o desenvolvimento da imunologia e estenderam o procedimento da vacinação de Jenner para outras doenças. Por volta de 1880, Louis Pasteur projetou uma vacina contra a cólera em galinhas e desenvolveu uma vacina anti-rábica. Tantos triunfos práticos levaram à busca dos mecanismos de proteção imune. Em 1890, Emil von Behring e Shibasaburo Kitasato descobriram que o soro dos indivíduos vacinados continha substâncias – as quais chamaram de anticorpos – que se ligavam especificamente aos agentes infecciosos.

Uma resposta imune específica, como a produção de anticorpos voltados para uma determinada atuação, é conhecida como uma resposta imune adaptativa, uma vez que é obtida durante a vida de um indivíduo como reação adaptativa à presença de patógenos

específicos em seu organismo. Isso diferencia essas respostas da imunidade inata, já conhecida na época em que von Behring e Kitasato descobriram os anticorpos, principalmente por meio dos trabalhos do grande imunologista russo Elie Metchnikoff. Metchnikoff verificou que muitos microorganismos podiam ser ingeridos e digeridos por células fagocitárias, chamadas macrófagos. Essas células estão imediatamente disponíveis para combater uma ampla variedade de patógenos, sem requerer exposição prévia, e são um componente essencial do sistema imune inato.

Na verdade, logo se tornou claro que anticorpos específicos podiam ser induzidos contra uma vasta gama de substâncias. Essas substâncias são conhecidas como antígenos (*anti* = contra; *gen* = gerar), porque podem estimular a geração de anticorpos. No entanto, o termo antígeno é usado hoje em um sentido mais amplo, descrevendo qualquer substância capaz de promover uma reação do sistema imunológico adaptativo.

2.2.1 Elementos do Sistema Imunológico

Tanto a imunidade inata como as respostas adaptativas dependem da atividade de células brancas do sangue, os leucócitos. No que se refere à imunidade inata, esta é mediada, principalmente, por granulócitos e macrófagos. Por sua vez, as respostas adaptativas dependem dos linfócitos. Em conjunto, os sistemas imunológicos inato e adaptativo contribuem para um sistema de defesa notavelmente efetivo.

As células que constituem o sistema imunológico se originam de precursores presentes na medula óssea. Todos os elementos celulares do sangue, inclusive as células vermelhas que transportam o oxigênio, as plaquetas que deflagram a coagulação sangüínea nos tecidos lesados e as células brancas do sistema imunológico, derivam das mesmas células precursoras ou progenitoras, as células-tronco hematopoiéticas da medula.

As células-tronco hematopoiéticas originam o progenitor linfóide comum, que é o precursor dos linfócitos B e T (células do sistema imune adaptativo), e o progenitor mielóide, que é o precursor dos granulócitos (neutrófilo, eosinófilo e basófilo), macrófagos, células dendríticas e mastócitos (células do sistema imune inato), como ilustrado na Tabela 2.1.

Tabela 2.1 Mediadores da imunidade inata e adaptativa e seus progenitores.

<table>
<tr><th colspan="2">Progenitor da Célula Mediadora</th><th colspan="2">Célula Mediadora</th><th>Imunidade</th></tr>
<tr><td rowspan="7">Célula-tronco hematopoiética</td><td rowspan="5">Progenitor Mielóide</td><td colspan="2">Macrófago</td><td rowspan="5">Inata</td></tr>
<tr><td colspan="2">Célula Dendrítica</td></tr>
<tr><td colspan="2">Mastócito</td></tr>
<tr><td rowspan="3">Granulócito</td><td>Neutrófilo</td></tr>
<tr><td>Eosinófilo</td></tr>
<tr><td>Basófilo</td></tr>
<tr><td rowspan="2">Progenitor Linfóide Comum</td><td colspan="2">Célula T</td><td rowspan="2">Adaptativa</td></tr>
<tr><td colspan="2">Célula B</td></tr>
</table>

Em relação ao sistema imune inato, os macrófagos, as células dendríticas e os neutrófilos possuem função fagocítica, sendo que os macrófagos e as células dendríticas maduras possuem ainda a habilidade de apresentar antígenos a outras células (Células apresentadoras de Antígenos, *Antigen Presenting Cell* – APC). Os mastócitos orquestram as respostas alérgicas, os eosinófilos parecem ser importantes principalmente na defesa contra infecções parasitárias e a função dos basófilos provavelmente é similar e complementar à dos eosinófilos e mastócitos (Tabela 2.2).

Tabela 2.2 Células Mielóides (adaptado de Janeway *et al.*, 2002).

Célula	Função Ativada
Macrófago	• Fagocitose e ativação de mecanismos bactericidas • Apresentação de antígeno
Célula Dendrítica	• Captura do antígeno nos locais periféricos • Apresentação de antígenos nos linfonodos
Neutrófilo	• Fagocitose e ativação de mecanismos bactericidas
Eosinófilo	• Morte de parasitas recobertos por anticorpos
Basófilo	• Similar e complementar à dos eosinófilos e mastócitos (função provável)
Mastócito	• Orquestra respostas alérgicas

Embora tanto os linfócitos B e T surjam na medula óssea, somente os linfócitos B ali se diferenciam; os linfócitos T migram para o timo para sofrer seu processo de amadurecimento. Portanto, os linfócitos B são assim denominados por sofrerem maturação na medula óssea (em inglês, *bone marrow*), enquanto os linfócitos T sofrem maturação no timo. Esses órgãos, medula óssea e timo, são denominados órgãos linfóides primários.

Uma vez completada sua maturação celular, os dois tipos de linfócitos entram na corrente sangüínea, migrando para os órgãos linfóides periféricos: os linfonodos, o baço e os tecidos linfóides associados às mucosas, como as amídalas, as placas de Peyer e o apêndice cecal. Os linfócitos estão em contínua recirculação entre esses tecidos, aos quais o antígeno é também encaminhado, vindo de todos os locais de infecção, primariamente dentro de macrófagos e células dendríticas. Os órgãos linfóides periféricos são os locais de ativação dos linfócitos pelos antígenos.

Os vasos do sistema linfático são os responsáveis por coletar o fluido extracelular dos tecidos, fazendo-os retornar para o sangue. Esse fluido extracelular é produzido continuamente pela filtração do sangue, e é chamado de linfa. Os vasos nos quais circula esse fluido denominam-se vasos linfáticos. Esses vasos podem ser aferentes quando drenam

a linfa, levando macrófagos e células dendríticas dos locais de infecção para os órgãos linfáticos (linfonodos), ou eferentes quando utilizados pelos linfócitos para deixar os linfonodos. A Figura 2.1 ilustra a distribuição dos tecidos linfóides no organismo humano e a Tabela 2.3 apresenta suas principais funções.

Os linfócitos B, quando ativados, diferenciam-se em células plasmáticas secretoras de anticorpos. Já os linfócitos T possuem duas classes principais. Uma dessas classes, quando ativada, diferencia-se em células T citotóxicas, que matam as células infectadas por antígenos. Já a segunda classe se diferencia em células T auxiliares (T_H) que ativam outras células, como os linfócitos e os macrófagos.

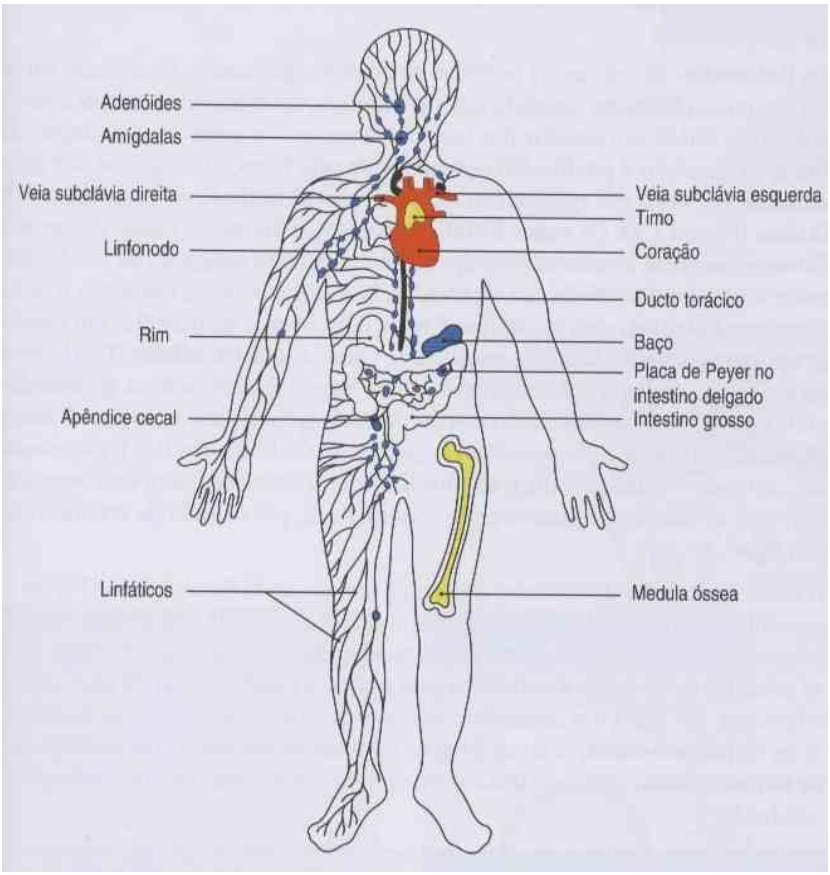


Figura 2.1 Distribuição dos tecidos linfóides no organismo humano (Janeway *et al*., 2002).

Tabela 2.3 Descrição dos Órgãos Linfóides.

Órgão Linfóide		**Função / Descrição**
Primário	Medula Óssea	Local da hematopoiese, ou seja, a geração dos elementos celulares do sangue, incluindo as hemácias, os monócitos, os leucócitos polimorfonucleares e as plaquetas. Nos mamíferos é o sítio de desenvolvimento das células B e a fonte de células-tronco que dão origem aos linfócitos T após a migração para o timo.
	Timo	Local de desenvolvimento das células T.
Secundário	Amídalas	Grandes agregados de células linfóides organizadas como parte do sistema imunológico associado a mucosas ou ao intestino.
	Adenóides	Formadas por tecido linfóide associado a mucosas e localizadas na cavidade nasal.
	Apêndice Cecal	Tecido linfóide associado ao intestino, localizado na porção inicial do cólon.
	Baço	Órgão que contém uma polpa vermelha, envolvida na remoção de células sangüíneas envelhecidas, e uma polpa branca de células linfóides, que responde aos antígenos levados ao baço pelo sangue.
	Vasos Linfáticos	Vasos de paredes finas que levam a linfa através do sistema linfático. Os vasos linfáticos podem ser aferentes quando drenam os fluidos dos tecidos (linfa), levando macrófagos e células dendríticas dos locais de infecção para os órgãos linfáticos (linfonodos), ou eferentes quando utilizados pelos linfócitos para deixar os linfonodos.
	Linfonodos	Órgão linfóide onde se iniciam as respostas imunes adaptativas. Os linfonodos encontram-se em muitas localizações, onde os vasos linfáticos se reúnem, entregando antígeno para as células apresentadoras, que os exibem aos muitos linfócitos recirculantes que migram através do linfonodo. Alguns desses linfócitos podem reconhecer o antígeno e responder contra o mesmo, iniciando uma resposta imune adaptativa.
	Placas de Peyer	Agregados de linfócitos ao longo do intestino delgado.

2.2.2 Características Básicas da Imunidade Adaptativa

Embora o sistema imunológico inato forme uma linha de defesa essencial contra muitos microorganismos comuns, nem sempre consegue eliminar os organismos infectantes, e há muitos patógenos que ele não consegue reconhecer. O mecanismo de reconhecimento usado pelos linfócitos na resposta imune adaptativa evoluiu para superar as limitações enfrentadas pelo sistema imune inato e permitir o reconhecimento de uma diversidade quase ilimitada de antígenos, de forma que patógenos arbitrários possam ser reconhecidos (Janeway *et al.*, 2002).

Tanto os linfócitos T como os linfócitos B carregam em sua superfície receptores de antígeno (receptores antigênicos) muito diferentes, capazes de reconhecer uma grande diversidade de antígenos. Em vez de carregar vários receptores, cada linfócito virgem que penetra na corrente circulatória é portador de receptores de antígeno com apenas uma única especificidade. A especificidade desses receptores, no entanto, é determinada por um único mecanismo genético especial que atua durante o desenvolvimento dos linfócitos na medula óssea e no timo, para gerar milhões de diferentes variantes dos genes codificadores das moléculas receptoras. Assim, embora um linfócito individual seja portador de um receptor de especificidade única, a especificidade de cada linfócito é diferente. Isso assegura que os milhões de linfócitos do organismo coletivamente possam dar origem a milhões de especificidades diferentes – o repertório de receptores de linfócitos de um indivíduo. Durante a vida de uma pessoa, esses linfócitos sofrem um processo de seleção natural; somente os que encontram um antígeno com o qual seu receptor pode interagir serão ativados para proliferar e se diferenciar em células efetoras. As células efetoras são os linfócitos que podem intermediar a remoção de germes patogênicos do organismo, sem a necessidade de ulterior diferenciação. Esse mecanismo seletivo foi proposto primeiramente por Macfarlane Burnet em meados de 1950. Ele propôs que, após a ligação do antígeno, a célula é ativada para proliferar e produzir uma numerosa progênie idêntica, conhecida como clone. Burnet deu assim à sua proposição o nome de teoria da seleção clonal.

A seleção clonal dos linfócitos é considerada o princípio mais importante da imunidade adaptativa. Seus postulados básicos são apresentados a seguir:

- Cada linfócito é portador de um só tipo de receptor de especificidade única;
- A interação de uma molécula estranha e um receptor de linfócitos capaz de ligar-se a essa molécula com alta afinidade leva à ativação linfocitária;
- As células efetoras diferenciadas, derivadas de um linfócito ativado, portarão receptores de especificidade idêntica à da célula parental da qual se originou o linfócito;
- Os linfócitos portadores de receptores específicos para moléculas próprias são destruídos em uma fase precoce do desenvolvimento linfóide (seleção negativa) e, assim, estão ausentes do repertório de linfócitos maduros.

A grande diversidade de receptores de linfócitos significa que pelo menos haverá uns poucos que possam se ligar a um antígeno estranho. No entanto, uma vez que cada linfócito tem um receptor diferente, o número de linfócitos que podem se ligar e responder a um determinado antígeno é muito pequeno. Para produzir linfócitos efetores específicos em número suficiente para combater uma infecção, um linfócito com o receptor de especificidade apropriada deve ser ativado e proliferar antes que sua progênie finalmente se diferencie em células efetoras. Esta expansão clonal é uma característica de todas as respostas imunes adaptativas.

Cabe ainda salientar que as células B, mas não as células T, são submetidas a um processo adicional (rearranjo das regiões variáveis das cadeias protéicas do receptor) após sua ativação. Esse processo denomina-se hipermutação somática e corresponde a uma alta taxa de mutações pontuais que gera uma diversidade adicional para o clone de células B produzido. Esse evento permite aumentar a afinidade da resposta imunológica, mas tais mutações afetam apenas células somáticas e não são herdadas pela transmissão por linha germinativa.

2.2.2.1 Seleção Negativa

Os linfócitos cujos receptores ligam-se fortemente aos antígenos próprios (células e moléculas do próprio organismo) são levados à morte (*deleção clonal*) ou a um estado de não-responsividade ao antígeno (*anergia clonal*). Isso se denomina *seleção negativa*. Os linfócitos fortemente auto-reativos são, dessa forma, removidos do repertório antes de se tornarem completamente maduros e propiciarem reações auto-imunes. Essa remoção pode ser tanto antes de emergirem dos órgãos linfóides primários (*seleção negativa central*), como durante a maturação nos órgãos linfóides periféricos (*seleção negativa periférica*).

Os linfócitos que formam a população de linfócitos maduros são, dessa forma, uma pequena fração daqueles produzidos na medula óssea ou no timo. No entanto, estas células expressam um amplo repertório de receptores capazes de responder para uma variedade praticamente ilimitada de estruturas. Este repertório fornece o material básico sobre o qual a seleção clonal atua na resposta imune adaptativa.

2.2.2.2 Memória Imunológica

Depois de ter sido ativado, um linfócito virgem leva de quatro a cinco dias para completar a expansão clonal e para se diferenciar em células efetoras. As células efetoras têm um tempo de vida limitado e, assim que o antígeno tenha sido removido, a maioria das células antígeno-específicas geradas pela expansão clonal dos pequenos linfócitos sofrerá morte celular programada (apoptose). Entretanto, algumas irão persistir após a eliminação do antígeno. Essas células são chamadas de *células de memória* e formam a base da *memória imunológica*, que garante uma resposta mais rápida e eficaz quando ocorre um segundo encontro com o patógeno.

As características da memória imunológica são facilmente observadas mediante a comparação da resposta induzida em um indivíduo após uma primeira infecção (*imunização primária*) com a resposta induzida no mesmo indivíduo por uma reinfecção (*imunização secundária*, ou de *reforço*). A *resposta secundária* de anticorpo ocorre após uma fase de latência mais curta, alcança um patamar nitidamente mais elevado e produz

anticorpos de maior afinidade ou com mais força de ligação ao antígeno, conforme se pode observar na Figura 2.2. Também na Figura 2.2 pode-se comparar a resposta secundária a um antígeno A com a resposta primária a um antígeno B, enfatizando inclusive a natureza paralela e distribuída da resposta imunológica.

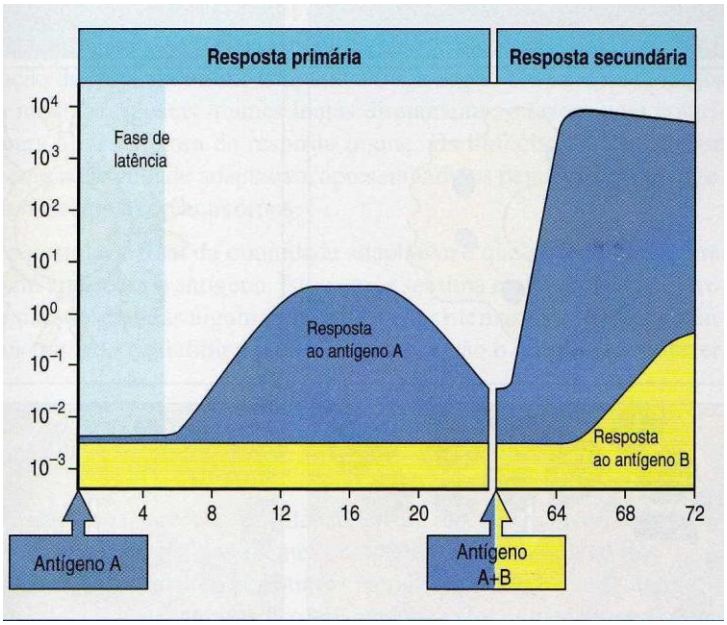


Figura 2.2 Curso de uma típica resposta de anticorpo (Janeway *et al*., 2002). Eixo x: dias; eixo y: anticorpo (μg/ml de soro).

É a memória imunológica que permite o êxito da vacinação e previne a reinfecção com patógenos que já tenham sido eliminados com sucesso por uma resposta imune adaptativa.

2.3 Modelos Embasadores

Os modelos a partir dos quais a Teoria do Perigo evoluiu se baseiam na distinção entre o próprio e o não-próprio (*Self-Nonself Discrimination* - SNSD). Embora existam muitas diferenças entre esses modelos e a Teoria do Perigo, há também pontos em comum que podem ser entendidos e comparados (Matzinger, 2001). Na seção 2.4, busca-se fazer uma breve comparação entre esses modelos e a Teoria do Perigo, ressaltando-se as principais afinidades e divergências.

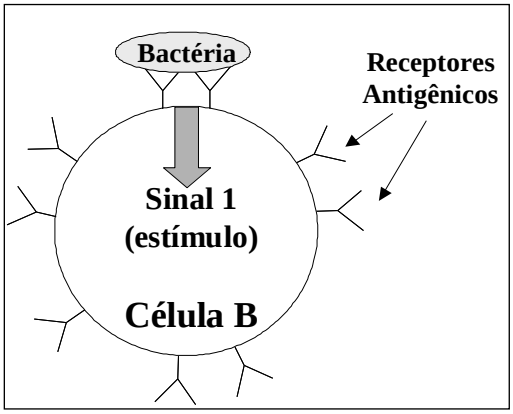
A visão SNSD foi inicialmente proposta em 1959 por Burnet em um modelo em que as células B carregavam receptores específicos para certos antígenos, sendo estes propagados por um processo de clonagem celular. O acionamento das células B se dava simplesmente pela interação desses receptores com o antígeno. O sinal gerado por essa interação é conhecido atualmente como sinal 1 (sinal de estímulo). Para garantir que a resposta imune fosse direcionada para o não-próprio, Burnet e Medawar incorporaram a sugestão de Lederberg que, no início da ontogênese, as células auto-imunes fossem deletadas (Lederberg, 1959; Billingham *et al.*, 1953; Burnet, 1959).

A seguir, em 1969, Bretscher e Cohn criaram o modelo do "Reconhecimento Associativo" (Bretscher & Cohn, 1970) em que um sinal de ajuda (sinal 2), enviado por uma célula T auxiliar (T_H), foi acrescentado para evitar que alguma célula B ativa, ao sofrer uma hipermutação, pudesse facilmente gerar uma reação auto-imune desenfreada. Eles sugeriram que o recebimento, por uma célula B, de um sinal 1 sem um sinal 2, levaria à sua deleção. Atualmente sabe-se que a célula B internaliza o antígeno fragmentando-o em peptídeos antigênicos e apresentado-o em sua superfície sob a forma de um complexo MHC/peptídeo à célula T auxiliar que, após reconhecer o complexo, envia o sinal de ajuda para as células B (Matzinger, 2001).

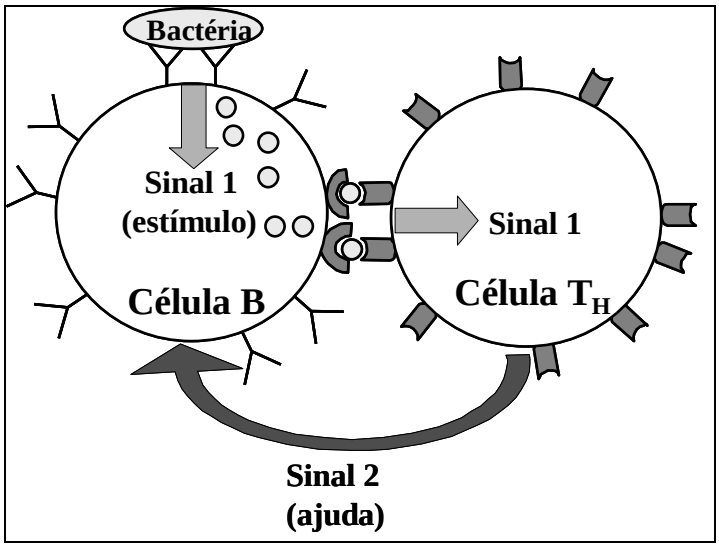
Um passo seguinte foi dado em 1974 por Lafferty e Cunningham (1975). Eles propuseram a introdução de uma nova célula (conhecida atualmente como Célula Apresentadora de Antígenos, *Antigen Presenting Cell* - APC) e um novo sinal (sinal de co-estímulo) enviado por essa célula. Somente pelo recebimento do sinal de co-estímulo, as células T auxiliares seriam ativadas (Lafferty & Cunningham, 1975). As células T auxiliares, dessa forma, não seriam tidas como constitutivamente ativas tal como no modelo anterior. Após sua ativação, as células T auxiliares se comportariam como no modelo do Reconhecimento Associativo. O modelo de Lafferty e Cunningham buscava justificar o porquê da alorreatividade (reação entre células imunes de um indivíduo e células de outro indivíduo da mesma espécie) ser mais forte que a xenorreatividade (reação entre células imunes de um indivíduo e células de outro indivíduo de uma outra espécie). Esse modelo não se casou totalmente à visão SNSD pelo fato de as APCs bem como o sinal de co-

estímulo, diferentemente das células T auxiliares e do sinal de ajuda, não estarem atrelados a um reconhecimento antigênico específico. Dessa forma, em 1989 Janeway propôs uma expansão do modelo SNSD, conhecido como modelo Não-próprio Infeccioso (*Infectious-Nonself Model* – INS) buscando esse casamento (Janeway, 1989, 1992). Nesse modelo, as APCs deveriam também ser ativadas. Essa ativação se daria por meio do reconhecimento de padrões moleculares associados a patógenos (*Pathogen-Associated Molecular Patterns* - PAMPs) na bactéria, feito pelos receptores de reconhecimento de padrões *(Pattern Recognition Receptors* - PRR). Uma vez ativadas, as APCs apresentariam os antígenos internalizados, sob a forma de um complexo MHC/peptídeo, às células T auxiliares e enviariam o sinal de co-estímulo às células T auxiliares. Seria possível a distinção entre "não-próprio infeccioso" e "próprio não-infeccioso". Assim, novamente a resposta do sistema imune se atrelava totalmente à distinção de padrões estáticos previamente concebidos.

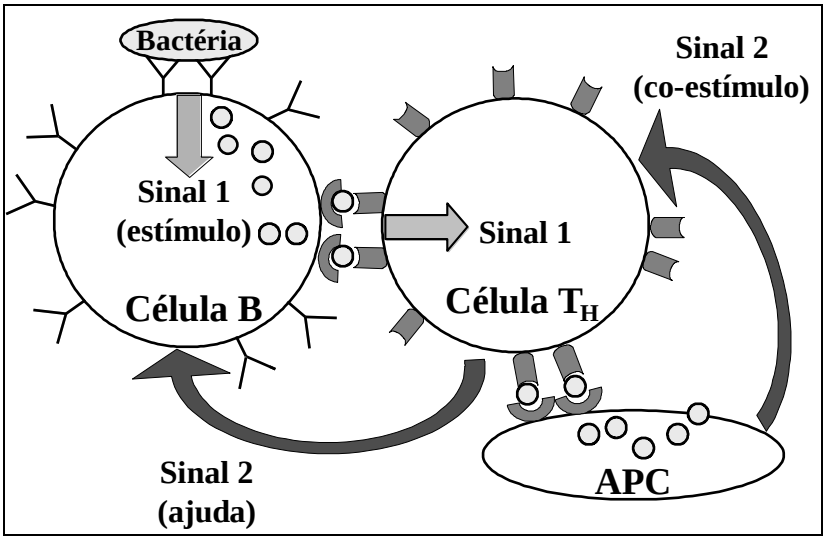
A Figura 2.3 baseia-se em ilustrações presentes em Matzinger (2001, 2002), e busca ilustrar as diferentes visões de atuação do sistema imune concebidas pelos modelos SNSD.



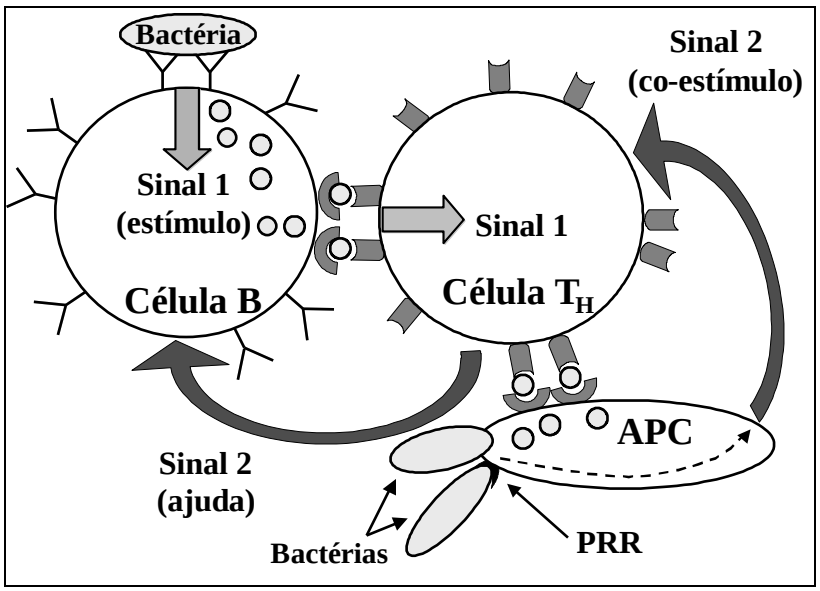
(a) Burnet, 1959. Antígeno no controle. O reconhecimento do antígeno (sinal 1) leva à ativação do linfócito



(b) Bretscher e Cohn, 1969. Modelo do Reconhecimento Associativo. A célula T auxiliar está no controle. O recebimento de um sinal 1 por uma célula B leva à sua deleção, mas a adição do sinal de ajuda (sinal 2) leva à sua ativação



(c) Lafferty e Cunningham, 1974. A Célula Apresentadora de Antígenos (APC) está no controle. As células T auxiliares morrem quando recebem um sinal 1, a não ser que sejam salvas, e ativadas, pelo sinal de co-estímulo (sinal 2) vindo de uma APC



(d) Janeway, 1989. Modelo Não-Próprio Infeccioso (INS). Os PRRs (Receptores de Reconhecimento de Padrões) estão no controle. As APCs não enviam o sinal de co-estímulo a não ser que sejam ativadas pelo reconhecimento de padrões moleculares associados a patógenos (PAMPS), via PRR

Figura 2.3 Histórico dos modelos SNSD.

2.4 A Teoria do Perigo

A Teoria do Perigo em muito se assemelha ao modelo INS de Janeway. Ambos se baseiam na idéia de um sinal de co-estímulo (sinal 2) combinado a um sinal de detecção de antígeno por um linfócito apropriado (sinal 1). Outro ponto crucial é que nos dois modelos a APC precisa ser ativada por seu ambiente, ou seja, o ambiente da APC está no controle da reação imune. No entanto, a Teoria do Perigo parte do princípio de que esse controle atrela-se a sinais endógenos e não exógenos, em oposição ao que o modelo INS e toda a visão SNSD propõem.

Para Matzinger, o que controla o disparo da reação imune são os sinais de perigo enviados pelos tecidos danificados ou submetidos a condições de estresse. Esses sinais de perigo seriam conhecidos como sinal 0 (Fuchs,1993; Matzinger, 1994, 1998; Matzinger & Fuchs, 1996). Células saudáveis não enviariam tais sinais, podendo inclusive enviar “sinais

tranqüilizadores" para as APCs locais. Células que sofrem necrose (morte de células devido a distúrbios físicos ou químicos) enviariam sinais de perigo, enquanto que células mortas por apoptose (morte programada de células) enviariam sinais para terem seus restos removidos.

Uma vez que a APC é ativada pelo sinal 0, ela se torna capaz de proceder ao estímulo do sistema imune adaptativo. A partir daí, o desencadear da resposta imune se dá pela presença ou falta dos sinais 1 e 2. Matzinger (1994) descreve a base do modelo por meio das *Leis da Linfótica*. Essas leis seriam aplicadas aos linfócitos em repouso. São elas:

- Um linfócito requer dois sinais para ser ativado. Um linfócito inativo morre sempre que recebe um sinal 1 sem um sinal 2 e é ativado sempre que recebe ambos os sinais. Um sinal 2 sem um sinal 1 é ignorado.

- Células T podem somente receber sinais 2 de APCs, e células B de células T ativas ou células de memória. Existe uma exceção à regra. Durante a fase inicial da seleção negativa, os linfócitos não são capazes de receber o sinal 2, não importa de qual fonte ele venha.

- Células T ou B ativas ignoram o sinal 2. Elas executam suas funções ao receber o sinal 1, sem ser necessária a presença do sinal 2. Depois de certo período de tempo, essas células morrem ou retornam ao estado de inatividade (se tornam células de memória).

A reação do sistema imune, segundo Matzinger (2001, 2002), frente a uma condição de perigo, pode ser observada na Figura 2.4.

Embora esse seja um modelo teórico, vários sinais de perigo já puderam ser constatados empiricamente (Gallucci & Matzinger, 2001). Por ser a essência do modelo, esse assunto é tratado a seguir com mais detalhes.

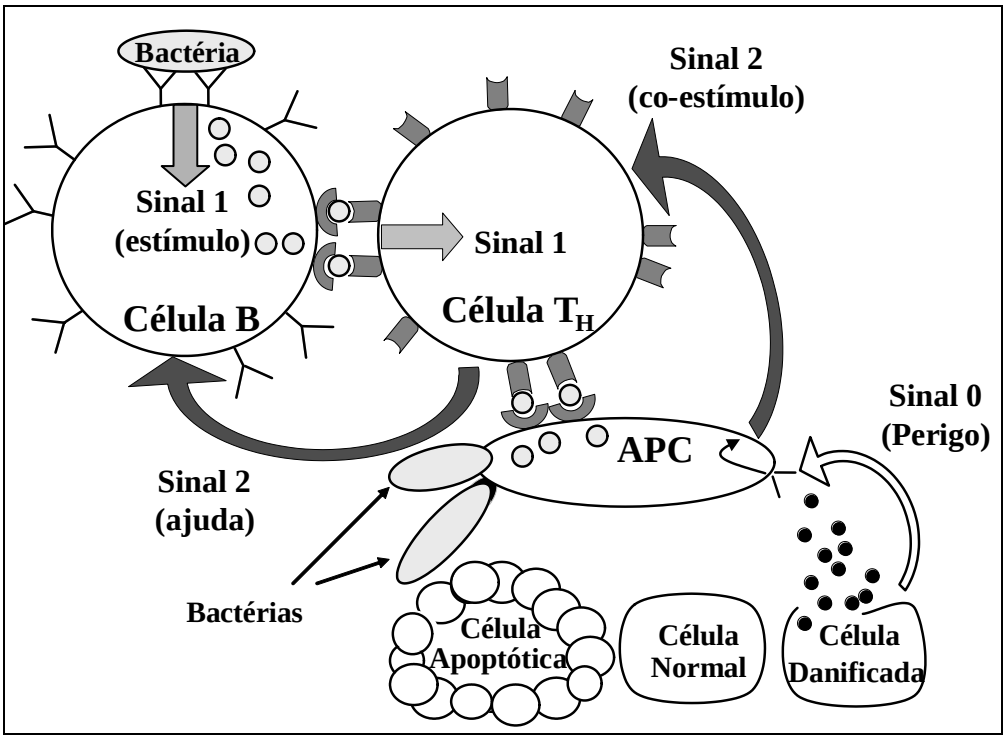


Figura 2.4 Teoria do Perigo: os tecidos estão no controle. As APCs são ativadas por sinais de perigo (sinal 0) emitidos por células danificadas ou submetidas a condições de estresse. Células saudáveis ou passando por um processo de morte apoptótica não enviam sinais de perigo.

2.4.1 Sinais de Perigo

Os sinais de perigo, enviados pelos tecidos do corpo, podem ser classificados em duas categorias: constitutivos e induzíveis (Gallucci & Matzinger, 2001). Os constitutivos são bastante importantes pelo fato de nem sempre uma célula poder ser induzida a uma sinalização antes de sua morte. Um exemplo de sinal constitutivo pode ser obtido pela própria reinterpretação da utilização de PRR sugerida por Janeway. Uma mitocôndria pode ser aceita como uma evolução de um processo de simbiose envolvendo uma bactéria primitiva. Ela não é encontrada fora de células saudáveis nem fora de células durante um processo de apoptose. Dessa forma, a detecção de uma mitocôndria por um PRR poderia ser entendida como uma possível sinalização constitutiva. Um outro exemplo de sinalização constitutiva seria a presença de manose (um monossacarídeo) em ambiente extracelular em grandes quantidades. Sabe-se que as APCs possuem receptores para manose que servem para auxiliar na percepção de antígenos. Essa substância é abundante em ambiente

intracelular. De fato, qualquer molécula intracelular presente em ambiente extracelular, e que seja detectável por APCs, pode ser tida como um sinalizador constitutivo de perigo. Sua detecção aponta para anormalidades naquele ambiente, causadas possivelmente por ruptura do tecido celular.

Seong & Matzinger (2004) apresentam uma outra possibilidade de sinalização constitutiva de perigo. Pelo fato da vida ter se desenvolvido na água, as porções hidrofóbicas das moléculas tendem a ficar em regiões protegidas, como, por exemplo, a membrana lipídica da célula. Uma vez expostas, essas porções podem servir como sinalizadoras de perigo para as APCs.

A segunda categoria de sinais, os induzíveis, são sinais emitidos por células coagidas, submetidas a um ambiente de estresse. Como exemplos, podem ser citados o Interferon-α, que é uma proteína emitida pela célula quando ocorre uma infecção viral, e proteínas de choque térmico (*Heat-Shock Protein* – HSP) que podem ser sintetizadas quando submetidas a vários tipos de estresse, entre eles o choque de temperatura.

2.5 Comparação entre os Modelos

Os modelos que foram apresentados nas seções 2.3 e 2.4 podem ser separados em três grupos fundamentais: (*i*) os modelos SNS básicos (anteriores ao modelo INS); (*ii*) o modelo INS de Janeway; e (*iii*) a Teoria do Perigo de Matzinger. Eles possuem divergências, mas também pontos em comum. Essas intersecções são a causa de tais propostas, principalmente o INS e a Teoria do Perigo, serem muitas vezes consideradas como fundamentadas em dicotomias equivalentes: ⟨próprio⟩ x ⟨não-próprio⟩ e ⟨inofensivo⟩ x ⟨perigoso⟩. Isso representaria, então, uma questão de pura semântica. Na verdade, esse raciocínio não se sustenta quando verificamos as particularidades de cada uma.

O ponto onde os três grupos se equivalem é na definição de patógenos externos perigosos e na definição de próprios que não oferecem perigo. Nessas situações, os três grupos podem se confundir. Porém, existem outras classificações como os não-próprios que não são infecciosos podendo ou não ser perigosos, não-próprios que são infecciosos mas

não são perigosos, e próprios que são perigosos. Esses são pontos de discordância entre os modelos. A Tabela 2.4 (inspirada na Figura 2 de Matzinger (2002)) mostra a comparação entre os três grupos de modelos.

Tabela 2.4 Indicação de quais circunstâncias levam os modelos SNS, INS e do Perigo a iniciarem uma resposta imune. São apresentadas duas formas de visualização: (a) enfatizando a comparação dos modelos; e (b) enfatizando o ambiente próprio para uma resposta imune. Os campos deixados em branco representam uma combinação em que nenhum dos modelos iniciaria uma resposta. Os campos preenchidos com traços representam combinações não previstas ou inexistentes. Os campos escuros ou indicando um modelo representam as combinações em que se iniciaria uma resposta imune.

(a) Ênfase nas diferentes condições de resposta imune de cada modelo.

	Próprio		**Não Próprio**			
	Sem Perigo	**Perigoso**			**Sem Perigo**	
	Não Infeccioso			**Infeccioso**		**Não Infeccioso**
SNS						
INS						
Perigo						

(b) Ênfase nos cenários que levam a uma resposta imune junto a cada modelo.

	Próprio		**Não Próprio**	
	Perigoso	**Sem Perigo**	**Perigoso**	**Sem Perigo**
Infeccioso	----------	----------	SNS / INS / Perigo	SNS / INS
Não Infeccioso	Perigo		SNS / Perigo	SNS

Embora todos os modelos defendam que haja a necessidade de separação entre grupos de antígenos durante o estágio efetor da resposta imunológica, de tal forma que o sistema imune possa eliminar ou tolerar esses grupos, cada modelo apresenta sua própria interpretação sobre o que deve ser tolerado ou eliminado. Seguindo a classificação sugerida

no início desta seção (seção 2.5), para o grupo (*i*), há apenas a necessidade de distinção entre o que é próprio e o que não é. Para o grupo (*ii*), acrescenta-se, a essa distinção, a separação entre o que é ou não infeccioso. Para a Teoria do Perigo, o grupo (*iii*), toda essa classificação é posta de lado, e assume-se a distinção entre o que é ou não perigoso.

Outro ponto de separação entre os modelos, além do que deve ser eliminado ou tolerado numa resposta imune, é o que controla ou provoca essa resposta. Para (*i*), o controle encontra-se a cargo do próprio sistema imune adaptativo (Burnet, 1959; Bretscher & Cohn, 1970) ou inato (Lafferty & Cunningham, 1975). Em (*ii*), o controle não é mais visto como uma propriedade constitutiva do sistema imune, ou seja, deve ser ativado anteriormente por um elemento não-próprio infeccioso. A Teoria do Perigo dá um passo além e propõe que o controle imunológico está, em última instância, a cargo das células comuns do corpo (Matzinger, 1998). Para tanto, o modelo postula que cada célula do corpo deve ter três funções imunológicas:

1. Iniciar respostas imunológicas: qualquer célula exposta a uma condição de estresse deve enviar um sinal de alarme para a ativação das APCs.
2. Induzir a tolerância: a Teoria do Perigo, bem como o SNS e o INS, entendem que células T ou B que recebem o sinal 1 sem um sinal de co-estímulo morrem. Nesse caso qualquer célula desempenhando suas condições fisiológicas normais pode induzir uma tolerância.
3. Definir uma classe efetora padrão de resposta: essa é uma extensão da Teoria do Perigo gerado por estudos em ”privilégios imunológicos” (Wilbanks & Streilein, 1997) e “tolerância oral” (Liu *et al.*, 1997). Embora se creia que a classe da resposta imune esteja casada ao patógeno, a Teoria do Perigo propõe que os tecidos possuem preferência sobre as classes efetoras. Por exemplo, o intestino e os olhos preferem IgA (Imunoglobulina A) como anticorpos efetivos.

Visualizando assim as propriedades de cada modelo, constatam-se suas particularidades e o fato desses modelos não estarem fundamentados em dicotomias equivalentes.

CAPÍTULO 3

Detecção de Falhas e Engenharia Imunológica

Este capítulo tem por objetivo apresentar o conceito de Engenharia Imunológica e introduzir aspectos relevantes de um sistema de detecção de falhas, com uma possível classificação e levantamento de características desejáveis. Também é proposta desse capítulo relacionar a engenharia imunológica à detecção de falhas, apresentando as principais fontes de inspiração que um sistema imunológico pode oferecer na elaboração de uma ferramenta de detecção e, por fim, apresentar alguns trabalhos relacionados ao tema detecção de falhas, que também utilizam como base a Teoria do Perigo.

3.1 Introdução

Qualquer sistema de engenharia deve evitar estados anormais que comprometam de alguma forma o seu pleno funcionamento. Esses estados de anormalidade, mesmo ocorrendo, podem ter seus efeitos minimizados caso sua detecção e diagnóstico se dêem dentro de um período de tempo reduzido. Com esse intuito, vários estudos ao longo das três últimas décadas foram feitos na área de automação em detecção e diagnóstico de falhas (FDD – *Automated Fault Detection and Diagnosis*). Nesta área de pesquisa, pode-se citar Willsky (1976) como o primeiro trabalho de maior relevância realizado. Mais recentemente, podem ser citados outros trabalhos como Venkatasubramanian *et al*. (2003a, 2003b, 2003c) e Kaipamula & Brambley (2005), que são publicações de referência para o tema.

Vários algoritmos imuno-inspirados já foram propostos para detecção de falhas e anomalias, incluindo segurança computacional e de rede. Isso se deve à iniciativa de aplicação direta da metáfora imunológica, onde as ferramentas produzidas exercem o papel

fundamental do SIH: a defesa de todo o sistema. Os trabalhos imuno-inspirados desenvolvidos nos últimos anos abordam em sua grande maioria a visão ⟨próprio⟩ x ⟨não-próprio⟩ para sua modelagem. Essa visão refere-se ao conceito de próprio, do reconhecimento do que é normal ou pertencente ao sistema, a fim de detectar o oposto, o não-próprio, ou a anomalia. Esse modelo pode alcançar níveis de desempenho elevados quando aplicado a cenários estáticos (González & Dasgupta, 2002). Contudo, em cenários dinâmicos, como o abordado nesse trabalho, seu desempenho pode não ser satisfatório. Isso se deve ao fato de possuir deficiências com relação à generalização ou adaptabilidade. Essas deficiências levam principalmente a um aumento no número de falsos positivos (FPs) durante o processo de detecção. A busca da maximização da capacidade de generalização durante o processo de treinamento inicial do sistema de detecção, ou a adaptabilidade ao longo do processo de detecção, diminuiria o número de FPs em sistemas que sofram modificações ao longo do processo.

A Teoria do Perigo abre a possibilidade de acrescentar aos algoritmos de detecção de falha mencionados características especialmente voltadas à adaptabilidade. Isso pelo fato do *sinal de perigo* auxiliar o algoritmo no estabelecimento de condições para manter ativo o processo de treinamento. Esse sinal ajuda a decidir se um novo dado coletado, que se encontra fora dos padrões de normalidade, deve ser visto como uma falha ou como um possível início na mudança do ponto de operação (comportamento de normalidade), sem a necessidade de intervenção humana.

Nas seções seguintes, serão apresentadas características de Sistemas de Detecção e Diagnóstico de Falhas, conceitos referentes aos Sistemas Imuno-Inspirados e como os Sistemas Imuno-Inspirados podem ser utilizados na detecção de falhas. Ao final, serão apresentados trabalhos recentes que utilizam a Teoria do Perigo, destacando suas principais propriedades e comparando-os com o trabalho proposto.

3.2 Detecção de Falhas

Um processo de FDD é descrito comumente como possuindo três passos: a detecção da falha, o isolamento da falha e a identificação da falha (Kaipamula & Brambley, 2005). O primeiro passo consiste na indicação de alguma anormalidade ocorrida no sistema. O isolamento da falha equivale à distinção entre tipos diferentes de falhas. Esse passo inclui ainda a determinação do local e do tempo da ocorrência da falha. O terceiro passo, identificação da falha, consiste basicamente no reconhecimento de atributos da falha, como o seu grau e sua evolução temporal. Esses dois últimos passos compõem o que se denomina *diagnóstico da falha*.

Esse trabalho foca o primeiro passo do processo de FDD, ou seja, *detecção de falhas*, embora possua características complementares que se atribuam especificamente ao *diagnóstico da falha*.

3.2.1 Características de um Sistema de Detecção de Falhas

Um sistema de detecção de falhas possui várias características importantes que devem ser consideradas em sua elaboração. Nem todas as características aqui abordadas são encontradas em todas as implementações e, dependendo da implementação, algumas características podem ser privilegiadas em detrimento de outras. Essas características também não fecham o espaço de propriedades que um sistema de detecção de falhas deve possuir. De qualquer forma elas são importantes para se verificar o desempenho, a confiabilidade, a generalidade e outras qualidades desejáveis para um sistema de detecção.

Embora as características a seguir se apliquem também a processos de diagnóstico, elas serão apresentadas como constituintes de processos de detecção de falhas. São elas:

Detecção Rápida de Falhas: o sistema de detecção deve responder rapidamente às ocorrências de anormalidade. Deve ser observado, contudo, que um sistema que é projetado para detectar falhas, particularmente mudanças abruptas, torna-se sensível a ruídos, gerando com freqüência alarmes falsos durante a operação normal (Venkatasubramanian *et al.*, 2003a).

Robustez: é desejável que um sistema de detecção de falhas seja robusto a vários tipos de ruídos. Mas de igual forma é desejável que o seu desempenho (velocidade de detecção de falhas) degrade aos poucos ao invés de cair abruptamente. Nota-se aí que deve haver um compromisso entre robustez e desempenho.

Adaptabilidade: processos podem sofrer mudanças de várias naturezas. Entre elas, podem ser citadas alterações nas entradas externas acarretadas por distúrbios ou mudanças das condições operacionais, e modificações estruturais devido a reajustes. É desejável, dessa forma, que um sistema de detecção de falhas seja adaptável a vários graus de mudança.

Características secundárias podem também ser exploradas com o intuito de aumentar a confiança do usuário no desempenho do sistema de detecção, como, por exemplo, informações sobre o processo de detecção e estimativas de erro.

Existem outras características que se aplicam exclusivamente aos passos relativos ao diagnóstico. Essas seriam basicamente:

Isolabilidade: capacidade de distinguir entre diferentes falhas.

Capacidade de Identificar Novidades: capacidade de não classificar um novo tipo de falha como uma operação normal ou como uma falha já conhecida.

Facilidade de Explicação: capacidade de explicar como a falha se originou e se propagou.

Estimação do erro de classificação: um importante requisito prático para um sistema de diagnóstico está em construir a confiança do usuário em sua fidedignidade. Isso pode ser facilitado se o sistema de diagnóstico prover uma estimativa do erro de classificação que pode ocorrer.

Devem ser considerados também os requisitos computacionais. Para que um algoritmo seja utilizável na prática, seus requisitos computacionais, como capacidade de processamento e memória necessários, devem estar dentro de uma margem aplicável.

3.2.2 Classificação de Algoritmos de Detecção de Falha

Várias abordagens podem ser utilizadas para se detectar falhas em um sistema. A principal diferença entre essas abordagens está no tipo de conhecimento utilizado pelo algoritmo. Considerando esse tipo de conhecimento necessário, ou seja, o discernimento entre o que é falha e o que não é, deve-se ter relacionados os sintomas observáveis em situações de falha ou normalidade do sistema. Um sistema de detecção de falhas pode ter esse conhecimento de forma explícita (por meio de uma tabela) ou pode inferi-lo por meio de uma fonte de conhecimento do domínio. Esse conhecimento do domínio pode ser desenvolvido por meio de um modelo do processo (algoritmos baseados em modelos quantitativos ou qualitativos), tendo como base o conhecimento de seus princípios básicos, ou por meio de experiências passadas com o processo (algoritmos baseados na história do processo).

A seguir, são mencionadas as principais características desses tipos de abordagens.

Métodos Baseados em Modelos Quantitativos

Os modelos quantitativos são conjuntos de relações matemáticas quantitativas baseadas na física ou nos aspectos fenomenológicos do processo (Venkatasubramanian *et al.*, 2003a; Katipamula & Brambley, 2005).

Dentre as vantagens dos modelos quantitativos, pode-se citar:

- são obtidos diretamente das variáveis que descrevem a dinâmica do processo, também denominadas variáveis de estado, facilitando assim a detecção de anomalias pelo monitoramento dessas variáveis.

Dentre as desvantagens:

- há a possibilidade de que o relacionamento entre as variáveis de estado seja complexo, exigindo um esforço significativo de modelagem;
- podem existir muitas variáveis de estado, sendo algumas de difícil mensuração;
- a atualização das variáveis de estado pode envolver erros de cálculo numérico que se refletem no resultado final.

Métodos Baseados em Modelos Qualitativos

Os modelos qualitativos são aqueles baseados em atributos nominais e em relações lógicas entre eles. As abordagens baseadas nesses modelos incluem sistemas baseados em regras e modelos baseados no conhecimento qualitativo da física ou dos aspectos fenomenológicos do processo. As regras referem-se a decisões intermediárias (antes de se chegar a uma detecção de falha) tomadas com base em informações qualitativas do processo, visando alcançar um veredicto de falha ou normalidade. Os modelos baseados no conhecimento qualitativo da física ou dos aspectos fenomenológicos do processo envolvem o conhecimento impreciso dessas questões, como por exemplo a utilização apenas da ordem de grandeza das variáveis de estado e a utilização de equações que não descrevem com precisão o relacionamento entre as variáveis utilizadas (Venkatasubramanian *et al.*, 2003b; Katipamula & Brambley, 2005). Dentre as vantagens dos modelos qualitativos, pode-se citar:

- podem chegar a conclusões sobre processos sobre os quais não se têm informações precisas;
- são simples no desenvolvimento e aplicação.

Dentre as desvantagens:

- embora esse método seja fácil de ser desenvolvido, é difícil garantir que todas as regras sejam sempre aplicáveis e encontrar um conjunto completo e consistente de regras, especialmente quando o sistema é complexo;
- à medida que novas regras são adicionadas para estender as regras existentes ou ajustar circunstâncias especiais, a simplicidade é perdida.

Métodos Baseados na História do Processo

Em contraste com as abordagens baseadas em modelos, onde o conhecimento *a priori* (quantitativo ou qualitativo) sobre o processo é necessário, os *métodos baseados na história do processo* baseiam-se somente em uma ampla quantidade de dados históricos do processo

para obter as informações necessárias (Venkatasubramanian *et al.*, 2003c; Kaipamula & Brambley, 2005).

O principal objetivo dos métodos baseados na história do processo é, de posse dos dados de entrada e saída, relacioná-los e transformá-los em conhecimento suficiente para a detecção, isolamento e identificação de falhas. Como exemplo de métodos baseados na história do processo, podem ser citados métodos que utilizam classificadores estatísticos, redes neurais artificiais, sistemas imunológicos artificiais, bem como outros algoritmos de reconhecimento de padrões. Dentre as vantagens dos métodos baseados na história do processo, pode-se citar:

- são adequados a problemas para os quais não se têm informações suficientes para modelá-los quantitativamente ou qualitativamente;
- são adequados a processos cujos dados são abundantes e acessíveis.

Dentre as desvantagens:

- é necessária uma grande quantidade de dados para criar o discernimento entre normalidade e falha;
- possuem limitações na capacidade de generalização, ou seja, na capacidade de discernir entre normalidade e falha para dados que se encontram fora da região amostrada e utilizada para treinar o algoritmo;
- podem possuir limitações de adaptabilidade caso não implementem um procedimento de treinamento permanente.

3.3 Sistemas Imunológicos Artificiais e Engenharia Imunológica

A partir de 1996, um paradigma denominado Sistemas Imunológicos Artificiais (SIAs), inspirado no Sistema Imunológico Humano, começou a ser consolidado. Naquele ano, realizou-se no Japão um workshop sobre sistemas baseados em imunologia (IBMS –

Immunity-Based Systems), onde se buscou definir essa nova linha de pesquisa e integrar os estudos anteriormente realizados.

A partir de então, vários trabalhos foram elaborados propondo definições e formulações genéricas para os Sistemas Imunológicos Artificiais. Dentre as definições, podem ser citadas:

> "Os Sistemas Imunológicos Artificiais são metodologias de manipulação de dados, classificação, representação e raciocínio que seguem um paradigma biológico plausível: o sistema imunológico humano" (WWW Starlab).
>
> "Um Sistema Imunológico Artificial é um sistema computacional baseado em metáforas do sistema natural" (Timmis, 2000).
>
> "Os Sistemas Imunológicos Artificiais são compostos por metodologias inteligentes, inspiradas no sistema imunológico biológico, para a solução de problemas do mundo real" (Dasgupta, 1998).
>
> "Sistemas Imunológicos Artificiais (SIA) são sistemas adaptativos inspirados na imunologia teórica e em funções, princípios e modelos imunes observados, os quais são aplicados na solução de problemas" (de Castro & Timmis, 2002).

A busca por um modelo imunológico artificial genérico, como em Hunt & Cooke (1996) e Hofmeyr & Forrest (1999; 2000), é algo de grande importância, pois propicia um melhor entendimento da metáfora biológica e até mesmo a utilização do modelo diretamente na solução de problemas reais. Contudo, a necessidade da obtenção final desse modelo, ainda não alcançado plenamente, não deve restringir o uso da inspiração biológica na construção de ferramentas de engenharia.

Com esse pensamento, objetivando a criação de uma linha de pesquisa derivada dos SIAs e que procura a construção de ferramentas computacionais visando a solução de problemas complexos, de Castro (2001) propõe a terminologia Engenharia Imunológica (EI):

> "A Engenharia Imunológica é um processo de meta-síntese, o qual vai definir a ferramenta de solução de um determinado problema baseado nas características do próprio problema, e depois aplicá-la na obtenção da solução. Ao invés de buscar a reconstrução parcial ou total do sistema imunológico tão fielmente quanto possível, a engenharia imunológica deve procurar desenvolver e implementar modelos pragmáticos inspirados no sistema imunológico que preservem algumas de suas propriedades essenciais e que se mostrem passíveis de implementação computacional e eficazes no desenvolvimento de ferramentas de engenharia."

Com essa visão, a inspiração biológica focada no SIH permite a elaboração de algoritmos computacionais em que somente as propriedades de interesse são exploradas, podendo inclusive ser combinadas a outras rotinas quaisquer, gerando soluções híbridas mais apropriadas ao contexto.

Dentre as aplicações já vislumbradas na literatura para os SIAs (Dasgupta & Balachandran, 2006), podem ser citadas:

- Reconhecimento de padrões;
- Aproximação de funções;
- Otimização de processos;
- Análise de dados e clusterização;
- Aprendizagem de máquina;
- Memórias associativas;
- Geração e manutenção de diversidade;
- Programação e computação evolutiva;
- Detecção de falhas e anomalias;

- Controle e *scheduling*;
- Segurança computacional e de rede;
- Geração de comportamentos emergentes.

3.4 Engenharia Imunológica na Detecção de Falhas

Como mencionado anteriormente, dentre as principais áreas de interesse de aplicação dos trabalhos imuno-inspirados, podem ser destacadas: detecção de falhas e anomalias, e segurança computacional e de rede. Isso pela utilização da principal metáfora imunológica que é a defesa do sistema. Para a elaboração de ferramentas de detecção de falhas imuno-inspiradas para sistemas tecnologicamente complexos, devem ser buscadas características apropriadas do SIH de tal forma que o mapeamento metafórico produza o resultado esperado. Para um melhor entendimento, algumas características que podem ser destacadas como de interesse e a descrição dessas sob a visão do SIH e do sistema de detecção de falhas são citadas a seguir (Banchereau *et al.*, 2000; Costa Branco *et al.*, 2003; Dasgupta & Forrest, 1996; Doherty & Christensen, 2000; Dutton *et al.*, 1998; Freitas & Rocha, 2000; Janeway *et al.*, 2002; Timmis *et al.*, 2000; Vargas, 2005; Zinkernagel, 2000).

Singularidade: o sistema imunológico de cada indivíduo é único, apesar das semelhanças que porventura existam entre eles. Cada sistema tecnológico, cada parte de um mesmo sistema tecnológico, ou mesmo sistemas tecnológicos iguais atuando em ambientes diferentes, possuem suas singularidades. Como exemplo, duas redes de computadores, que possuam uma mesma configuração, podem interpretar respectivamente uma mesma seqüência de eventos como uma falha e como uma normalidade. Isso devido ao histórico passado de eventos próprios que cada uma dessas redes possui.

Detecção Imperfeita e Mutação: por não exigir uma identificação absolutamente precisa de cada patógeno, o sistema imunológico se torna mais flexível e aumenta sua abrangência de detecção. Contudo, quando um patógeno é detectado, um mecanismo de hipermutação aguça a identificação. Inspirando-se nessa característica, pode ser feita uma busca inicial, mais rústica, para identificar as condições de falhas, e por meio de hipermutações e geração

de novos detectores fazer uma aproximação mais precisa do perfil das falhas. Esse processo permite uma redução no número de detectores.

Aprendizado e Memória: os SIHs são capazes de aprender e memorizar as estruturas dos patógenos. A utilização dessas características em um sistema de detecção de falhas é importante, especialmente para melhorar as características de d*etecção rápida de falhas* quando uma situação de falha já ocorrida volta a acontecer e *adaptabilidade* diante de novas situações de falha.

Auto-regulação: o SIH participa do processo de homeostase do organismo, ou seja, procura o estado fisiológico normal do mesmo. No caso direto dos linfócitos, homeostase refere-se àquele indivíduo não-infectado que possui uma contagem normal dessas células. Em um sistema computacional, como o detector de falhas aqui apresentado, essa propriedade consiste na habilidade do sistema de se adaptar a um ambiente dinâmico. Tanto o estado interno quanto o externo, definidos por variáveis sistêmicas, podem regular e controlar o sistema em questão com o intuito do mesmo encontrar seu ponto de operação.

3.5 Trabalhos Relacionados à Detecção de Falhas Utilizando a Teoria do Perigo

A maioria dos trabalhos de detecção de falhas, utilizando SIAs, realizados até o momento, aborda a visão ⟨próprio⟩ x ⟨não-próprio⟩. Essa visão remete ao conceito do que é próprio, ou pertencente ao sistema, a fim de detectar o oposto, ou o que não é próprio, ou seja, a anomalia procurada. Como exemplo de sistemas sob essa perspectiva, podem ser citados Anchor *et al.* (2002), Balthrop *et al.* (2002a, b), Boudec & Sarafijanovic (2003), Canham & Tyrrell (2002), Dasgupta *et al.* (2004), Dasgupta & González (2002), González & Dasgupta (2003), Hofmeyer & Forrest (1999), Kephart (1994), Sarafijanovic & Boudec (2003), Somayaji *et al.* (1996). Esses exemplos referem-se também a sistemas de detecção de intrusão, que possuem um comportamento semelhante aos de detecção de falha. Aickelin *et al.* (2004) apresentam um estudo comparativo entre os principais trabalhos sob a visão ⟨próprio⟩ x ⟨não-próprio⟩ implementados até aquela data.

Os algoritmos sob a visão ⟨próprio⟩ x ⟨não-próprio⟩ tendem a possuir uma estrutura estática, produzida a partir de um treinamento de dados inicial, e não atualizada durante o processo de detecção.

Só mais recentemente os algoritmos de detecção baseados na Teoria do Perigo começaram a ser estudados. Aickelin *et al.* (2003) propõem em seu trabalho a utilização da Teoria do Perigo na detecção de intrusões. Enfatiza a idéia de que a correlação entre vários sinais fornece um método de embasamento para a resposta imunológica. Esse trabalho propõe a identificação de alertas do tipo apoptótico que corresponderiam a ações legítimas do sistema, a pré-requisitos de um ataque e a alertas do tipo necrótico que corresponderiam aos danos reais gerados por ataques com sucesso (ver analogia biológica na seção 2.4). As ações tomadas pelo sistema de detecção de intrusão se baseiam no equilíbrio entre esses dois tipos de alertas.

Sarafijanovic & Boudec (2004) propõem a utilização de um sistema de detecção de mau funcionamento em redes móveis *ad hoc* com a visão da Teoria do Perigo. De forma simplificada, essa proposta se baseia na utilização de um timo virtual onde antígenos que possuem autorização para participarem do mesmo (antígenos não relacionados a sinais de perigo) por meio de um processo de seleção negativa permanente, geram detectores para o sistema. Esses detectores co-estimulados por sinais de perigo identificam o mau funcionamento na rede.

Em Bentley *et al.* (2005) é apresentado o paradigma do Tecido que se baseia também na Teoria do Perigo. Nesse trabalho, a detecção de falhas se dá por meio de duas camadas. Uma camada seria o SIA e uma camada intermediária entre o problema e o SIA seria o tecido. O tecido é montado dinamicamente tendo como base os dados de entrada. Ele procura espelhar a situação de normalidade. A informação passada pelo tecido não é refinada, é uma pré-informação. Quando uma anomalia é detectada pelo tecido, esse passa para a camada SIA, para essa fazer o refinamento da detecção. Como na Teoria do Perigo, o controle da resposta imunológica encontra-se a cargo dos tecidos (ver Figura 2.4).

Um algoritmo aplicável na detecção de anomalias inspirado na Teoria do Perigo e no funcionamento das células dendríticas é apresentado em Greensmith *et al.* (2005). Os parâmetros dos dados de entrada são associados a sinais do tipo PAMPs (*Pathogen-Associated Molecular Patterns*), Sinais de Perigo e Sinais de Segurança. Baseado na combinação dos sinais de entrada, há a possibilidade de serem geradas células dendríticas maduras ou semimaduras. As células maduras possuem um efeito estimulador, enquanto as semimaduras um efeito supressor (em relação à deflagração de uma resposta imunológica). Essas duas funções (estimuladora e supressora) permitem a geração de uma saída classificatória dos dados de entrada, sendo possível assim a utilização desse algoritmo na distinção entre situações de falha ou normalidade. Também em Kim *et al.* (2005) é apresentado um trabalho de defesa computacional baseado na Teoria do Perigo. Ele toma como fundamento os dois algoritmos apresentados anteriormente em Bentley *et al.* (2005) e Greensmith *et al.* (2005).

Essencialmente, o que se pode destacar nos algoritmos baseados na Teoria do Perigo frente aos baseados na visão ⟨próprio⟩ x ⟨não-próprio⟩ é a possibilidade de se trabalhar com sistemas dinâmicos de forma robusta, ou seja, sem um incremento significativo no número de falsos positivos frente às variações no comportamento do sistema e sem a necessidade da intervenção humana constante.

No capítulo 5, após a apresentação do algoritmo de detecção de falhas proposto pelo presente trabalho, é feita uma comparação desse com os acima mencionados que fazem uso da Teoria do Perigo como embasamento teórico.

CAPÍTULO 4

Chamadas Telefônicas e a Inspiração Biológica na Detecção de Falhas

Este capítulo tem por objetivo apresentar modelos de chamadas telefônicas, tanto para redes de pacotes como para redes de circuito. Apresenta as principais mensagens e parâmetros utilizados em chamadas comuns, bem sucedidas, enfatizando os parâmetros que podem ser utilizados por um algoritmo de detecção de falhas. Por fim, busca mostrar como a inspiração na Teoria do Perigo pode se relacionar com as chamadas telefônicas de forma a se obter resultados de alta qualidade na detecção de falhas.

4.1 Introdução

Os sistemas de telefonia podem ser classificados como tendo uma arquitetura baseada em redes de circuito, em redes de pacotes ou em redes híbridas. As redes de pacotes utilizam protocolos como o RTP (*Real-Time Transport Protocol*) e o RTCP (*Real-Time Transport Control Protocol*) que viabilizam a comunicação via voz e buscam manter a qualidade do serviço telefônico em tempo real. Há ainda a necessidade de outros protocolos, como, por exemplo, o SIP (*Session Initiation Protocol*), que estabelece uma sessão em que a comunicação de voz mencionada se dará.

Tratando-se das redes de circuito, embora não haja a necessidade de um protocolo para a transmissão da voz, como ocorre nas redes de pacotes, faz-se necessária a utilização de um protocolo que viabilize o estabelecimento da chamada telefônica. O que ocorre é que essas redes tradicionais (redes telefônicas de circuito) possuem uma rede de sinalização acoplada a elas. Essa rede de sinalização utiliza protocolos específicos, como o protocolo

SS7 (*Signaling System Number 7*), que servem tanto para o estabelecimento de chamadas como para o oferecimento de serviços não orientados à conexão.

Os protocolos adotados para o estabelecimento de uma chamada (ou sessão) telefônica utilizam informações comuns que são bastante relevantes no controle dessas comunicações entre usuários. Com base nisso, é possível se ter um sistema de gerenciamento que analise essas informações e localize falhas ocorridas na rede. De fato, essas análises existem em toda rede telefônica e geralmente são feitas com base na contagem de eventos indesejáveis.

Deve-se observar, contudo, que nem todo tipo de falha é facilmente identificado. Dependendo da seqüência de eventos observados, algumas mudanças no comportamento da rede podem ser tidas como falhas. Há ainda a possibilidade de se considerar chamadas normais como pertencentes a um conjunto de chamadas que geram determinada falha. Um outro ponto é que nem todo sistema de gerenciamento verifica as tendências a falha, o que pode ser bem interessante para manter o funcionamento da rede. Esse trabalho se propõe a acrescentar as propriedades supracitadas a um sistema de detecção de falhas em uma rede telefônica.

Este capítulo busca introduzir o conceito de uma chamada telefônica. Para tanto, introduz as principais arquiteturas e protocolos utilizados. Mostra também a troca de informações necessárias para o estabelecimento das chamadas, ressaltando a possibilidade de detecção de falhas pela observação dessas informações. Por fim segue a explanação da inspiração biológica, no contexto telefônico, para a detecção de falhas utilizando a Teoria do Perigo.

4.2 Chamadas Telefônicas

Dependendo da arquitetura ou dos protocolos utilizados, as chamadas telefônicas podem apresentar peculiaridades. Embora possam existir outras arquiteturas e outros eventos além dos abordados nesta seção, pretende-se mostrar aqui, de forma sintética, passos que podem

ocorrer ao longo de chamadas telefônicas comuns, tanto em uma rede de circuitos como em uma rede de pacotes.

4.2.1 Arquitetura de Redes Telefônicas: Redes Telefônicas Comutadas por Circuito e por Pacotes

Uma das arquiteturas possíveis de redes telefônicas utilizada atualmente é a rede comutada por circuito. Essa rede é composta por centrais telefônicas interconectadas fisicamente em estrutura predominantemente hierárquica, possuindo no entanto conexões fora da hierarquia principal e privilegiando assim conexões diretas entre pontos da rede. Os usuários encontram-se geralmente conectados às centrais de mais baixa hierarquia (Davidson & Peters, 2000). A Figura 4.1 ilustra uma estrutura como a citada anteriormente.

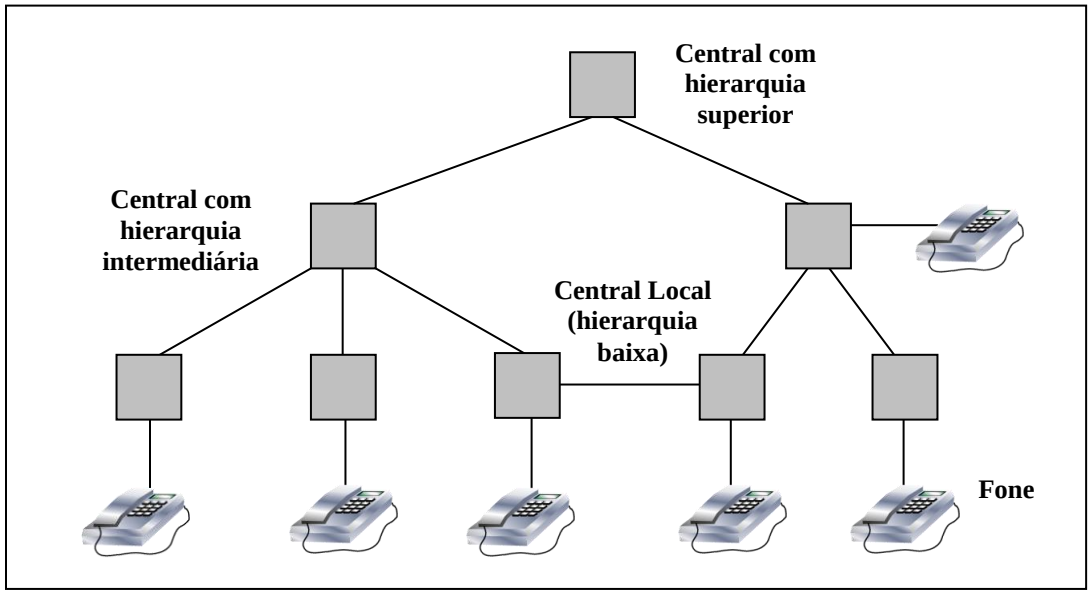


Figura 4.1 Traços gerais de uma arquitetura tipicamente utilizada por uma rede telefônica comutada por circuitos.

O estabelecimento de uma chamada entre dois usuários em uma rede telefônica de circuito implica no estabelecimento de um caminho dedicado entre os usuários. Para tanto, há a utilização de protocolos de sinalização específicos entre o usuário e a central telefônica a que se encontra conectado, e entre diversos elementos da rede. Essas sinalizações podem ser tanto analógicas como digitais. Tomando-se especificamente a sinalização digital intra-

rede (excluindo-se a sinalização entre o usuário e a rede), uma arquitetura possível é a utilização de um canal de sinalização para transmitir a sinalização de vários canais de voz (sinalização por canal comum). Nesse último caso, a rede de sinalização pode ficar dissociada logicamente da rede de voz. A Figura 4.2 mostra uma arquitetura onde os circuitos de voz encontram-se dissociados logicamente dos enlaces de sinalização. A sinalização SS7 permite tal arquitetura.

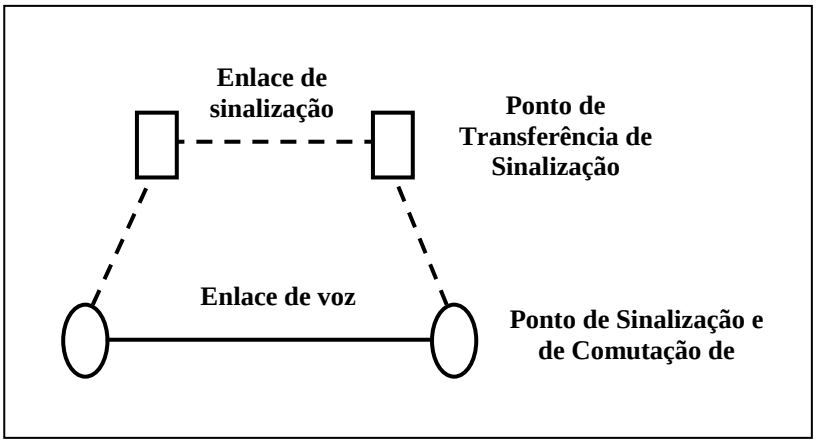


Figura 4.2 Sinalização não-associada ao circuito de voz.

Na rede telefônica, cada nó possui um endereçamento. Para a sinalização SS7, esses nós podem ser tanto centrais telefônicas como simplesmente pontos de transferência de sinalização. Ao se fazer uma chamada, inicialmente o endereço do destino é fornecido à central de origem. Essa central troca mensagens com o próximo elemento da rede de acordo com sua tabela de roteamento. Cada elemento, consultando também sua própria tabela de roteamento, procura sinalizar para o próximo elemento, buscando sempre o fechamento de um circuito adequado para o estabelecimento da chamada. Na seção 4.2.3, são apresentadas mensagens SS7 típicas em uma chamada telefônica.

Diferentemente das redes de circuito, para a realização de uma chamada em uma rede telefônica de pacotes não há a necessidade do estabelecimento de um caminho dedicado entre os usuários. A voz digitalizada é particionada em pacotes que são endereçados ao usuário final. Esses pacotes são transmitidos pela rede percorrendo rotas diversas e sendo remontados no destino. Essas rotas são escolhidas pelos nós da rede de

acordo com a disponibilidade de recursos e estratégias de roteamento adequados no momento da transmissão do pacote.

As redes IP são consideradas atraentes para esse tipo de aplicação pelo fato de transportarem os dados da aplicação fim-a-fim sem nenhum interesse real no *payload*. A rede utiliza protocolos específicos para o estabelecimento de sessões e, após o estabelecimento, para a transmissão dos pacotes de voz. Na seção 4.2.2, são apresentadas sinalizações típicas em uma sessão de voz sobre IP.

Atualmente, as redes telefônicas de pacotes e de circuitos possuem conexões formando redes híbridas, viabilizando assim o acesso entre os diversos tipos de terminais. A Figura 4.3 apresenta uma interconexão possível entre redes de circuitos e de pacotes. Nesse exemplo são mostrados os terminais comuns de usuários (fone) e os terminais SIP (fone SIP). Os fones podem ser conectados diretamente a uma rede de circuitos ou por meio de um *gateway* residêncial a uma rede IP. Na Figura 4.3 são apresentados outros elementos possíveis de serem utilizados, como os *gateways* (GW), que viabilizam a interconexão das diferentes redes, e os *softswitches*, que exercem funcionalidades de controle de chamadas.

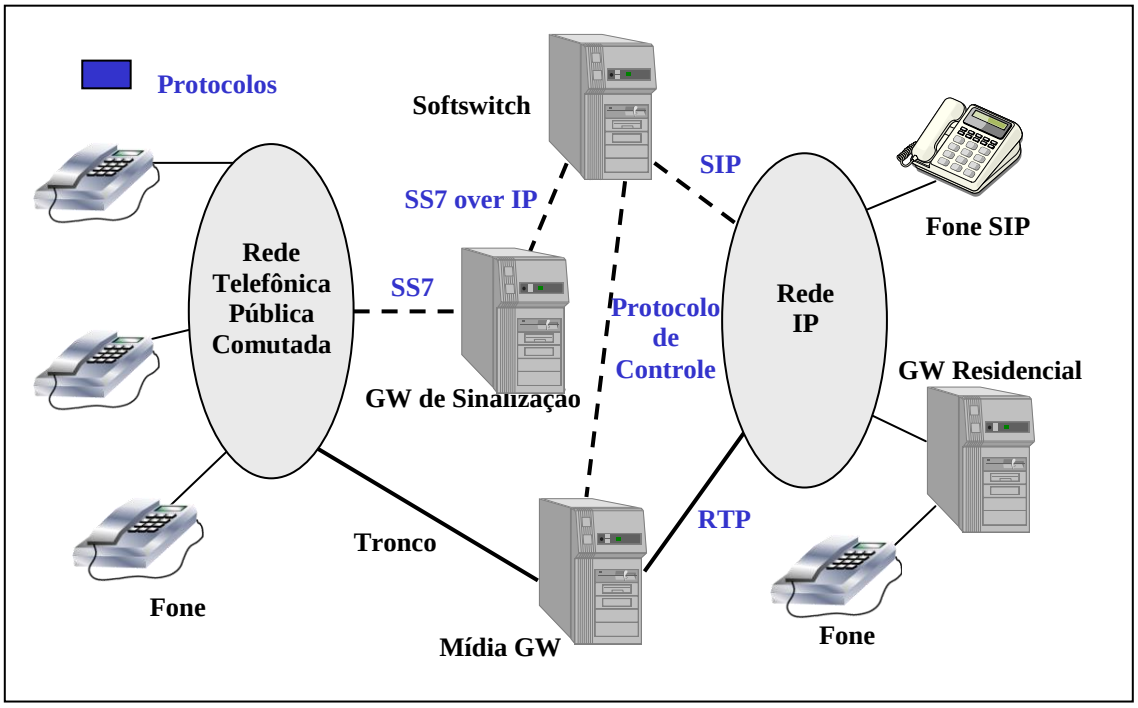


Figura 4.3 Rede telefônica híbrida (interconexão das redes de circuitos e pacotes).

4.2.2 Chamada Telefônica em uma Rede IP Utilizando o Protocolo SIP

Os protocolos SIP (*Session Initiation Protocol*) e SDP (*Session Description Protocol*) são protocolos de controle de sessão. São equivalentes funcionalmente ao H.225.0 e o H.245 do ITU-T (*International Telecommunications Union – Telecommunications Standardization Sector*). O SIP foi definido inicialmente pela MMUSIC (*Multiparty Multimedia Session Control*) grupo de trabalho do IETF (*Internet Engineering Task Force*) na RFC 2543 (Faynberg *et al.*, 2000; Hersent *et al.*, 2002).

O protocolo SIP foi projetado para estabelecer e terminar sessões multimídia. Sua sintaxe é similar ao HTTP (*Hypertext Transfer Protocol*). Porém, diferentemente do HTTP, o SIP foi desenvolvido com o intuito de endereçar usuários humanos, por essa razão o URI (*Uniform Resource Identifier*) se assemelha mais a um endereço de e-mail do que a um endereço de uma página *Web*. É importante notar também que, com o intuito de integrar a RTPC (Rede Telefônica Pública Comutada) com a Internet, a mensagem SIP também pode utilizar outras URIs (como URL – *Universal Resource Locator* – de telefone, definida pelo IETF).

As entidades SIP comunicam-se utilizando *transações*. O SIP denomina uma transação como sendo o conjunto formado por um *pedido* e pelas *respostas* que esse pedido dispara. O iniciador de um pedido é chamado de cliente SIP, e a entidade que responde, de servidor SIP. As mensagens compartilham um número Cseq (*command sequence*) comum, com uma exceção: a mensagem ACK usa a mesma Cseq que a transação a que ela se aplica, embora ela mesma já seja uma aplicação por si só.

São pedidos SIP:

- ACK: pedido enviado por um cliente para confirmar que ele recebeu uma resposta final do servidor, como, por exemplo, 200 OK.
- BYE: pedido enviado ou pelo agente de origem ou pelo de destino para interromper uma chamada.

- Cancel: pedido que pode ser enviado para interromper um outro pedido que foi enviado previamente enquanto o servidor não tiver enviado ainda uma resposta final.
- Invite: pedido usado para iniciar uma chamada.
- Options: um cliente envia esse pedido para um servidor para informar-se de suas capacidades.
- Register: os clientes podem registrar sua localização corrente (um ou mais endereços) com esse pedido.

Os servidores SIP respondem a um pedido SIP com uma ou mais respostas SIP. A maioria das respostas (2xx, 3xx, 4xx, 5xx e 6xx) é resposta final e termina a transação SIP. As respostas 1xx são provisórias e não terminam a transação SIP. As seis categorias de códigos de status foram definidas da seguinte forma:

- 1xx (Informativo): recebeu o pedido, continuando a processá-lo.
- 2xx (Sucesso): a ação foi recebida com sucesso, entendida e aceita.
- 3xx (Redirecionamento): uma ação melhor deve ser tomada a fim de completar o pedido.
- 4xx (Erro do cliente): o pedido contém um erro de sintaxe ou não pode ser completado nesse servidor.
- 5xx (Erro do servidor): o pedido contém erro de sintaxe ou não pode ser completado nesse servidor.
- 6xx (Falha global): o pedido é inválido em qualquer servidor.

A Figura 4.4 apresenta a troca de mensagens com sucesso entre duas entidades em uma chamada telefônica sobre IP (*Internet Protocol*).

Figura 4.4 Troca de mensagens SIP entre duas entidades.

O convite para uma sessão é acompanhado pelo protocolo SDP definido na RFC 2327. SDP fornece o formato de descrição dos endereços *unicast* e *multicast*, o número e tipos de *streams* envolvidos (áudio, vídeo, dados, controle), os codecs envolvidos (ou seja, os tipos de *payload* a serem carregados pelo protocolo de transporte), o protocolo de transporte a ser utilizado (por exemplo, RTP), a porta UDP (*User Datagram Protocol*), entre outras informações. Nota-se que, uma vez estabelecida a sessão, um outro protocolo, e.g. RTP, deve ser utilizado para o transporte da voz na chamada telefônica sobre IP.

4.2.3 Chamada Telefônica em uma RTPC Utilizando ISUP

O objetivo geral da SS7 (*Signalling System Number 7*) é prover meios confiáveis de transferência de informação no suporte a controle de chamadas, controle remoto, gerência e administração. É definido pelo ITU-T em vários documentos da série Q, especialmente a Q.7xx (Faynberg *et al.*, 2000).

Existem dois tipos de usuários SS7:

1. Aplicações que usam transações relacionadas ao serviço (mas não relacionadas ao circuito) entre centrais telefônicas e bases de dados.

2. Aplicações de comutação que se baseiam na troca de informações relacionadas a circuitos a fim de estabelecer, testar, manter e desconectar circuitos telefônicos.

As aplicações do segundo grupo usam ou o protocolo TUP (*Telephone User Part*) ou ISUP (*Integrated Services Digital Network User Part*). Esses protocolos utilizam o protocolo de roteamento MTP (*Message Transfer Part*).

Em uma rede de circuitos, pode-se utilizar o protocolo ISUP para realizar a sinalização entre centrais telefônicas. Entidades ISUP endereçam umas às outras utilizando um esquema de identificação do protocolo MTP (identificação da central), acrescido da identificação do circuito que será utilizado entre as centrais.

O modelo de chamadas ISUP possui três fases:

1. estabelecimento da chamada;

2. conversação (troca de dados);

3. desconexão da chamada.

Como existe um número elevado de mensagens ISUP, segue uma apresentação das categorias de mensagens:

- *Forward setup*. Nessa categoria, as mensagens estão envolvidas no estabelecimento de uma chamada com características particulares, no sentido da parte chamada.
- *Backward setup*. Nessa categoria, as mensagens completam o estabelecimento da chamada (quando possível) no sentido da central que contém a parte chamada para a que contém a parte chamadora. Possui os procedimentos de tarifação.
- *General setup*. As mensagens nesta categoria carregam informações adicionais relacionadas à chamada, necessárias ao estabelecimento dessa chamada.

- *Call supervision*. Nessa categoria, as mensagens são notificações de eventos relacionados ao andamento das chamadas, tal como a notificação da necessidade de intervenção por um operador.
- *Circuit supervision*. As mensagens nessa categoria são notificações dos eventos relacionados a circuitos alocados para a chamada.
- *Circuit group supervision*. As mensagens nessa categoria referem-se mais a grupos de circuitos do que a circuitos individuais, e são usadas para gerenciamento da rede.
- *In-call modification*. Nessa categoria, as mensagens dão suporte a modificações nas características de chamadas existentes ou solicitam um meio (facilidade) particular.
- *End-to-end*. As mensagens nesse grupo incluem sinalizações usuário a usuário independente de mensagens de controle de chamadas

A Figura 4.5 apresenta uma chamada básica e bem sucedida entre duas centrais telefônicas, sendo utilizadas as seguintes mensagens ISUP:

Initial Address Message **(IAM):** Mensagem enviada para frente para indicar tomada de um circuito de saída e para transmitir o número e outras informações relativas ao roteamento e tratamento da chamada.

Address Complete Message **(ACM):** Mensagem enviada para trás indicando que todos os sinais de endereço necessários para o roteamento da chamada para a parte chamada já foram recebidos.

Answer Message **(ANM):** Mensagem enviada para trás indicando que a chamada foi atendida.

Release Message **(REL):** Mensagem enviada em qualquer direção indicando a liberação do circuito por alguma causa especificada.

Release Complete Message **(RLC):** Mensagem enviada em resposta à mensagem REL.

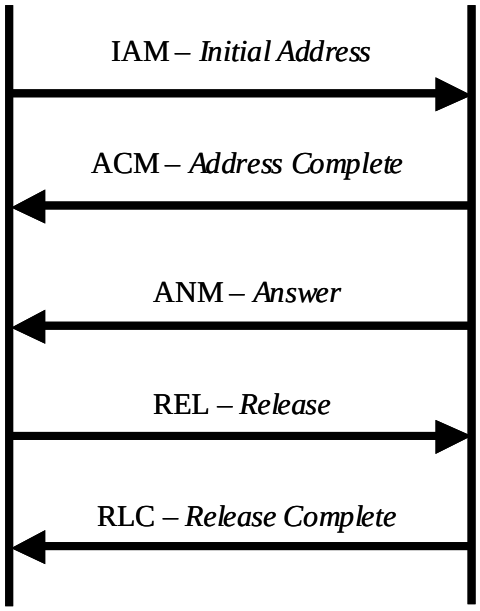


Figura 4.5 Troca de mensagens ISUP entre duas centrais.

As informações fundamentais utilizadas em uma chamada ISUP são: o endereço e categoria da parte chamante; o endereço, categoria e status da parte chamada; o meio de transmissão requerido para a chamada e a causa da liberação da chamada.

4.3 Modelagem do Sistema de Detecção de Falhas com Base nas Chamadas Telefônicas

Dependendo do tipo de rede e de protocolo específico utilizado, pode-se compor uma analogia entre o fluxo de mensagens e as células e moléculas do organismo. Essa abordagem no entanto pode não se demonstrar adequada a redes com o perfil híbrido (ver Figura 4.3) pelo fato de não se conseguir um modelamento homogêneo, utilizável em todo o sistema sob análise.

Contudo, independentemente da sinalização ou arquitetura utilizada, um sistema telefônico que se pretenda supervisionar possui vários parâmetros obtidos por meio de seus

protocolos, sejam eles quais forem. Com base nessas informações, é possível se criar um modelamento homogêneo, mesmo em uma rede híbrida.

A partir dos parâmetros das chamadas observadas, são modeladas estruturas que descrevem o comportamento habitual da rede telefônica. Fazendo um paralelo imunológico, esse seria o funcionamento saudável de um organismo. Continuando então a observação das chamadas de entrada, caso uma dessas se desvie do comportamento mencionado, isso equivaleria ou a uma mudança no comportamento da rede ou a ocorrência de uma falha. Na analogia imunológica, isso seria a detecção de um antígeno que poderia ou não oferecer perigo ao organismo. Tratando-se então da rede telefônica sob análise, pode-se confirmar a presença da falha com a ocorrência do sinal de perigo. No caso biológico paralelo, o sinal de perigo indicaria a necessidade de uma resposta imunológica diante de um antígeno que oferece perigo ao organismo.

Devido à possibilidade de se ter redes híbridas, ou de se utilizar vários protocolos dentro de uma mesma rede, os parâmetros utilizados nesse passo (modelagem do sistema de detecção de falhas) devem ser preferencialmente os essenciais a qualquer tipo de chamada (ver seções anteriores) e que busquem cercar algum sintoma de anomalia, e.g. identificação de um grupo de circuitos com falha. Devem assim ser salientados como parâmetros essenciais o endereçamento passado pelo protocolo (tanto de origem como de destino) e a duração da chamada obtida pela gerência da chamada. Outros parâmetros podem e devem ser acrescentados, contanto que de fato possam ser observados em qualquer chamada, não gerando dessa forma distorções na análise.

4.3.1 Perigo nas Redes de Telefonia

Os Sinais de Perigo em uma rede telefônica podem ser buscados fazendo uso da analogia biológica apresentada em Gallucci & Matzinger (2001). Como mencionado no Capítulo 1, esses sinais são classificados como constitutivos e induzíveis. Os constitutivos são aqueles acarretados pela própria morte necrótica da célula. Pode-se perceber esses sinais quando qualquer molécula intracelular não presente normalmente em ambiente extracelular, e que

seja detectável por APCs, é percebida por alguma APC. A segunda categoria de sinais, os induzíveis, são sinais emitidos por células coagidas, submetidas a uma situação de estresse.

Comparando uma chamada telefônica a uma célula e fazendo uma analogia com a teoria supracitada, pode-se ter sinais de perigo constitutivos como sendo as características anormais observadas no comportamento da rede, geradas pelo desvio abrupto de operação das chamadas finalizadas (células mortas). Como exemplo desse tipo de sinalização, pode-se citar especialmente uma taxa elevada de chamadas com duração próxima de zero e com status de finalização genérico. Na verdade, uma vez que o perfil global das chamadas, ou seja, o perfil das chamadas de todo o ambiente ou local de observação é traçado, qualquer variação significativa nesse perfil, que se observe após a finalização das chamadas, pode ser considerada um sinal de perigo constitutivo.

Continuando a analogia, pode-se observar também desvios de comportamento padrão ao longo das chamadas. Como exemplo, pode-se citar mudanças excessivas de rota e de recursos utilizados no estabelecimento da chamada. Nesses casos, são estabelecidas sinalizações durante a vida das células. Sinalizações acarretadas pela submissão dessas a condições de estresse, ou seja, sinais de perigo induzíveis.

Um sinal de perigo, embora relacionado às chamadas monitoradas, pode também possuir vínculos de mais difícil percepção, não obtidos diretamente pela observação das chamadas em questão. Por exemplo, podem ser utilizados como sinais de perigo alarmes de *hardwares* das centrais telefônicas, *hardwares* esses cujo mau funcionamento influencie na finalização de chamadas.

Os parâmetros analisados para a geração dos sinais de perigo devem ser escolhidos com cautela. Eles devem ser capazes de indicar alterações de comportamento, alterações no ambiente de observação. Dessa forma, esses parâmetros devem possuir um comportamento estável e conhecido.

A geração do sinal de perigo como uma mudança no comportamento global do sistema pode transmitir a idéia de que o mesmo é suficiente para determinar uma falha. O próprio sinal de perigo seria essa indicação. Mas, embora se conheça o comportamento

padrão da rede analisada, podem ocorrer de fato desvios de comportamento que não impliquem diretamente em falhas e sim em simples anomalias. Como exemplo, uma rede pode ter em determinado tempo um incremento abrupto no número de chamadas com duração próximo de zero. Esse incremento pode referir-se à rejeição da chamada por parte do usuário por razões normais (e.g. simplesmente a chamada de fato é curta) ou devido ao fato de queda de qualidade na transmissão de voz (razão essa considerada como uma falha).

Uma outra questão é a separação entre o que é falha e o que não é, diante de uma situação de perigo. A simples contagem dos eventos que acarretaram o sinal de perigo não consegue indicar a diferença entre as chamadas que de fato estão relacionadas a uma falha e aquelas que não estão. O algoritmo proposto busca, por meio de um processo de aprendizado, fazer a distinção entre os eventos analisados. Nesse processo, procura-se acompanhar a evolução do comportamento das tentativas de chamada, independentemente da velocidade dessa evolução, e busca-se por meio de mecanismos baseados na Teoria do Perigo evitar a intervenção humana no processo de síntese de um detector de falha. A Figura 4.6 busca ilustrar o que pode ocorrer quando da geração de um sinal de perigo. Nesse exemplo, há um não completamento, com causa não diagnosticada, que deve ser considerado como uma ocorrência comum naquele ambiente. Porém, com a ocorrência de outros três não completamentos, com causa igualmente não diagnosticada, é disparado um sinal de perigo. Nota-se então que, embora o sinal tenha sido deflagrado pela ocorrência dos quatro não completamentos, apenas três deles referem-se a uma falha.

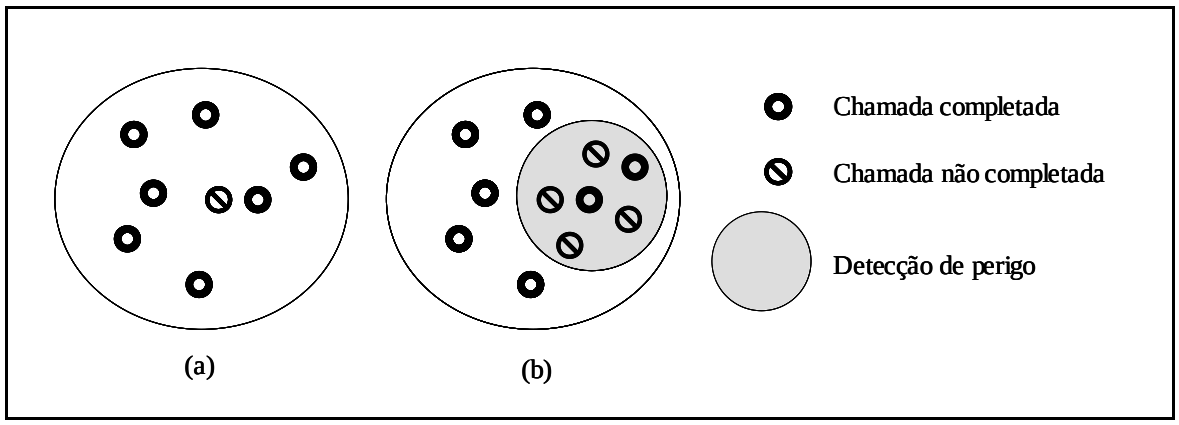


Figura 4.6 (a) Comportamento global padrão do sistema observado. (b) Ocorrência de anomalia.

Os resultados obtidos, ou seja, as indicações de falhas ocorridas no sistema, buscam complementar as informações obtidas pelos alarmes gerados que usam a simples contagem de eventos. Essa complementação de informações atua: (*i*) na indicação de continuidade de falhas já alarmadas, (*ii*) na indicação de falhas ou propensão a falhas ainda não alarmadas, (*iii*) na associação de tentativas de chamadas a um alarme específico, a fim de saber quais foram de fato as chamadas que possam ter originado um alarme, e ainda (*iv*) na associação de tentativas de chamada (associadas a um alarme) a uma dada região de parâmetros de tal forma que facilite a busca da causa da falha gerada.

A proposta desse trabalho, de utilizar a Teoria do Perigo na detecção de falhas em redes telefônicas, se mostra como uma medida inovadora com propriedades únicas. Somente projetos tão recentes, como mencionado nas seções 3.5 e 5.3, utilizaram a Teoria do Perigo na tentativa de detectar anomalias e falhas em sistemas. Embora esses trabalhos baseiem-se ainda na mesma teoria biológica, o modelamento sugerido por cada trabalho apresenta suas características próprias. O algoritmo proposto nessa dissertação busca se aproximar o mais fielmente da Teoria do Perigo e acrescentar rotinas complementares para aumentar a robustez do processo de detecção de falhas.

CAPÍTULO 5

O Algoritmo de Detecção de Falhas Baseado na Teoria do Perigo

Este capítulo visa apresentar o algoritmo de detecção de falhas, baseado na Teoria do Perigo de Polly Matzinger, com seus módulos e respectivas propriedades. Serão evidenciadas as analogias utilizadas entre o Sistema Imunológico Humano e um sistema telefônico, essenciais à elaboração do software em questão. Este capítulo mostra também o recurso de votação, adicionado aos módulos bio-inspirados, utilizado para aumentar a robustez do algoritmo. Por fim, busca fazer uma comparação entre o algoritmo proposto e trabalhos relacionados.

5.1 Introdução

Este capítulo mostra a elaboração de um sistema de detecção de falhas a partir da analogia entre condições anormais de operação em uma rede telefônica e o processo de reconhecimento antigênico sob uma situação de perigo, proposta por Matzinger (1994), no sistema imunológico humano.

Uma vez visualizado que no contexto biológico o objetivo primário desse sistema se assemelha aos objetivos do algoritmo almejado, o mapeamento direto dos elementos biológicos em elementos computacionais mostra-se como o principal requisito para a viabilização da proposta. Com isso, utilizaram-se as principais características dessa interpretação biológica na elaboração do algoritmo em questão.

Busca-se apresentar a implementação do algoritmo mostrando cada módulo utilizado e suas respectivas funções. Mostra-se também a especificação das variáveis e seus valores utilizados nos ensaios realizados. É exposta inicialmente uma visão geral do algoritmo, seguida pela apresentação da função utilizada na determinação da afinidade

entre as chamadas de entrada e os detectores. A seguir, são apresentadas as rotinas bio-inspiradas de ativação de detectores e detecção de falhas, morte de detectores, desativação de detectores e renovação da população de detectores. É também apresentado o processo de votação que visa aumentar a robustez do algoritmo. A seguir, é mostrada uma síntese do algoritmo por meio de uma seqüência de passos e por meio de diagramas de fluxo. Por fim, é feita uma comparação entre o algoritmo proposto e trabalhos relacionados, mencionados inicialmente na seção 3.5, mostrando os contrapontos mais relevantes.

5.2 Implementação do Algoritmo de Detecção de Falhas

Um algoritmo de detecção de falhas que toma como base a verificação de anomalias em um sistema de chamadas telefônicas pode encontrar tais anomalias sem detectar as falhas de fato (nem toda anomalia é uma falha). Caso toda anomalia seja considerada uma falha, ter-se-ia um acréscimo no número de falsos positivos (FPs). O que acontece é que a detecção de uma anomalia pode significar simplesmente a detecção de uma chamada fora dos padrões tidos como mais comuns para aquele sistema telefônico, naquele momento. Esta chamada pode, no entanto, não ser uma falha. Aqui se encontra um ponto relevante para a aplicação da Teoria do Perigo. Caso se consiga a definição de um sinal de perigo (sinal 2) adequado, o número de FPs pode ser diminuído. Isso pelo fato desse sinal de perigo buscar confirmar se a ocorrência de uma anomalia se refere de fato a uma falha.

Pela observação do significado dos elementos relacionados ao estabelecimento de uma chamada telefônica, e observação dos elementos constituintes de uma resposta imunológica, é possível extrair um mapeamento conforme descrito na Tabela 5.1.

Tabela 5.1 Analogia entre o SIH e o Sistema de Detecção de Falhas proposto.

Célula normal	Chamada telefônica dentro do padrão de normalidade da rede telefônica
Antígeno	Chamada telefônica fora do padrão de normalidade da rede telefônica
Linfócito T /B	Detector
APC	Módulo verificador de perigo (alarme)
Morte apoptótica	Terminação normal de chamada
Morte necrótica / sinal 0	Terminação anormal de chamada
Sinal 1 (estímulo)	Identificação de antígeno pela população de detectores

Com base nas analogias apresentadas, o princípio geral do algoritmo pode ser entendido da seguinte forma: uma chamada telefônica é tida como falha toda vez que um detector reconhecê-la e receber a confirmação, ou já possuir a informação, de que algum perigo está ocorrendo. Caso uma chamada (antígeno) seja reconhecida, porém esse detector não esteja de alguma maneira ciente da existência de perigo, essa chamada não é tida como falha. Chamadas não reconhecidas pelos detectores são sempre consideradas normais.

Dessa forma, uma vez que se perceba a presença de um antígeno por meio de um detector, um sinal 1 será disparado.

O sinal 2, assim como no caso biológico, deve alarmar uma situação de perigo. Em telefonia, várias situações podem ser alarmadas, por exemplo, situações de congestionamento, não completamento de chamadas, falhas de hardware, etc. No presente caso, a taxa de não completamento de chamadas foi escolhida pelo fato de ser uma medida que pode ser obtida diretamente dos dados dos protocolos de comunicação em telefonia, em especial dos dados considerados nesse ensaio. Assim sendo, nesse ensaio, o sinal 2 sempre

é diparado toda vez que $NC > L_{NC}$, onde NC é a taxa de não completamento de chamadas e L_{NC} é o limiar para a taxa de não completamento.

5.2.1 Afinidade entre Detectores e Tentativas de Chamadas

Em princípio, foram escolhidas três variáveis de grande significância para o estabelecimento e controle de qualquer tipo de chamada telefônica, que seriam: *origem* (origem da chamada), *destino* (destino da chamada) e *duração* (duração da chamada). Foi também escolhida uma quarta variável genérica, denominada *funcionalidade*, com o intuito de representar qualquer funcionalidade global observável em uma rede telefônica que seja relevante no controle de suas chamadas. Como exemplo de funcionalidades podem ser mencionados serviços oferecidos pela rede telefônica bem como a alocação de diferentes recursos necessários para a realização da chamada.

Cada chamada telefônica é representada no sistema de detecção de falhas por um antígeno, o qual deve ser comparado com a população de detectores do sistema. Tanto os antígenos quanto os detectores são modelados como *strings* heterogêneas, ou seja, compostas de atributos numéricos (no caso, *origem*, *destino* e *duração*) e atributo nominal (no caso, *funcionalidade*). Os atributos *origem* e *destino* foram considerados como sendo numéricos, embora pudessem ser considerados como atributos nominais.

Foi então definida uma função que indica a afinidade entre os detectores e os antígenos. Essa função indica se um antígeno está dentro ou fora de uma região ao redor do detector, a qual é estabelecida como sendo a região de afinidade. Essa abordagem baseia-se no conceito de espaço de formas introduzido por Perelson & Oster (1979). Tal conceito busca descrever quantitativamente a ligação entre células do sistema imunológico e antígenos, sendo que para que haja uma ligação é necessário um grau de afinidade mínimo entre as moléculas envolvidas. Essa afinidade baseia-se, entre outros fatores, na distribuição de cargas eletrostáticas e complementaridade de grupos químicos. Considerando que esses fatores que avaliam quantitativamente a afinidade entre moléculas sejam generalizados e representados em um espaço de dimensão *m*, pode-se entender esse espaço como sendo o

espaço de formas *S*. Ao redor de um ponto representado nesse espaço de formas pode-se então representar qual a região de afinidade para que se dê a ligação intermolecular (reconhecimento antigênico).

Para se medir a afinidade entre os detectores e os antígenos, pode-se em princípio utilizar vários tipos de função dependendo do espaço de atributos. Caso o espaço de atributos seja contínuo e real, dentre as funções possíveis, pode-se utilizar a distância euclidiana. Para tanto, os detectores e antígenos devem ser definidos como um ponto *p* em um espaço *S* ($p \in S \subseteq R^m$, onde *S* é o espaço de formas e *m* sua dimensão).

A distância euclidiana corresponde a:

$$D_E = \sqrt{\sum_{i=1}^{m} \left(Dt_i - Ag_i\right)^2} \qquad \textbf{(Equação 5.1)}$$

sendo $Dt = \langle Dt_1, Dt_2, \cdots, Dt_m \rangle$ e $Ag = \langle Ag_1, Ag_2, \cdots, Ag_m \rangle$ as coordenadas do detector e do antígeno, respectivamente.

Caso o espaço de formas apresentado, *S*, utilize a distância euclidiana, ele é denominado *espaço de formas euclidiano* (Segel & Perelson, 1988; De Boer *et al*., 1992; Smith *et al*., 1997).

Para um espaço de atributos nominal (*espaço de formas de Hamming*), a afinidade entre duas cadeias simbólicas pode se dar pela comparação direta de seus elementos (distância de Hamming):

$$D_H = \sum_{i=1}^{m} \delta_i \text{ , onde } \delta_i = \begin{cases} 1 & \text{se } Dt_i \neq Ag_i \\ 0 & \text{outros casos} \end{cases} \qquad \textbf{(Equação 5.2)}$$

Considerando a utilização do espaço de Hamming na verificação da afinidade entre antígenos e detectores modelados, são mostradas na literatura várias formas de se tratar essa questão. Tem-se, por exemplo, a afinidade entre os elementos medida pelo número de bits complementares (Forrest & Perelson, 1992; Hajela & Lee, 1996; Hightower *et al*., 1996),

ou pela quantidade de r-bits complementares consecutivos (Forrest *et al*., 1994; Dasgupta & Forrest, 1996).

Devido às características híbridas do espaço de formas, foi utilizada uma função heterogênea. A função é baseada em intervalos de afinidade para a comparação dos atributos numéricos (distância euclidiana), e na igualdade entre os atributos quando são considerados os atributos nominais (distância de Hamming). Essa função denomina-se HEOM (*Heterogeneous Euclidean-Overlap Metric*) (Wilson & Martinez, 1997), e é dada a seguir, sendo *i* o índice do i-ésimo atributo do detector (Dt_i) e do antígeno (Ag_i):

$$d_i = \begin{cases} 1 & \text{se } Dt_i \text{ ou } Ag_i \text{ são desconhecidos, senão} \\ \delta_i & \text{se } i \text{ é nominal, senão} \\ dif\_norm_i & \end{cases} \qquad \textbf{(Equação 5.3)}$$

onde δ_i é apresentado na Equação 5.2 e dif_norm_i é a diferença normalizada dada por:

$$dif\_norm_i = \frac{|Dt_i - Ag_i|}{\max_i - \min_i} \qquad \textbf{(Equação 5.4)}$$

sendo $\max_i$ e $\min_i$ os valores máximos e mínimos, respectivamente, observados para o atributo *i* do detector e antígeno em questão.

A distância total (considerando todos os *m* atributos) entre um detector e um antígeno é dada por:

$$D_{HEOM} = \sqrt{\sum_{i=1}^{m} d_i^{\,2}} \qquad \textbf{(Equação 5.5)}$$

5.2.2 Ativação de Detectores e Detecção de Falhas

Quando um sinal 2 é disparado, devem ser identificados os antígenos que se encontram em uma zona de perigo. Essa zona se refere a uma região de alcançabilidade do efeito do sinal 2.

A Figura 5.1 apresenta uma ilustração desse conceito. Ao ser disparado um sinal 2, cada detector que estiver dentro da zona de perigo deve verificar se identifica algum antígeno por meio de sua região de afinidade.

Para um melhor entendimento, deve-se observar cada cenário apresentado pela Figura 5.1. No cenário (a), existe um detector e ao redor dele várias chamadas representadas no espaço de análise. Esse detector reconhece chamadas que estão dentro de sua região de afinidade. Caso isso aconteça, ele assume ter recebido o sinal 1. No cenário (a), embora haja o reconhecimento de duas chamadas pelo detector, nenhuma das chamadas do espaço de análise é considerada como falha. Isso pelo fato de não ter sido detectado nenhum sinal de perigo.

No cenário da Figura 5.1 (b), detecta-se um evento que gera um sinal de perigo. Isso implica numa zona de efeito, ou seja, próximo a esse perigo detectado estabelece-se uma zona de perigo. Caso um detector esteja dentro dessa zona de perigo, ele assume ter recebido o sinal 2. No cenário (b), o detector recebe o sinal 2 sem receber o sinal 1, pois não há nenhuma chamada dentro de sua região de afinidade. Logo, também não é verificada nenhuma falha nesse cenário.

Finalmente, na Figura 5.1 (c), há o recebimento dos dois sinais (sinal 1 e sinal 2) pelo detector. O detector encontra-se na zona de perigo, recebendo assim o sinal 2, e detecta duas chamadas dentro de sua região de afinidade, recebendo assim o sinal 1. Logo, o detector é ativado e essas últimas chamadas detectadas por sua região de afinidade são chamadas com diagnóstico de falha.

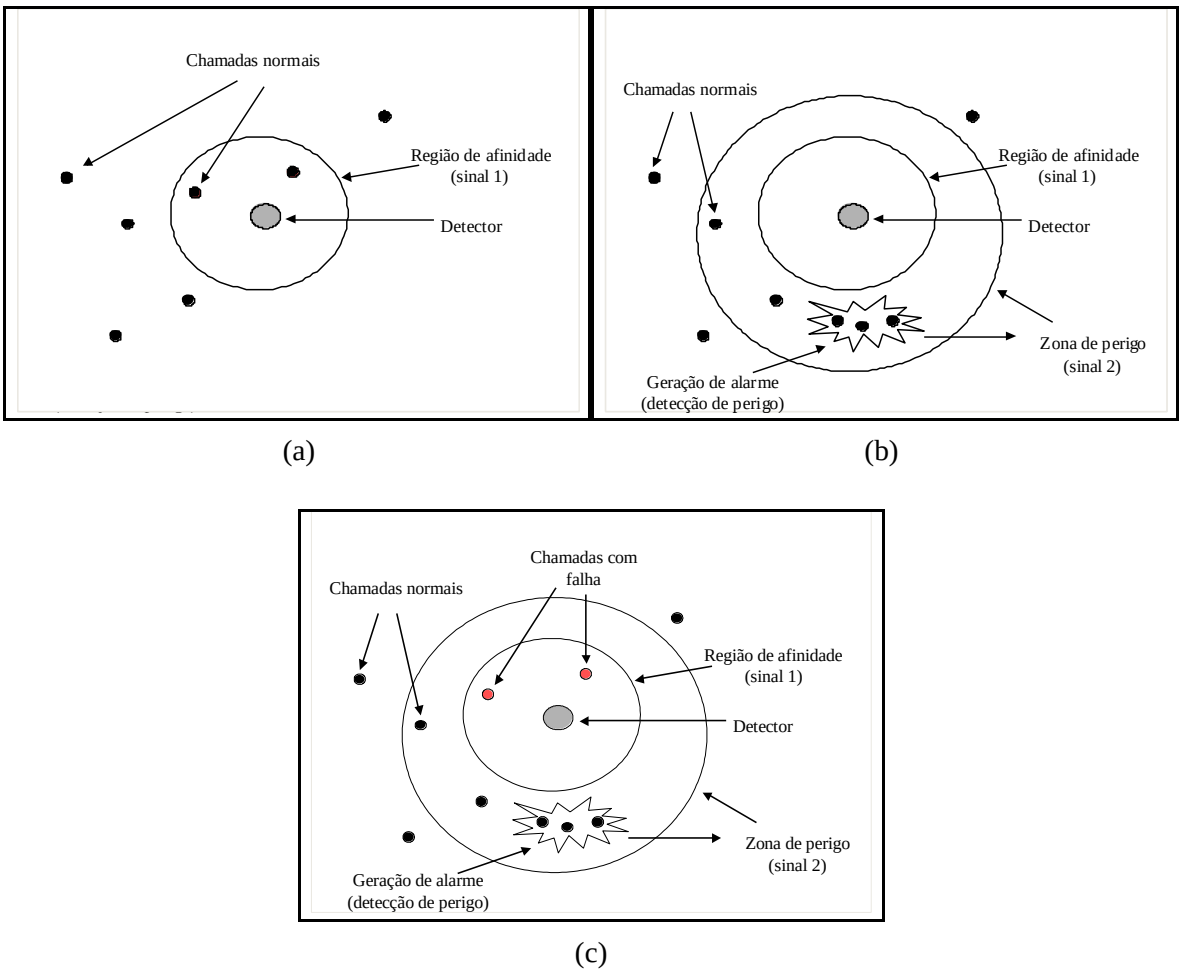


Figura 5.1 Cenários de comportamento de um detector na presença de chamadas. (a) detector recebendo somente o sinal 1; (b) detector recebendo somente o sinal 2; (c) detector recebendo o sinal 1 e o sinal 2.

Deve-se definir uma região que expresse uma relação de causa com o sinal 2. A região a ser adotada será uma região temporal, ou seja, uma vez que se tenha a indicação de um sinal 2, a região a ser analisada será o último intervalo de tempo *td*, verificando-se assim os detectores que acarretaram o disparo de um sinal 1 dentro desse espaço de tempo. Provavelmente, os antígenos (tentativas de chamadas) com alta afinidade a esses detectores contribuíram para o disparo do sinal 2.

Como a zona de perigo é temporal, a combinação da região de afinidade dos diversos detectores com a zona de perigo, na detecção das falhas, pode ser entendida como uma seqüência de estados do espaço representativo das tentativas de chamadas telefônicas,

onde alguns desses estados possuem indicação de perigo. Nesse espaço, existem detectores que verificam se as chamadas ocorridas pertencem ou não ao comportamento tido como padrão e verificam a ocorrência de falhas conforme já mencionado. Caso haja ocorrência de chamadas fora do comportamento considerado padrão, persistentemente, sem a ocorrência do sinal de perigo, o algoritmo interpreta que o comportamento padrão para as chamadas deve ser alterado.

Durante a adaptação do sistema, a fim de trabalhar com um novo perfil de detector (devido às mudanças no perfil das chamadas), o sinal de co-estímulo torna-se essencial para evitar que as chamadas anormais que irão compor o novo perfil do sistema (passarão a ser comportamento normal) sejam declaradas como falhas. Além disso, o sinal de co-estímulo evita a geração de FPs em resposta à ocorrência de chamadas anormais difusas que não geram falhas no sistema.

A Figura 5.2 busca apresentar o comportamento global do algoritmo. Inicialmente, de acordo com o quadro (a), detectores dispostos em pontos específicos do espaço de atributos descrevem o comportamento tido como fora do padrão para as chamadas telefônicas. Nesse exemplo, o espaço é formado apenas por dois atributos (variáveis genéricas A e B). Embora esses detectores descrevam esse comportamento, pela observação dos Quadros (a) e (b) nota-se a ocorrência de tentativas de chamadas insistentemente na região coberta por um dos detectores, porém, sem serem consideradas como falhas. Isso pelo fato de não ocorrer nenhum sinal de perigo dentro do intervalo de tempo referente aos Tempos 1 e 2. Esse evento é tido então como uma alteração de comportamento e os detectores são rearranjados no espaço de variáveis para descrever o novo comportamento do sistema, conforme se nota no Quadro (c). Então, num espaço de tempo seguinte (Quadro (d)), são detectadas novas tentativas de chamadas por detectores, porém, com a ocorrência de um sinal de perigo concomitante. Com isso, ocorre a ativação dos detectores que reconheceram as tentativas de chamadas mencionadas, e essas tentativas de chamadas são tidas como falhas.

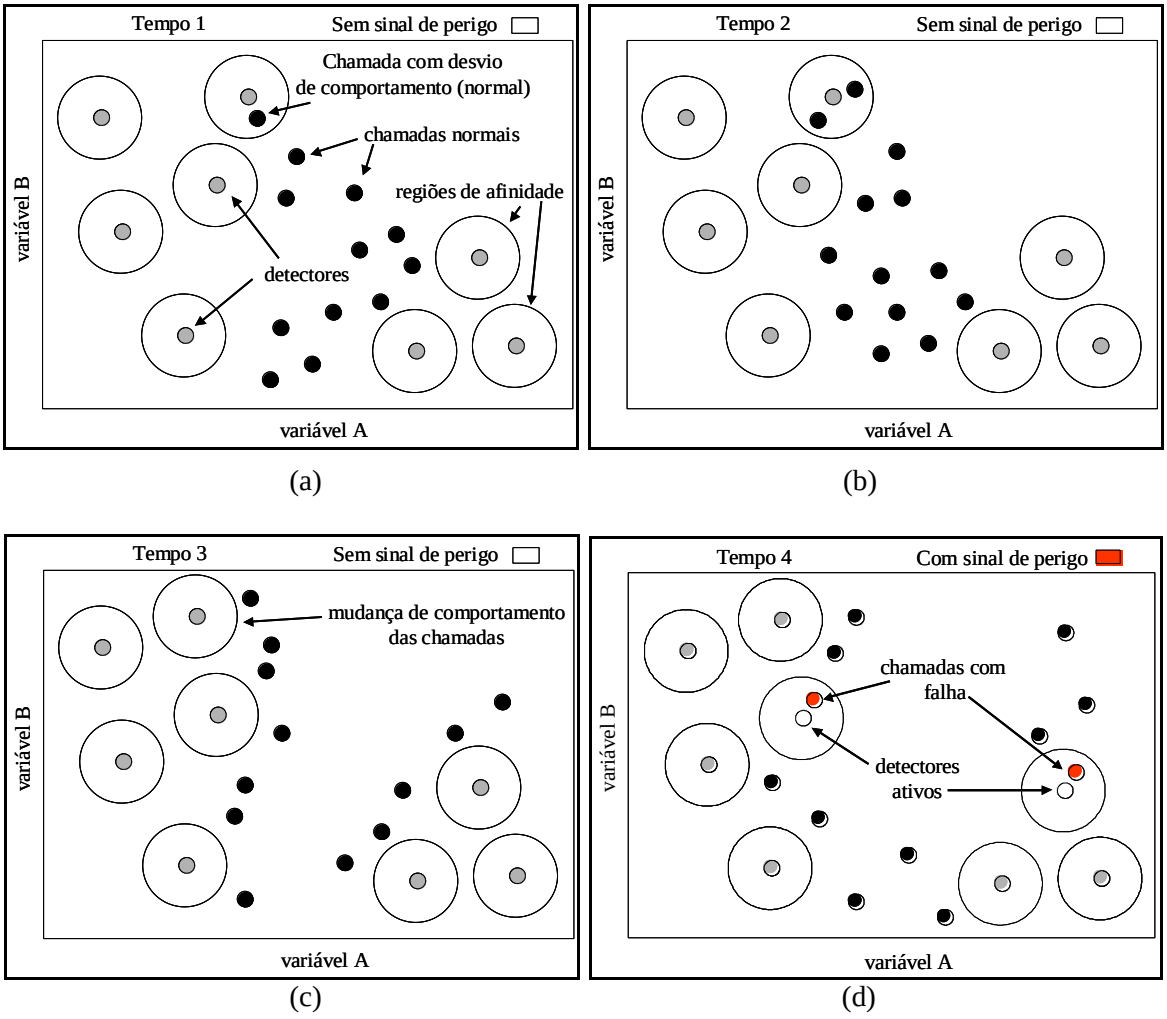


Figura 5.2 Seqüência de chamadas representadas em espaço de duas variáveis. (a) conjunto de detectores indicando a ocorrência apenas de chamadas normais, embora uma dessas esteja fora do comportamento padrão; (b) idem ao item (a), ocorrendo a persistência de chamadas fora do comportamento padrão; (c) mudança no perfil de detectores devido à persistência de ocorrência de chamadas normais em regiões de afinidade de detectores, nos Tempos 1 e 2, sem gerar falhas; (d) detecção de falhas ocasionada pela verificação de chamadas fora do padrão e pela ocorrência de indicação de perigo.

Conforme se busca exemplificar na Figura 5.3, um detector ativo não precisa do sinal 2 para indicar uma falha, basta identificar um antígeno (tentativa de chamada). Nota-se nessa figura que, no Tempo 5, dois detectores percebem três chamadas fora do comportamento tido como padrão. Por não ocorrer nesse intervalo de tempo nenhum sinal de perigo, apenas as chamadas detectadas pelos detectores ativos são tidas como falha.

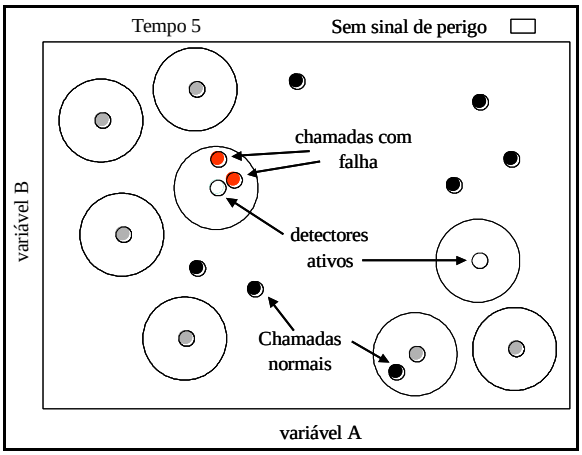


Figura 5.3 Detecção de falha fora da zona de perigo por um detector ativo.

5.2.3 Morte de Detectores

Como nas Leis da Linfótica da Teoria do Perigo, um detector deve ser eliminado caso receba um sinal 1 sem a ocorrência de um sinal 2. Do ponto de vista do algoritmo, essa eliminação é bastante significativa uma vez que esses detectores são potenciais geradores de falsos positivos (FPs).

Quando da ocorrência de mudanças no perfil das tentativas de chamada, especialmente em mudanças aceleradas, embora a rotina de renovação de detectores busque se adaptar a essas modificações, pode ocorrer o casamento entre detectores e células de entrada não relacionadas efetivamente a sinais de perigo. Caso ocorra, então, em um tempo próximo, o disparo de um sinal 2 vinculado a algum outro acontecimento, de tal forma que a zona de perigo seja ampla o suficiente para englobar o primeiro detector mencionado (o que já recebeu o sinal 1 sem um sinal 2, porém não possui vínculo com nenhum sinal de perigo), será gerado um co-estímulo para esse detector. Com isso teremos a ocorrência de um FP. Sendo assim, é necessário um ajuste adequado da zona de perigo para minimizar tais ocorrências.

5.2.4 Desativação de Detectores

Após um período de tempo T_{At} decorrido, um detector ativo se torna inativo. Esse ponto da Teoria do Perigo é observado tendo em vista o caráter adaptativo do sistema. Um antígeno que indica um perfil de tentativa de chamada com falha, em determinado período da análise, pode vir a ser uma tentativa de chamada comum, após um determinado período de tempo. Esses detectores que passaram pelo estado ativo passam a ser detectores de memória, viabilizando assim uma resposta secundária efetiva, ou seja, caso o sistema seja submetido a uma situação de falha semelhante, em que tentativas de chamadas com o mesmo perfil estejam envolvidas, a detecção de falha se torna mais rápida e eficaz (ver analogia biológica na seção 2.2.2.2).

5.2.5 Votação

Como forma de aumentar a confiabilidade no diagnóstico do algoritmo, utiliza-se uma rotina de votação. São utilizados múltiplos sistemas de detecção que devem ser executados em paralelo. Esses sistemas de detecção analisam independentemente as mesmas tentativas de chamadas entrantes. Uma tentativa de chamada é considerada efetivamente como falha caso um determinado percentual dos sistemas de detecção que fazem a análise a classifiquem como tal. Apesar de receberem os mesmos estímulos (por estarem monitorando o mesmo sistema telefônico), os sistemas de detecção não reagem da mesma forma e cada um vai apresentar uma população de detectores distinta das demais, visto que a geração de detectores é estocástica. Dessa forma, o uso de alguns sistemas de detecção, fazendo a mesma análise em paralelo, aumenta a robustez do algoritmo e justifica o emprego de um processo de votação. Processos semelhantes de votação são empregados em aprendizado de máquina, mais especificamente em *ensemble* de classificadores (Matan, 1996; Bauer & Kohavi, 1998; Gangardiwala & Polikar, 2005).

5.2.6 Renovação da População de Detectores

O algoritmo busca explorar a característica de adaptabilidade pela formação constante de uma população de detectores maduros inspirado no algoritmo de seleção clonal (Ayara *et al.*, 2002; de Castro & Von Zuben, 2002).

Como forma de se obter um bom desempenho, o algoritmo gerencia a geração de detectores, fazendo com que essa população seja suficiente para tratar o espaço de observação. Nesse contexto, o número de detectores passa a ser flutuante, adaptando-se à complexidade do cenário de operação a ser tratado.

Essencialmente, busca-se gerar ao longo do tempo detectores mais adaptados. Para tanto, mede-se a afinidade de cada detector ao conjunto das últimas tentativas de chamadas (Afinidade Global). Essa medida é então normalizada servindo como parâmetro na mutação. Os detectores são então clonados. Daí, seguindo o princípio da seleção negativa (Janeway *et al.*, 2002), quanto maior a afinidade do detector com os antígenos (deve-se notar que a grande maioria dos antígenos corresponde ao comportamento de normalidade), maior é a taxa de mutação que seus clones sofrem, pois se buscam detectores capazes de identificar situações de anormalidade. A seguir, é observada a afinidade dos detectores com a população de antígenos. Somente os detectores que não apresentem alta afinidade à população de antígenos e nem a outros detectores são mantidos (Figura 5.4 (a)). A questão de não apresentar afinidade a outros detectores visa aumentar a eficiência da população de detectores. Nota-se que, por esse processo de seleção negativa, os detectores tendem a se acomodar no espaço complementar ao dos antígenos, descrevendo assim o comportamento de anormalidade desejado.

Observam-se, assim, pontos importantes relativos à renovação constante da população de detectores, tais como: (*i*) busca pelo amadurecimento e diversidade da população, baseando-se no total de detectores; e (*ii*) adaptação do tamanho dessa população ao espaço de observação.

Esse procedimento de renovação é bastante significativo, uma vez que proporciona um amadurecimento global da população de detectores, seguindo a velocidade das

modificações que o perfil das tentativas de chamadas do sistema em observação sofre ao longo do tempo.

5.2.7 Síntese do Algoritmo

Conforme buscou-se apresentar nas seções anteriores deste capítulo, o algoritmo proposto consiste basicamente na interação de uma população de detectores com cada tentativa de chamada telefônica do sistema sob análise. Ao interagirem com a população de detectores, as tentativas de chamadas, ou antígenos, permitem que a população aprenda qual o comportamento normal do sistema. Dessa forma, caso ocorra alguma chamada fora dos padrões de normalidade aprendido, a população de detectores gera um sinal de detecção de anormalidade (sinal 1), e fica à espera de uma confirmação (sinal 2) de que aquela anormalidade é de fato uma falha. Caso seja confirmada a falha pela presença de um sinal 2, a população de detectores utiliza o padrão daquela falha (conjunto de características das tentativas de chamadas que geraram a falha) para a identificação de outras falhas.

Inicialmente é gerada, randomicamente, uma população de detectores (ver Figura 5.4(a)). Esses detectores encontram-se a princípio inativos. Cada tentativa de chamada do sistema (antígenos) passa então a interagir com essa população de detectores. Caso uma tentativa de chamada seja reconhecida por um detector inativo, gera-se um sinal 1 e o contador de afinidade (*Caff*), relativo àquele detector, é incrementado. Caso uma tentativa de chamada não seja reconhecida por um detector inativo, ela é adicionada à população de antígenos (a população de antígenos deve espelhar apenas os casos de tentativas de chamada normais).

A seguir, é verificado se a ocorrência da tentativa de chamada sob análise gerou um sinal 2 no sistema. Caso tenha gerado, e dentro da zona de perigo estabelecida (*td*) haja algum detector com o contador de afinidade indicando um número de detecções maior que o limiar para ativação do detector (*Caff* > *LCaff*), ocorre a ativação daquele detector e é gerada uma indicação de falha. Caso, dentro da zona de perigo mencionada, não exista nenhum detector com *Caff* > *LCaff*, a tentativa de chamada sob análise é adicionada à

população de antígenos. Deve-se notar que a zona de perigo *td* é contada retroativamente a partir do momento da tentativa de chamada sob análise.

Caso a ocorrência de uma tentativa de chamada não tenha gerado um sinal 2, verifica-se se tal ocorrência implica na eliminação de algum detector. Para tanto, é necessário que um detector possua seu contador de afinidade com valor superior ao limiar para eliminação do detector (*Caff* > *LCelim*) e esse detector não seja um detector de memória.

A interação das tentativas de chamadas com os detectores inicialmente inativos pode ativá-los. Com isso a população de detectores passa a ter tanto detectores ativos quanto inativos. Caso uma tentativa de chamada interaja com um detector ativo, verifica-se se o mesmo a reconhece. Caso isso ocorra, é gerado um sinal 1 (detecção de uma anormalidade) seguido de uma indicação de falha. Caso o detector ativo não reconheça a tentativa de chamada, não é gerado um sinal 1 e, conseqüentemente, não há nenhuma indicação de falha.

Decorrido um tempo T_{At} da ativação de um certo detector, esse é desativado e passa a ser um detector de memória.

Ao longo do tempo, alguns detectores vão sendo eliminados e há a necessidade de se criar novos detectores (ver Figura 5.4(c)). Esse processo de eliminação e criação faz com que a população de detectores aprenda o comportamento padrão do sistema analisado. Para a geração de novos detectores, utiliza-se um processo de clonagem com hipermutação da população de detectores. Processo similar ao CLONALG (de Castro & Von Zuben, 2002). Os novos detectores gerados interagem com a população de antígenos e com a população de detectores. Caso esses novos detectores sejam reconhecidos pelas populações citadas, são eliminados. Caso contrário, são adicionados à população de detectores. Isso deve ser feito para que não haja detectores com comportamento errôneo (identificando chamadas normais como sendo falhas), nem detectores com comportamentos similares.

Na simulação feita por esse trabalho, o sinal 2 é gerado pela verificação das características das tentativas de chamadas (ver Figura 5.4(b)). Caso uma seqüência de tentativas de chamadas implique em uma taxa de não complentamento de chamadas maior que o limiar para taxa de não completamento de chamadas ($NC > L_{NC}$), é gerado um sinal 2.

De forma simplificada, o algoritmo, para cada sistema de detecção votante, pode ser visualizado nos seguintes passos:

1. geração inicial randômica de detectores.
2. caso se receba o sinal 1 sem o sinal 2, o detector é (ou os detectores são) eliminado(s) (seleção negativa).
3. caso seja necessário, geram-se novos detectores para cobrir o espaço de defesa, utilizando clonagem com hipermutação de forma similar ao CLONALG (de Castro & Von Zuben, 2002).
4. caso se recebam os sinais 1 e 2, o detector inativo é (ou os detectores inativos são) ativado(s).
5. após um intervalo de tempo T_{At}, um detector ativo se torna inativo.
6. um detector ativo detecta falha com a presença do sinal 1 sem a necessidade do sinal 2.
7. um sinal 2 recebido sem um sinal 1 é ignorado.

O algoritmo é apresentado com mais detalhes nos fluxogramas da Figura 5.4.

As seguintes variáveis devem ser consideradas para o entendimento dos passos que constituem o fluxo da Figura 5.4:

D: distância entre dois elementos (detectores e antígenos)

Laff: limiar de afinidade (região de reconhecimento ao redor do detector)

Caff: contador de afinidade (verifica a quantidade de vezes que o sinal 1 é disparado)

LCaff: limiar para ativação do detector

LCelim: limiar para eliminação do detector

Nag: quantidade máxima de antígenos da população de antígenos

td: período de tempo após a ocorrência do sinal 1 para que, pela ocorrência do sinal 2, o detector seja ativado

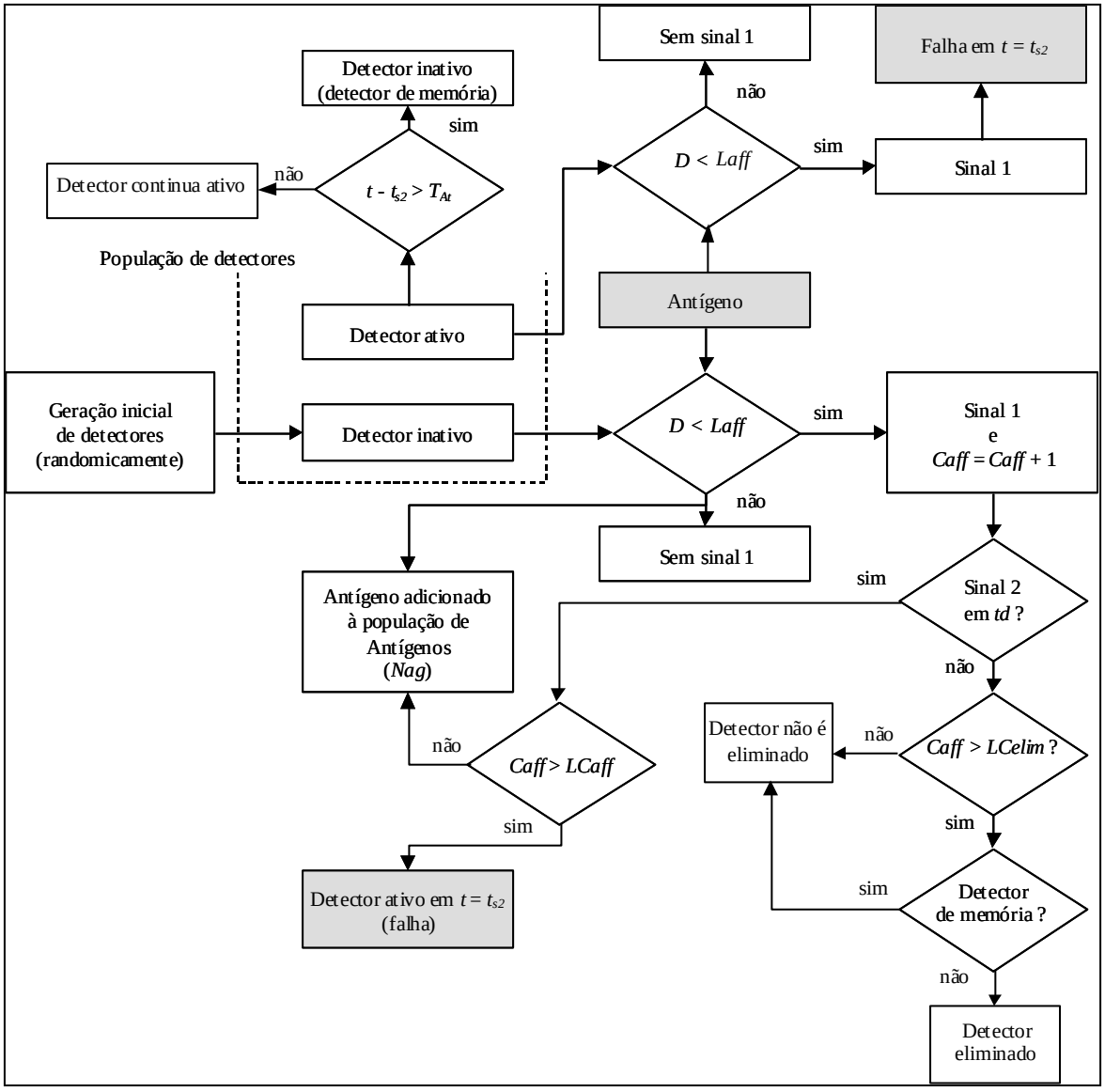
t_{s2}: tempo de disparo do sinal 2 ou detecção de falha

T_{At}: limiar de tempo para a desativação de um detector ativo

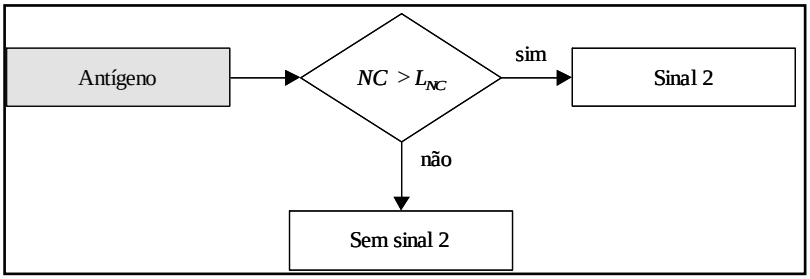
t: tempo da tentativa de chamada

NC: taxa de não completamento de chamadas

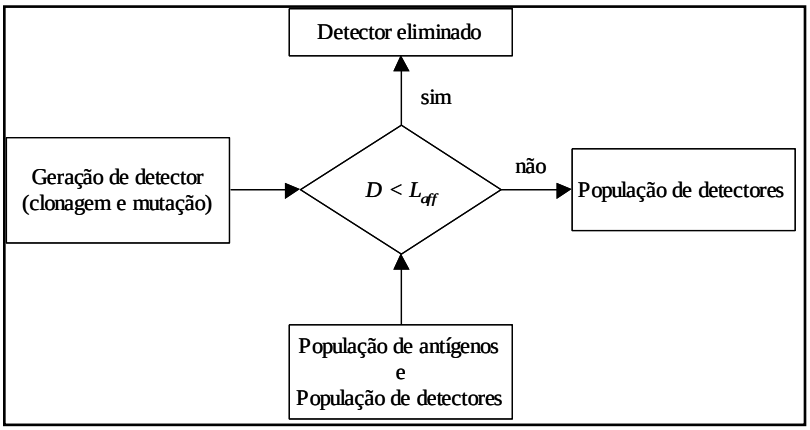
L_{NC}: limiar para taxa de não completamento de chamadas



(a) visão geral do algoritmo, mostrando o procedimento de ativação de detectores e detecção de falhas em cada tentativa de chamada.



(b) geração de alarme (sinal 2).



(c) geração de novos detectores ao longo do tempo

Figura 5.4 Detecção de falhas baseado no paradigma da Teoria do Perigo.

Como observações gerais, deve-se salientar que após a geração inicial de detectores (Figura 5.4(a)) essa população é constantemente renovada, como indicado na Figura 5.4(c). O fluxo na Figura 5.4(a) mostra o que ocorre quando um antígeno aparece no sistema (geração de uma tentativa de chamada) e o que ocorre quando o tempo de atividade de um detector expira. Nota-se ainda que um antígeno, para ser adicionado à população de antígenos, deve estar relacionado ao comportamento padrão das tentativas de chamadas (Figura 5.4 (a)). Isso se torna necessário pelo fato dos detectores serem gerados a partir da busca pelo espaço complementar ao dos antígenos (Figura 5.4 (c)). A Figura 5.4(b) mostra como o sinal 2 é gerado.

5.3 Comparação do Algoritmo de Detecção de Falhas Proposto com os Trabalhos Relacionados

Esta seção tem por objetivo fazer uma comparação de alto nível entre os algoritmos baseados na Teoria do Perigo, apresentados inicialmente na seção 3.5, e o algoritmo aqui proposto, buscando assim destacar suas principais diferenças. Alguns desses trabalhos são bastante incipientes, não oferecendo resultados práticos de um modelo implementado ou mesmo não propondo nenhum modelo a ser implementado de fato. Nesse último caso, somente idéias, ainda que interessantes, são lançadas para uma utilização futura inspirada na Teoria do Perigo.

Como ponto de diferenciação inicial em relação aos trabalhos mencionados na seção 3.5, o algoritmo aqui proposto apresenta um processo de votação na detecção de falhas, tornando-o mais robusto. Apresenta também uma análise baseada na observação de superfícies formadas a partir da área de afinidade dos detectores ativos, proporcionando uma verificação da isolabilidade das falhas e estimação do erro de classificação.

As comparações específicas para cada trabalho mencionado anteriormente são apresentadas a seguir.

Aickelin *et al.* (2003)

Em Aickelin *et al.* (2003), é lançada a idéia da utilização da Teoria do Perigo na detecção de intrusões em redes de computadores. Porém, não é apresentado nenhum modelo de fato. O trabalho consiste basicamente na apresentação da idéia da utilização de múltiplos alertas (apoptóticos e necróticos) que, combinados, poderiam vir a disparar uma ação imunológica, ou seja, a detecção de uma intrusão no sistema supervisionado. Essa idéia de diferenciação de alarmes não é explorada na proposta deste capítulo sendo considerada como uma proposta de continuidade do estudo.

Sarafijanovic & Boudec (2004)

Em Sarafijanovic & Boudec (2004), a relação entre os sinais 1 e 2 (ver seção 2.4) é objetiva. Embora na detecção de falhas utilizando a Teoria do Perigo esses sinais devam

estar relacionados, a sua relação pode não ser perceptível aparentemente. Em Sarafijanovic & Boudec (2004), uma vez que ocorra um sinal 2 ocasionado por um evento qualquer, a geração de detectores para a constatação de um sinal 1 também é diretamente relacionada à percepção daquele evento. No algoritmo apresentado neste capítulo, embora os sinais 1 e 2 possuam um relacionamento necessário para a constatação de uma falha, esse relacionamento pode se dar de forma aparentemente subjetiva. Isso pelo fato de, enquanto o sinal 2 é gerado pela observação direta de um evento qualquer, o sinal 1 é gerado sempre pela observação de uma modificação abrupta e concentrada espacialmente (espaço de atributos observados) no comportamento interpretado como sendo normal. Essa interpretação do comportamento tido como normal pode ser sempre modificada ao longo do processo de análise. Enquanto no algoritmo de Sarafijanovic & Boudec (2004) a interpretação de normalidade vincula-se ao não relacionamento dos objetos observados com o sinal 2, no algoritmo proposto neste capítulo a normalidade relaciona-se ao comportamento habitual da rede. Deve ser ressaltado que a interpretação do que é ou não uma mudança abrupta ou espacialmente concentrada também é definido pelos parâmetros do algoritmo. Essa abordagem diferenciada do algoritmo aqui proposto permite a detecção de falhas ocasionadas por eventos de percepção mais complexos, como por exemplo diminuição no número de chamadas realizadas devido à queda de qualidade nos enlaces de voz.

Em Sarafijanovic & Boudec (2004), também existe o conceito de células, ou detectores, de memória. Uma vez que um detector é ativado, pode se tornar um detector de memória e esse não precisa receber o sinal 2 para a detecção. Já no algoritmo apresentado neste capítulo, busca-se a utilização de um conceito de células de memória mais próximo do teórico. Após um determinado tempo de ativação, um detector pode ser desativado e então se tornar um detector de memória. Para uma nova resposta imunológica (detecção de falha), esse detector de memória deve ser novamente ativado pela presença do sinal 2. Isso se torna interessante devido às mudanças de comportamento que as chamadas podem ter. A desativação e possível reativação de um detector de memória implica na redução de potenciais falsos positivos.

Bentley *et al.* (2005)

A diferença principal entre Bentley *et al.* (2005) e o algoritmo proposto neste capítulo encontra-se na interpretação da ocorrência de um sinal de perigo. Seu algoritmo busca montar um tecido que espelha os padrões de normalidade do sistema sob análise. Para Bentley *et al*. (2005), a geração de um sinal de perigo (sinal 2) seria acarretado sempre pelo aparecimento e desaparecimento de um antígeno fora dos níveis de normalidade. Ou seja, uma vez observados os padrões de entrada e se obtido ao longo da história do processo um comportamento estável, compõe-se assim o tecido desejado, que espelha os padrões de normalidade (salubridade). Caso apareça um padrão que não se assemelhe aos anteriores e nem esteja acompanhado de padrões posteriores semelhantes ao mesmo, esse padrão não consegue assegurar seu lugar no tecido proposto, e seu desaparecimento é tido como uma morte necrótica. Conseqüentemente, é gerado um sinal de perigo e analisada a proximidade espacial e temporal desse sinal com outras células para identificar as células "não saudáveis".

No algoritmo apresentado neste capítulo, deve-se identificar um sinal 1 por uma mudança abrupta no comportamento do sistema. Essa mudança, no entanto, pode ser prolongada, propondo-se assim a identificar falhas que acarretem, de igual forma, uma mudança de comportamento prolongada no sistema sob análise. Esse sinal (sinal 1) utiliza uma medida de proximidade espacial para fazer essa identificação inicial. O sinal de perigo gerado aponta para um desvio de comportamento global do sistema (e.g. taxas de congestionamento e não-completamento), podendo ainda não estar visivelmente relacionadas aos antígenos (chamadas) analisadas (e.g. alarmes diversos em centrais telefônicas). O sinal de perigo utiliza uma análise de proximidade temporal. O casamento dos dois sinais aponta para a falha. Como já mencionado, essa abordagem permite a identificação de falhas prolongadas.

O algoritmo deste capítulo apresenta ainda o conceito de células de memória visando melhorar o tempo de resposta secundária (ver seção 2.2.2.2).

Greensmith *et al.* (2005)

Greensmith *et al.* (2005) apresentam um novo paradigma, o das células dendríticas. A partir dos dados de entrada, é gerada uma saída classificatória onde antígenos (dados de entrada sob análise) são apresentados por uma célula dendrítica que possui ou uma função estimuladora ou supressora, possibilitando assim a classificação desses antígenos. No caso de um algoritmo de detecção de falhas, essa classificação seria normalidade ou falha. O exemplo mostrado no trabalho utiliza, em princípio, dados estáticos.

Uma limitação apresentada no trabalho de Greensmith *et al.* (2005) seria a geração de erros de classificação caso os dados de entrada mudem de estado (e.g. normalidade e falha) múltiplas vezes em rápida sucessão. Isso pelo fato das células dendríticas utilizadas no algoritmo apresentarem múltiplos antígenos, e esses dentro de um mesmo contexto (estímulo e supressão), podendo, de fato, alguns desses antígenos apresentados possuírem uma outra classificação necessária.

O algoritmo proposto neste capítulo não está sujeito a essa limitação citada, ou seja, propõe-se a detectar falhas em antígenos intercalados múltiplas vezes e em rápida sucessão com antígenos normais (uma vez encontrado os padrões de chamadas com falha, os eventos de falha ou normalidade podem ser detectados independentemente de seu posicionamento na seqüência de chamadas).

Outros pontos cobertos pelo algoritmo deste capítulo, como por exemplo memória, não foram mencionados pelo modelo da célula dendrítica. Há porém no modelo de Greensmith *et al.* (2005) a interessante abordagem da utilização de vários sinais, não somente os utilizados no trabalho desta dissertação. Esses sinais, com efeito não somente estimulador, como o sinal de perigo, mas também supressor, como o sinal de segurança, podem ser objetos de estudo posterior para aumentar a eficiência e aplicabilidade do algoritmo proposto neste capítulo, embora possam tornar a operação do sistema como um todo mais complexa e, portanto, de gerenciamento e calibração mais difícil.

Kim *et al.* (2005)

Kim *et al.* (2005) utilizam como base os algoritmos propostos por Bentley *et al.* (2005) e Greensmith *et al.* (2005) na defesa computacional. Em seu trabalho, são possíveis diferentes tipos de resposta devido à diferenciação feita entre sinais de perigo. É apresentado no artigo de Kim *et al.* (2005) um esquema proposto para o seu algoritmo, embora a implementação e análise não sejam consideradas. Algumas características de diferenciação entre o algoritmo deste capítulo e o de Kim *et al.* (2005) devem fazer parte das análises acima, que se referem a Bentley *et al.* (2005) e Greensmith *et al.* (2005). Porém, para uma comparação mais efetiva, torna-se necessária uma descrição mais detalhada dos procedimentos utilizados pelo algoritmo em questão bem como uma análise inicial de seus resultados obtidos, não disponibilizados no trabalho referenciado.

CAPÍTULO 6

Resultados Obtidos

Este capítulo apresenta os resultados obtidos baseando-se em simulações efetuadas com o algoritmo. Busca validar as proposições e mostrar como o sistema se comportou nos ensaios realizados. Por último, esse capítulo apresenta o comportamento do algoritmo sem a utilização do sinal de perigo, procurando assim contrastar as visões da Teoria do Perigo com a visão Próprio/Não-Próprio.

6.1 Introdução

A implementação do algoritmo teve como objetivo validar as proposições e analisar a emergência de suas principais características. Os aspectos mais relevantes a serem monitorados são a taxa de falsos positivos (FPs), a eficiência do algoritmo na detecção de falhas, o comportamento do algoritmo frente à variação de seus principais parâmetros e as questões envolvendo a adaptabilidade dos detectores ao longo do tempo.

Ao longo do estudo desse algoritmo, foram implementados vários cenários com diversos parâmetros e valores. Os resultados apresentados nesse capítulo referem-se a um caso particular que consegue exemplificar o bom desempenho obtido com o algoritmo.

Juntamente com a implementação do algoritmo (ver Figura 5.4), implementaram-se várias rotinas para a análise dos resultados obtidos. Para a utilização do algoritmo na simulação, devem ser fornecidos inicialmente como dados de entrada os limiares utilizados (ver seção 5.2.7). À medida que se roda a simulação, é ecoado na tela o número da chamada sob análise, se algum detector foi ativado e se alguma falha foi detectada. As rotinas de análise verificam a sensibilidade do algoritmo frente à variação de parâmetros específicos (ver seção 6.2), resumem o comportamento do algoritmo frente aos não-

completamentos de chamadas (ver Figuras 6.5 e 6.6) e ainda esboçam o comportamento médio do algoritmo em uma região de falhas, traçando uma superfície formada pela soma das regiões de afinidade dos detectores ativos (ver Figuras 6.7 e 6.8). Todas as rotinas foram implementadas em Matlab.

Foram definidos os seguintes intervalos de valores para os atributos de chamada (ver comentário sobre os atributos na seção 5.2.1):

- *Origem*: 0 – 11 (índices dos nós na rede telefônica)
- *Destino*: 0 – 11 (índices dos nós na rede telefônica)
- *Duração*: 0 – 359 (em segundos)
- *Funcionalidade*: 0 – 3 (admite quatro tipos diferentes de chamadas)

Os intervalos de valores para os atributos de chamadas foram escolhidos de forma a propiciar um estudo de caso simples, sem sobrecarga de processamento, que fosse capaz de demonstrar as propriedades computacionais almejadas pelo algoritmo. A escolha de um espaço de análise com mais atributos e com atributos de maior dimensão, embora pudesse representar melhor uma rede real, poderia implicar na necessidade de uma população de detectores de dimensão bastante elevada e de um aumento sobremodo expressivo do esforço computacional. A redução desse espaço, contudo, não implica na perda das propriedades computacionais necessárias para a identificação de falhas, alcançáveis pelo algoritmo proposto. A utilização de equipamentos especializados (com capacidade de processamento adequado) ou estratégias computacionais específicas, como por exemplo uma análise hierarquizada de falhas (ver seção 7.3), podem ser necessárias na utilização, em casos reais, do algoritmo proposto.

Foram definidos os seguintes parâmetros do algoritmo:

- limiar para ativação do detector (*LCaff*): 3 tentativas de chamadas
- limiar para eliminação do detector (*LCelim*): 3 tentativas de chamadas

- quantidade máxima de antígenos da população de antígenos (*Nag*): 50 antígenos
- limiar de tempo para desativação do detector ativo (T_{At}): 60 tentativas de chamadas (o tempo médio de 60 tentativas de chamadas)
- limiar para taxa de não completamento de chamada (L_{NC}): 3 tentativas de chamadas

Os valores atribuídos aos parâmetros do algoritmo foram obtidos interativamente de forma a propiciar um bom desempenho do algoritmo e um esforço computacional reduzido.

Foram geradas 500 chamadas, sendo que as variáveis *origem*, *destino* e *funcionalidade* dessas chamadas seguiram uma distribuição uniforme, enquanto a variável *duração* seguiu, para as chamadas de 1 a 70, uma distribuição assimétrica enviesada para a direita com média 70 segundos (distribuição com a média deslocada para a direita da mediana) e, para as chamadas de 71 a 500, uma distribuição simétrica centrada em 180 segundos (o ponto médio do intervalo 0 – 359). A forma das distribuições utilizadas para a variável duração não é de fato o mais relevante. Quaisquer distribuições poderiam ser utilizadas no ensaio, sendo a utilização de duas distribuições diferentes, seguidas temporalmente (até a chamada 70 uma distribuição e a partir da 71 outra distribuição), uma forma de se verificar a adaptação dos detectores ao fim do processo. Em sistemas reais, a distribuição adotada pode de fato não ocorrer. No entanto, o mais significativo para a análise é a mudança do comportamento. Esse fenômeno qualitativo pode realmente ocorrer e isso é o que a simulação objetiva verificar.

No conjunto de chamadas geradas, foi inserida uma situação de falha após a chamada de número 400. Toda chamada gerada nas origens de 1 a 6 e que utilizassem a funcionalidade 2 não seriam completadas (teriam duração igual a 0). Fora isso, na geração das chamadas de 1 a 500 qualquer outro não completamento poderia de fato ocorrer, porém isso deveria ser tido como um comportamento normal do sistema (e.g. não completamento por falta de resposta do usuário no destino). As chamadas são geradas randomicamente e algumas delas podem mostrar esse perfil. Elas ocorrem de forma dispersa, com duração

igual a zero e possivelmente com origem, destino e funcionalidade diferente daqueles escolhidos para gerar as falhas. Então, vários processos executando o algoritmo em paralelo diagnosticaram as chamadas do ensaio por meio de votação (ver seção 5.2.5). É evidente que a situação de falha acima descrita não é informada ao sistema de detecção, sendo usada apenas no processo de geração das 500 chamadas.

6.2 Análise de Sensibilidade

Inicialmente, verificou-se o comportamento do algoritmo tendo como foco o comportamento temporal de suas principais variáveis. Diante de seu embasamento na Teoria do Perigo, onde a presença do sinal 2 (sinalização de perigo) adiciona novos ganhos às teorias anteriores embasadas na distinção exclusiva entre o próprio e o não próprio, buscou-se observar o comportamento do algoritmo frente à variação do parâmetro *td* (período de tempo após a ocorrência do sinal 1 para que, pela ocorrência do sinal 2, o detector seja ativado), ou seja, frente à variação da zona de perigo.

Para a análise da rotina de votação, foram escolhidas as principais variáveis desse processo, a saber: o número de processos votantes e o limiar para votação. O limiar para votação consiste em um percentual de votos mínimo para que uma tentativa de chamadas possa ser considerada uma falha. Por exemplo, caso o limiar para votação seja 10%, significa que é necessário que pelo menos 10% dos processos votantes indiquem que aquela chamada é uma falha para que essa chamada possa ser classificada como uma falha.

Foram atribuídos os seguintes valores aos parâmetros mencionados:

- *td* (*zona de perigo*): 15, 20 e 60 tentativas de chamadas (o tempo médio de 15, 20 e 60 tentativas de chamadas)
- *número de processos votantes*: 1, 2, 5, 10, 25 e 50 processos votantes
- *limiar para votação*: intervalo de 0 a 50%

Pôde-se então observar o percentual de falhas detectadas e o percentual de falsos positivos gerados. Os resultados referentes ao percentual de falhas detectadas são

mostrados nas Figuras 6.1(a), 6.2(a) e 6.3(a), para extensões de zona de perigo (*td*) iguais a 15, 20 e 60 chamadas, respectivamente. Os resultados referentes ao percentual de falsos positivos gerados são mostrados nas Figuras 6.1(b), 6.2(b) e 6.3(b), para extensões de zona de perigo (*td*) iguais a 15, 20 e 60 chamadas, respectivamente.

A seguir, é apresentado um exemplo com o intuito de facilitar a compreensão das Figuras 6.1 a 6.3. Caso se tenha uma zona de perigo com extensão de 20 chamadas, 10 processos votantes e um limiar de votação de 30% dos votos, deve-se tomar o gráfico (iv) da Figura 6.2(a) para análise do percentual de falhas detectadas (percentual referente a toda a seqüência de 500 chamadas do ensaio) e o gráfico (iv) da Figura 6.2(b) para análise do percentual de FPs gerados (percentual referente a toda a seqüência de 500 chamadas do ensaio). O número de processos votantes igual a 10 e o limiar de votação igual a 30%, implicam na utilização de 10 processos analisando em paralelo a seqüência de chamadas do sistema, e na necessidade de pelo menos 3 desses processos indicarem simultaneamente a ocorrência de uma falha, para que essa falha seja considerada pelo sistema de detecção (caso não se tenha pelo menos 3 votos a chamada é tida como normal). Analisando-se então os gráficos mencionados, pode-se observar que o comportamento médio do algoritmo implica na detecção de 90% das falhas e na geração de 0% de falsos positivos para os limiares utilizados nesse exemplo.

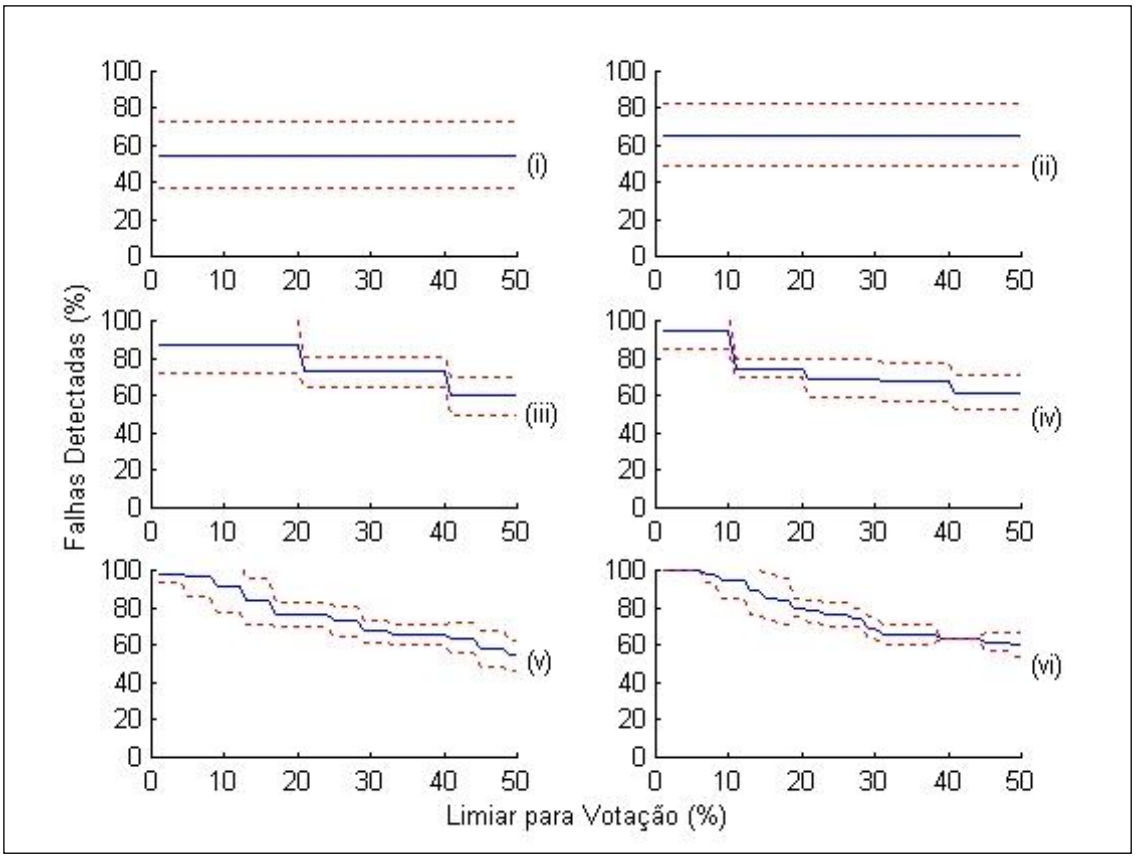
Com base nos gráficos das Figuras 6.1 a 6.3, pode-se inferir sobre o comportamento médio do algoritmo frente à extensão da zona de perigo. Para esboçar o comportamento médio mencionado, cada gráfico das Figuras 6.1 a 6.3 apresenta o valor médio para 5 execuções do algoritmo (linha contínua nos gráficos) e o intervalo de confiança (ver Miller & Freund, 1985) de 95% (linhas pontilhadas nos gráficos). Deve-se observar que os limiares superior e inferior dos intervalos de confiança são simétricos em relação à media. Porém, uma vez que os valores nos gráficos em questão variam de 0% a 100%, o limiar inferior e superior podem ficar limitados, respectivamente, a esses valores. Quanto maior a zona de perigo, maior é o número de falhas detectadas. Porém, isso também implica em um aumento no número de FPs gerados. O que ocorre nesse ensaio é uma mudança repentina de comportamento após a chamada de número 70. Tentativas de chamadas que eram

consideradas anormais passam a ser consideradas normais e existe um certo tempo para que o algoritmo possa se adaptar completamente a essa mudança. Com a zona de perigo aumentando seu tamanho, ela passa a ativar os detectores que entendem as novas tentativas de chamadas (normais) ainda como anormalidades.

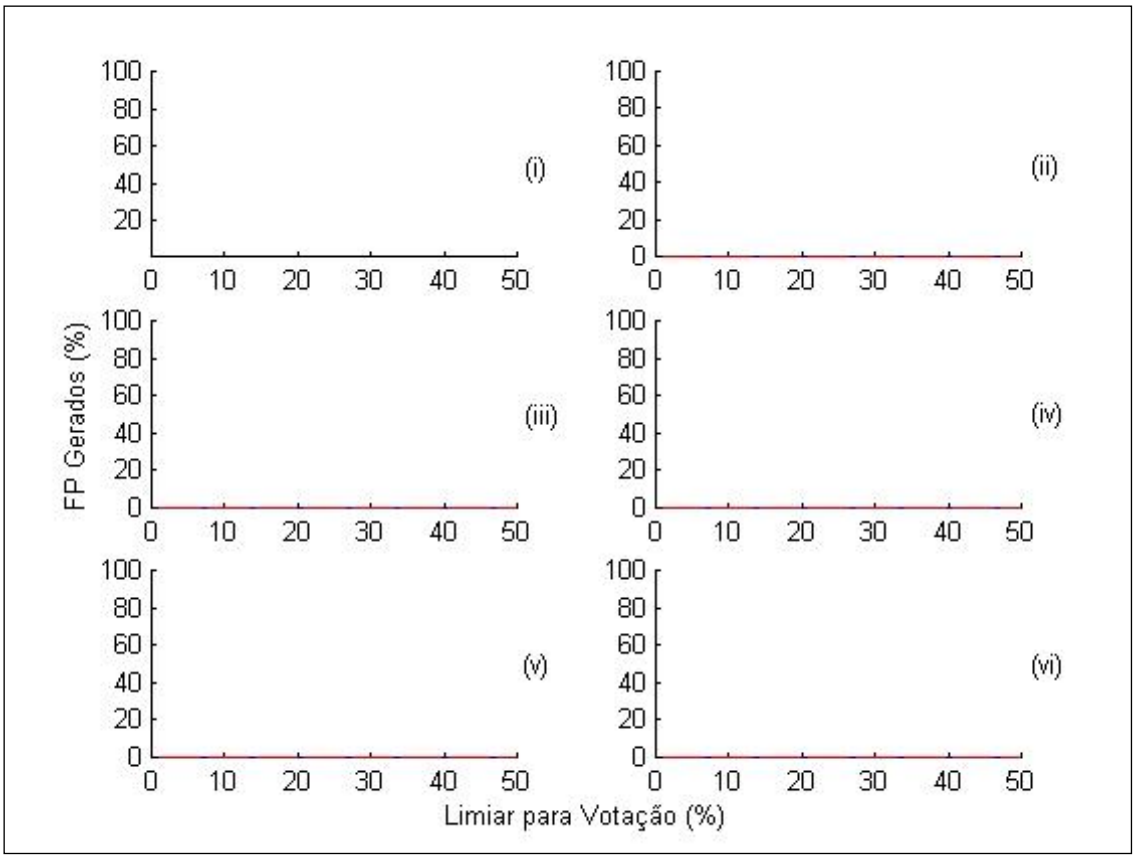
Existe um ponto a partir do qual aumentar a zona de perigo implica apenas em aumentar o número de FPs. Tal propriedade pode ser visualizada na comparação das Figuras 6.2 e 6.3. A Figura 6.2 possui uma zona de perigo com extensão de 20 chamadas. Na Figura 6.3 aumenta-se a extensão da zona de perigo para 60 chamadas e basicamente o que se observa é o aumento no número de FPs, não havendo um aumento significativo no percentual de falhas detectadas. Na análise realizada, as zonas de perigo correspondentes às extensões de aproximadamente 15 ou 20 tentativas de chamadas obtiveram o melhor desempenho.

Com relação à rotina de votação, nota-se que o aumento no número de processos votantes também implica em aumento no número de falhas detectadas, mas também no número de FPs gerados. Tomando-se como exemplo a Figura 6.3 e considerando um limiar para votação em 10%, nota-se na Figura 6.3(a) o aumento do percentual de falhas detectadas de um valor próximo a 50%, quando existe apenas 1 processo em execução, para um valor próximo a 100% de falhas detectadas, quando há 50 processos votantes. Em contrapartida, na Figura 6.3(b), o valor de FPs passa de aproximadamente 10% para aproximadamente 20%. Esse é o comportamento padrão que se nota nos gráficos. Observa-se também que existe um ponto a partir do qual não compensa aumentar o número de processos votantes. Isso pelo fato de não se ter um aumento significativo no número de falhas detectadas, por se ter um aumento no número de FPs gerados e ainda pela necessidade de se aumentar o esforço computacional significativamente para se obter esses pequenos incrementos na detecção de falhas. Como exemplo, pode-se observar novamente a Figura 6.3(a). Nota-se que a diferença no percentual de falhas detectadas de quando se tem 10 processos votantes para quando se tem 50 processo votantes não é significativa, embora a diferença no custo computacional o seja. Nota-se ainda nesse exemplo que na Figura 6.3(b), para valores baixos de limiares para votação, o número de FPs gerados

quando se tem 50 processos votantes é maior em relação a quando se tem apenas 10 processos. Na análise realizada, com 10 processos votantes já se consegue números significativos de detecção de falhas com baixo número de FPs.

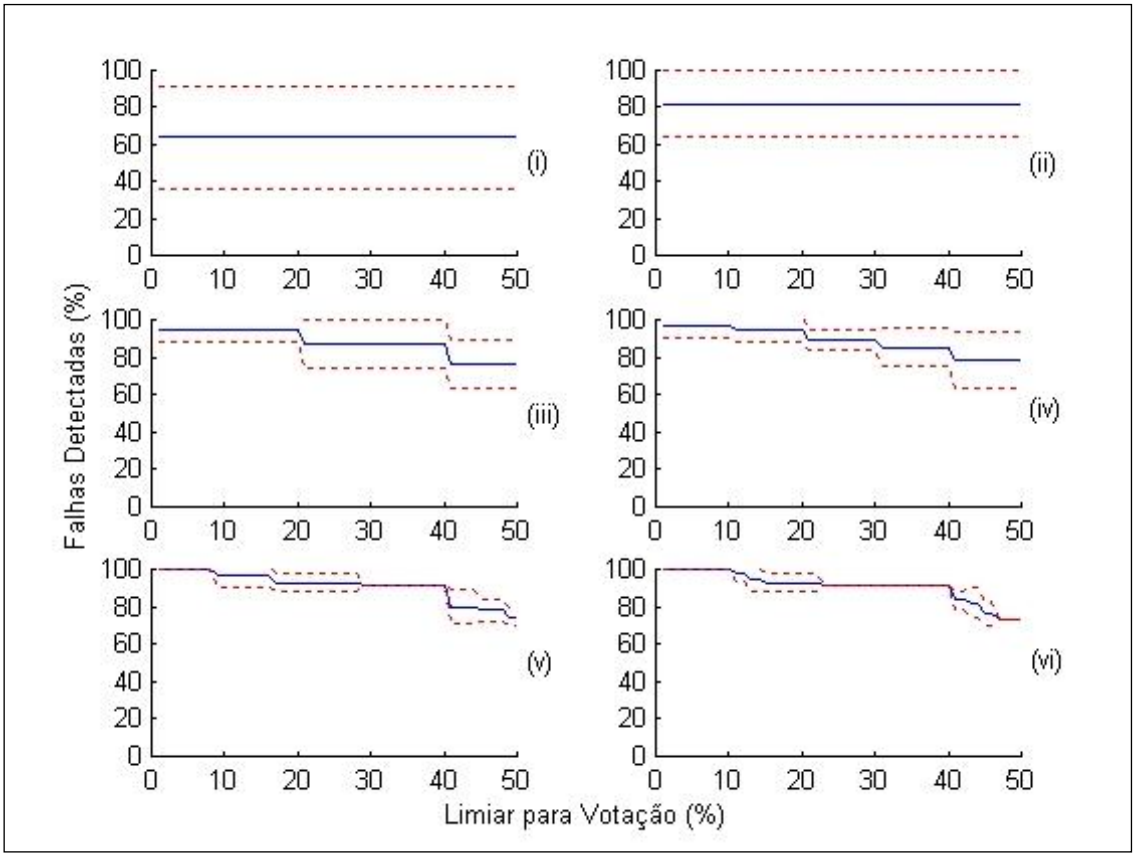


(a)

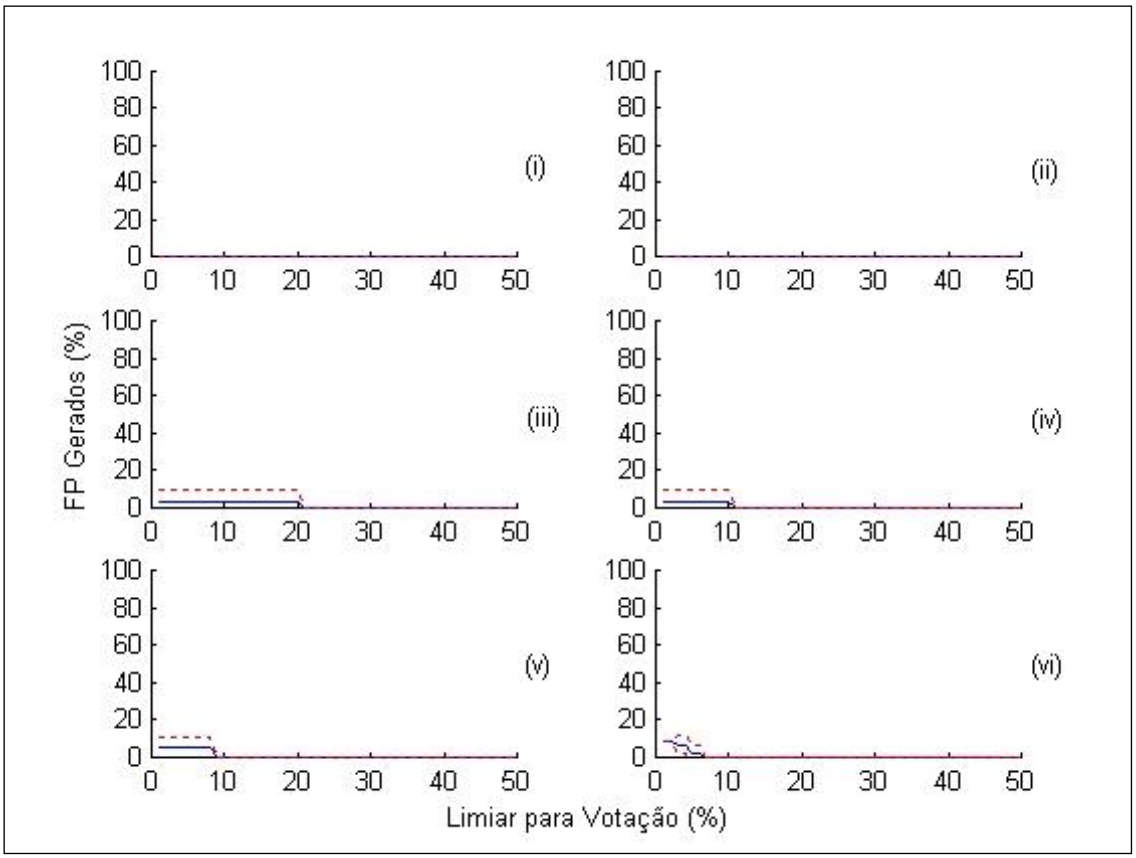


(b)

Figura 6.1 Comportamento médio do algoritmo no teste realizado com a extensão da *zona de perigo* (*td*) igual a 15 chamadas e o *número de processos votantes* igual a 1, 2, 5, 10, 25 e 50. Tanto em (a) quanto em (b), os gráficos representam os seguintes números de processos votantes: (i) 1 processo; (ii) 2 processos; (iii) 5 processos; (iv) 10 processos; (v) 25 processos e (vi) 50 processos. Os gráficos de (a) apresentam o *percentual de falhas detectadas* (0 – 100%) em função do *limiar para votação* (0 – 50%). Os gráficos de (b) apresentam o *percentual de falsos positivos gerados* (0 – 100%) em função do *limiar para votação* (0 – 50). Linha contínua: valor médio obtido em 5 execuções do algoritmo. Linhas pontilhadas: intervalo de confiança de 95%.

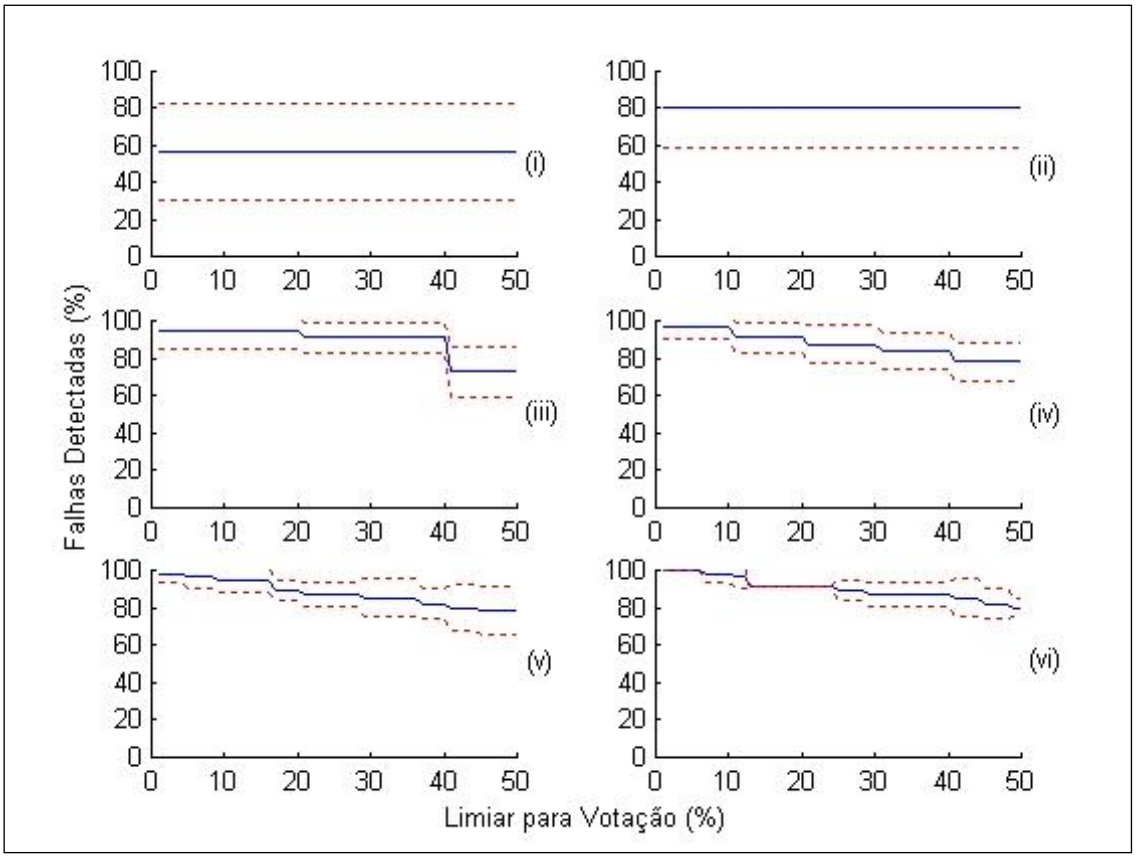


(a)

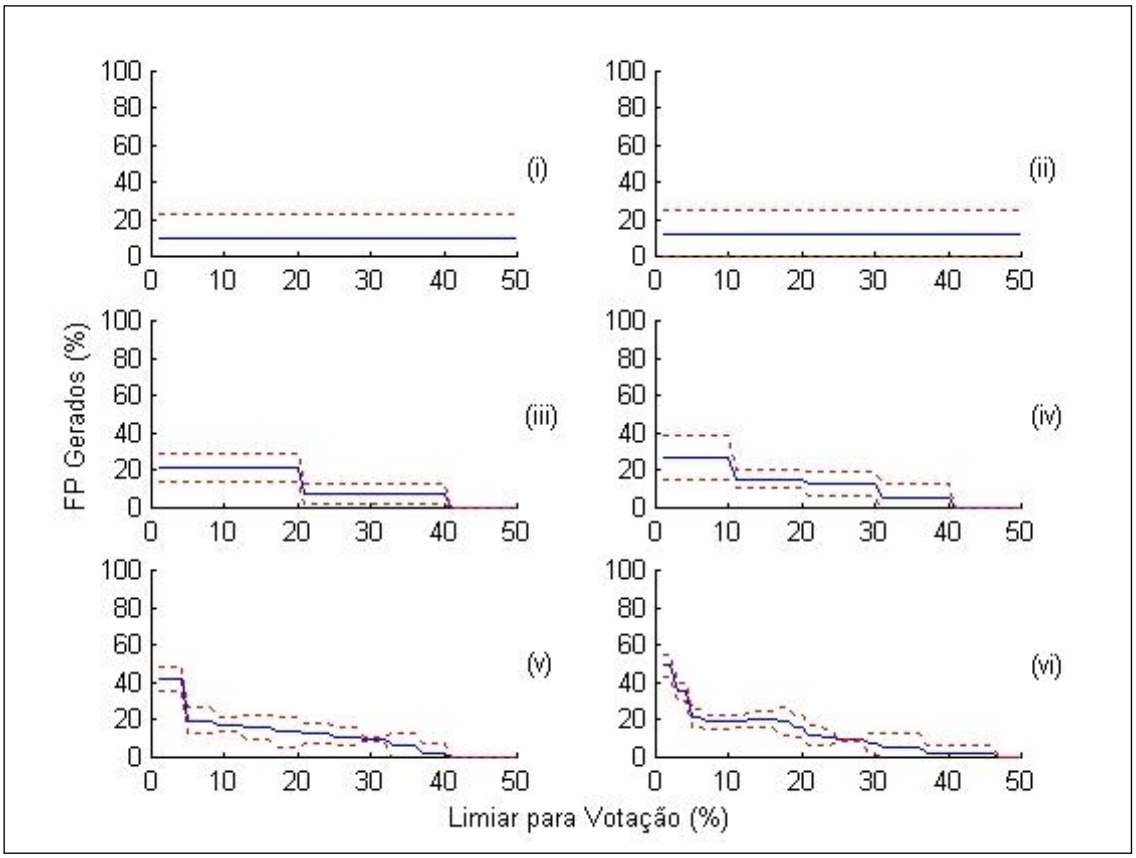


(b)

Figura 6.2 Comportamento médio do algoritmo no teste realizado com a extensão da *zona de perigo* (*td*) igual a 20 chamadas e o *número de processos votantes* igual a 1, 2, 5, 10, 25 e 50. Tanto em (a) quanto em (b), os gráficos representam os seguintes números de processos votantes: (i) 1 processo; (ii) 2 processos; (iii) 5 processos; (iv) 10 processos; (v) 25 processos e (vi) 50 processos. Os gráficos de (a) apresentam o *percentual de falhas detectadas* (0 – 100%) em função do *limiar para votação* (0 – 50%). Os gráficos de (b) apresentam o *percentual de falsos positivos gerados* (0 – 100%) em função do *limiar para votação* (0 – 50). Linha contínua: valor médio obtido em 5 execuções do algoritmo. Linhas pontilhadas: intervalo de confiança de 95%.



(a)



(b)

Figura 6.3 Comportamento médio do algoritmo no teste realizado com a extensão da *zona de perigo* (*td*) igual a 60 chamadas e o *número de processos votantes* igual a 1, 2, 5, 10, 25 e 50. Tanto em (a) quanto em (b), os gráficos representam os seguintes números de processos votantes: (i) 1 processo; (ii) 2 processos; (iii) 5 processos; (iv) 10 processos; (v) 25 processos e (vi) 50 processos. Os gráficos de (a) apresentam o *percentual de falhas detectadas* (0 – 100%) em função do *limiar para votação* (0 – 50%). Os gráficos de (b) apresentam o *percentual de falsos positivos gerados* (0 – 100%) em função do *limiar para votação* (0 – 50). Linha contínua: valor médio obtido em 5 execuções do algoritmo. Linhas pontilhadas: intervalo de confiança de 95%.

Ainda com respeito à rotina de votação, observando-se o limiar para votação nota-se que a consideração da maioria dos processos votantes como limiar para a detecção de falhas não é razoável. O limiar para votação deve ser baixo. No presente caso, menos de 25% dos votos devem ser considerados como suficientes para que uma falha seja detectada.

6.3 Adaptabilidade, Confiabilidade no Diagnóstico e Comportamento Geral

Para ilustrar a adaptação dos detectores ao longo do processo, foi observada a distribuição de antígenos e detectores de acordo com o parâmetro duração, no experimento realizado.

Foram utilizados os valores dos parâmetros considerados na seção 6.1, e definidos ainda os seguintes parâmetros:

- *td* (*zona de perigo*): 20 tentativas de chamadas (o tempo médio de 20 tentativas de chamadas)
- *número de processos votantes*: 50 processos
- *limiar para votação*: 25% (o total de votos deve exceder esse limiar para a tentativa de chamada ser considerada uma falha)

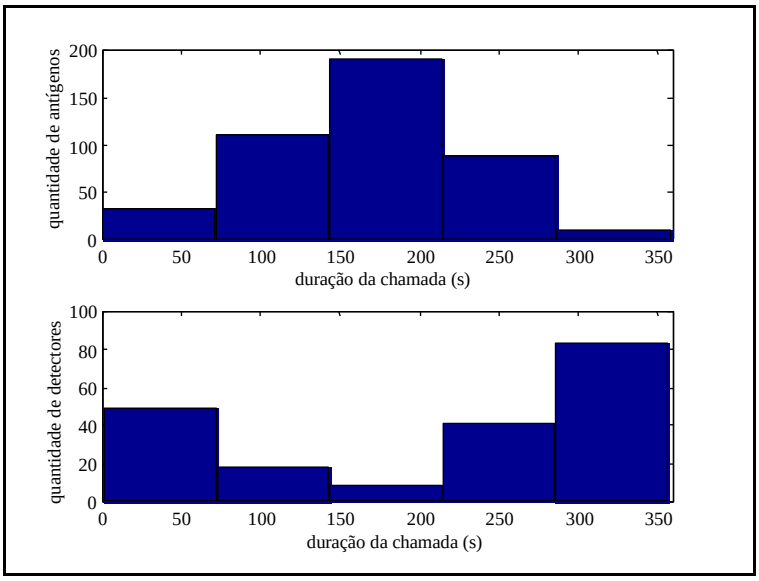


Figura 6.4 Distribuição de antígenos e detectores segundo o parâmetro *duração*. População de antígenos no intervalo de chamadas 71 a 500 e população de detectores ao final das 500 chamadas.

A Figura 6.4 ilustra a emergência do comportamento esperado. Observando-se a variável *duração*, foi possível notar esses detectores se acomodando no espaço complementar ao das chamadas geradas. Conforme descrito na seção 5.2.6, a geração constante de detectores baseia-se em um processo de seleção negativa. Com isso busca-se

ocupar o espaço complementar ao dos antígenos (chamadas normais) com os detectores (elementos que descrevem o comportamento de anormalidade do sistema).

Nas Figuras 6.5 e 6.6, é possível observar o melhor caso encontrado entre todas as execuções realizadas do algoritmo, analisado cada processo votante isoladamente. As Figuras 6.5 e 6.6 representam o mesmo caso de detecção executado. Porém, para uma melhor visualização da região de falhas, a Figura 6.6 apresenta somente as chamadas dessa região (seqüência de chamadas entre 400 e 500). Essas figuras apresentam as chamadas não completadas dentre as 500 chamadas geradas (gráfico (a) das Figuras 6.5 e 6.6), os alarmes gerados por excesso de não completamentos (sinal 2) (gráfico (b) das Figuras 6.5 e 6.6), a ativação de detectores e detecção de falha (sinal 1 + sinal 2) (gráfico (c) das Figuras 6.5 e 6.6) e a detecção de falha após ativação do detector (sinal 1 recebido com o detector ativo) (gráfico (d) das Figuras 6.5 e 6.6). O gráfico (b) das Figuras 6.5 e 6.6, como mencionado, indicam os alarmes gerados por excesso de não completamento de chamadas, sendo esse um processo padrão de detecção de falhas utilizado pelas redes telefônicas. O conjunto de todas as falhas detectadas pelo algoritmo, tanto na Figura 6.5 quanto na Figura 6.6, é obtido pela sobreposição dos gráficos (c) e (d). As falhas possíveis de serem detectadas são todas as falhas mostradas no gráfico (a) dessas figuras (não completamentos a partir da chamada de número 400) a partir do primeiro alarme indicado no gráfico (b). Dessa forma, pode-se observar, nesse exemplo, que, a partir da primeira falha detectável, não se pode detectar todas as falhas pela simples observação da taxa de não completamento de chamadas (gráfico (b) de ambas as figuras), porém pode-se detectá-las pela utilização do algoritmo (sobreposição dos gráficos (c) e (d) de ambas as figuras).

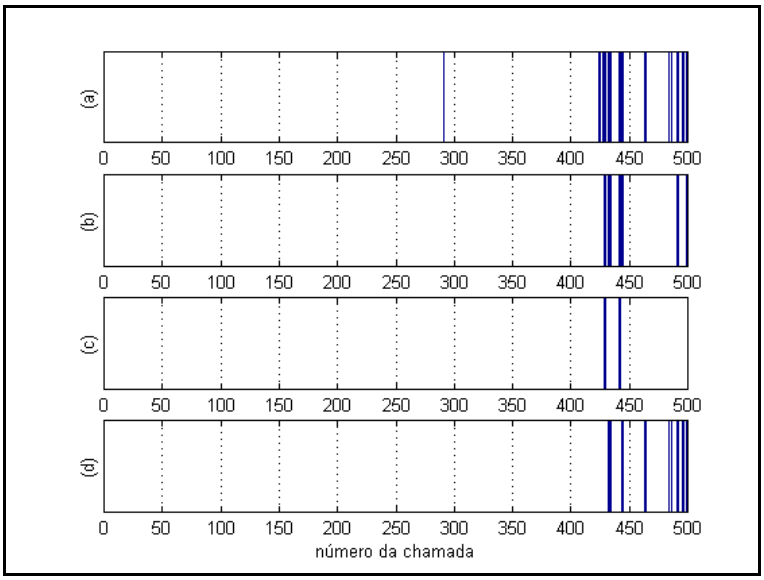


Figura 6.5 Seqüência de chamadas de 0 a 500: (a) chamadas não completadas; (b) alarmes (sinal 2); (c) ativação de detectores e geração de sinal de falha (sinal 1 + sinal 2); (d) geração de sinal de falha (sinal 1 com detector ativo).

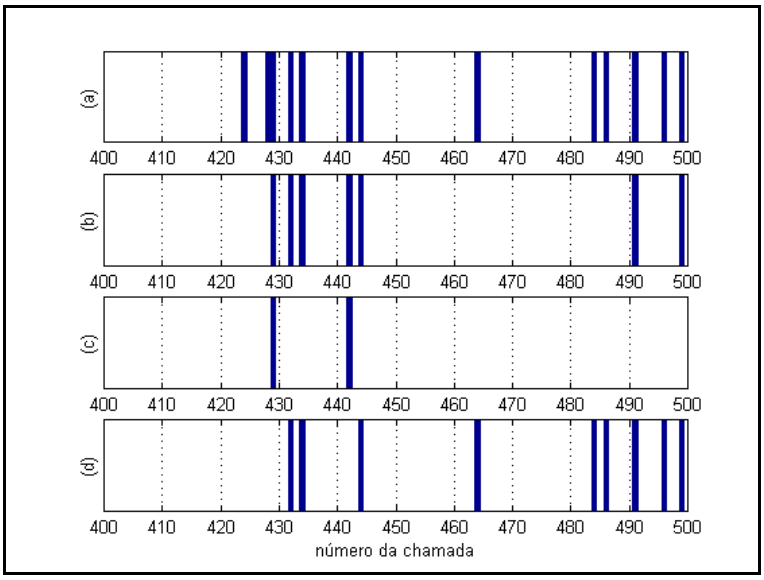


Figura 6.6 Seqüência de chamadas de 400 a 500 da Figura 6.5: (a) chamadas não completadas; (b) alarmes (sinal 2); (c) ativação de detectores e geração de sinal de falha (sinal 1 + sinal 2); (d) geração de sinal de falha (sinal 1 com detector ativo).

Pode-se notar que a chamada de número 291 (Figura 6.5), embora fosse uma anormalidade (não era uma falha), em nenhum dos 50 processos votantes foi detectada como uma falha, devido à falta do sinal 2. Isso era outro comportamento esperado.

A Tabela 6.1 mostra os resultados para os valores de variáveis utilizados aqui nesta seção. Conforme os dados apresentados, o algoritmo detectou 91% do total de falhas. Dentre as falhas detectáveis há aquelas que já são indicadas pelo sinal 2 (que pode ser um alarme qualquer vinculado à chamada), porém há aquelas que o sinal 2 não consegue detectar (ver Figuras 6.5 e 6.6). Nesse ensaio, 100% das falhas não detectáveis pelo sinal 2 foram detectadas pelo algoritmo. Outro resultado significativo, conforme indicado na terceira linha da Tabela 6.1, foi que em todos os 50 processos votantes houve sempre alguma falha não detectável pelo alarme 2 e detectada pelo algoritmo. Isso significa que, mesmo sem considerar o processo de votação, o algoritmo, nesse ensaio, sempre detectou alguma falha não detectável pelo sinal 2. Para uma melhor compreensão desse último resultado, deve-se enfatizar que, mesmo o algoritmo detectando 100% das falhas não detectáveis pelo sinal 2, por se tratar de um processo de votação, existe a possibilidade de alguns processos votantes não detectarem falha alguma. No entanto, isso não ocorreu. Como mencionado, todos os processos votantes detectaram ao menos 1 falha não detectável pelo sinal 2. Por fim, a taxa de falsos positivos foi 0%, completando o comportamento desejado.

Tabela 6.1 Resultados encontrados no teste com o algoritmo proposto considerando 50 processos votantes. Resultados consideram as falhas possíveis de detecção (falhas após o primeiro alarme – sinal 2).

Percentual de falhas não detectáveis pelo alarme (sinal 2), mas detectada pelo algoritmo	100 %
Percentual total de falhas detectadas	91,0 %
Percentual de processos votantes que tiveram alguma falha, não detectável pelo alarme (sinal 2), detectada pelo algoritmo	100 %
Percentual de falsos positivos	0%

Como resultado adicional, buscou-se observar o comportamento dos detectores ativados para a detecção das falhas. Foi gerado assim um gráfico (Figura 6.7) que indica o grau de cobertura gerado por esses detectores. Cada processo que participa da votação na busca pelas falhas ativa um grupo de detectores. Cada um desses detectores possui uma região de atuação (região de reconhecimento) onde reconhece os antígenos. O gráfico

apresentado é formado pela soma das áreas cobertas por esses detectores. A área de observação é no plano origem pelo destino, considerando a duração da chamada igual a zero e a funcionalidade igual a 2 (situação de falha). Nesse gráfico, as estrelas correspondem às falhas. A elevação da superfície corresponde ao grau de cobertura, i.e., quanto mais elevada for a superfície, maior a incidência de detectores ativos na região do plano origem-destino. O comportamento do algoritmo corresponde ao esperado, i.e., a região que os detectores ativos buscam cobrir é a de maior concentração de falhas, sendo que as regiões que não devem ser cobertas (não possuem falhas), pelo fato de poderem gerar FPs, possuem baixa taxa de cobertura.

Deve ser ressaltado também que se encontra aqui uma forma de se obter qualitativamente uma representação do nível de confiança no resultado diagnosticado. Quanto mais próximo do pico, maior a incidência de detectores ativos. Logo, maior é a possibilidade de se ter um diagnóstico de falha correto.

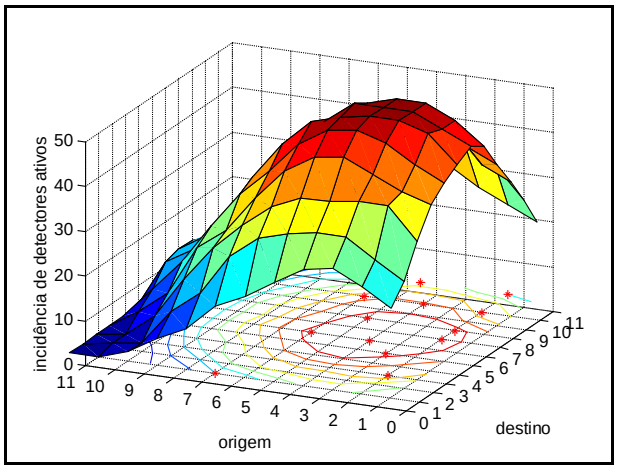


Figura 6.7 Comportamento médio do algoritmo de detecção de falhas no teste com uma região de falha. Estrelas correspondem às falhas. Quanto mais elevada a superfície, maior a incidência de detectores ativos na região.

6.4 Detecção de Múltiplas Falhas e Isolabilidade

Para se verificar ainda a possibilidade de se ter uma separação entre grupos de tentativas de chamadas cujas falhas tenham sido originadas por motivos diferentes, fez-se um outro teste onde foram geradas duas regiões de falhas distintas no espaço de parâmetros das tentativas

de chamadas. Traçou-se um gráfico do nível de cobertura dos detectores, como mencionado na seção anterior, e observou-se, nesse, a distinção das regiões pela formação de dois picos (Figura 6.8). As regiões que tiveram alta incidência de tentativas de chamadas, suficiente para disparar o sinal 1 de um detector, e que se acomodaram dentro de uma zona de perigo, foram cobertas pelos picos e podem ser identificadas na Figura 6.8.

É importante notar que existem regiões na Figura 6.8 que, embora possuam uma concentração de chamadas com duração igual a zero, não foram cobertas pelo algoritmo. Isso se deu pelo fato dessas tentativas de chamadas estarem dispersas, afastadas temporalmente umas das outras. Dessa forma, essas tentativas de chamadas, com características semelhantes, não ocorreram em tempo próximo o suficiente umas das outras para, somadas, dispararem um sinal 1. Elas são assim vistas pelo algoritmo como um comportamento normal. Este é um ponto importante, pois mostra a capacidade do algoritmo em diferenciar chamadas que, embora possuam características semelhantes, possuem diagnósticos diferentes.

Podemos ver assim claramente pelas Figuras 6.7 e 6.8 a possibilidade de se buscar a causa das falhas, mesmo na existência de múltiplas fontes de falhas.

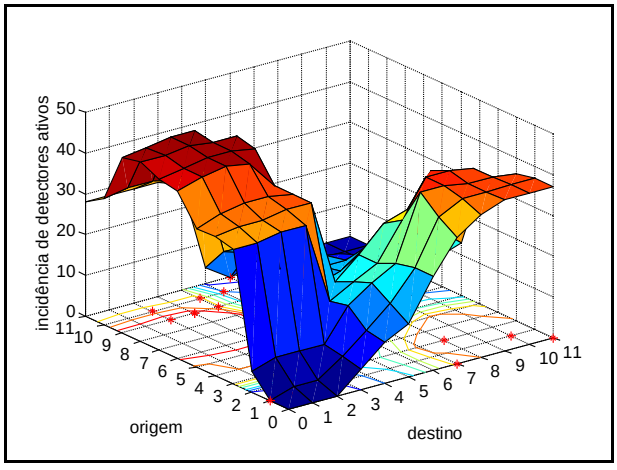


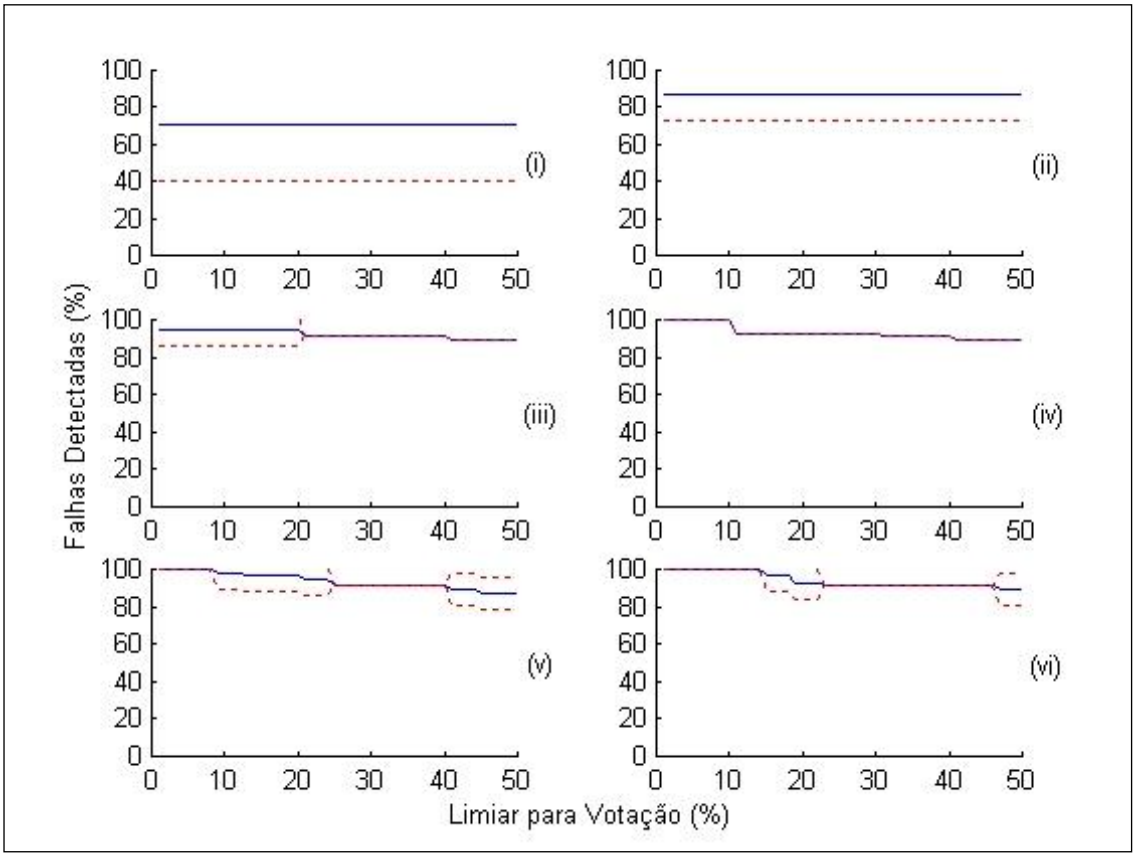
Figura 6.8 Comportamento médio do algoritmo de detecção de falhas no teste com duas regiões de falha. Estrelas correspondem às falhas. Quanto mais elevada a superfície, maior a incidência de detectores ativos na região.

6.5 Não Utilização do Sinal de Perigo

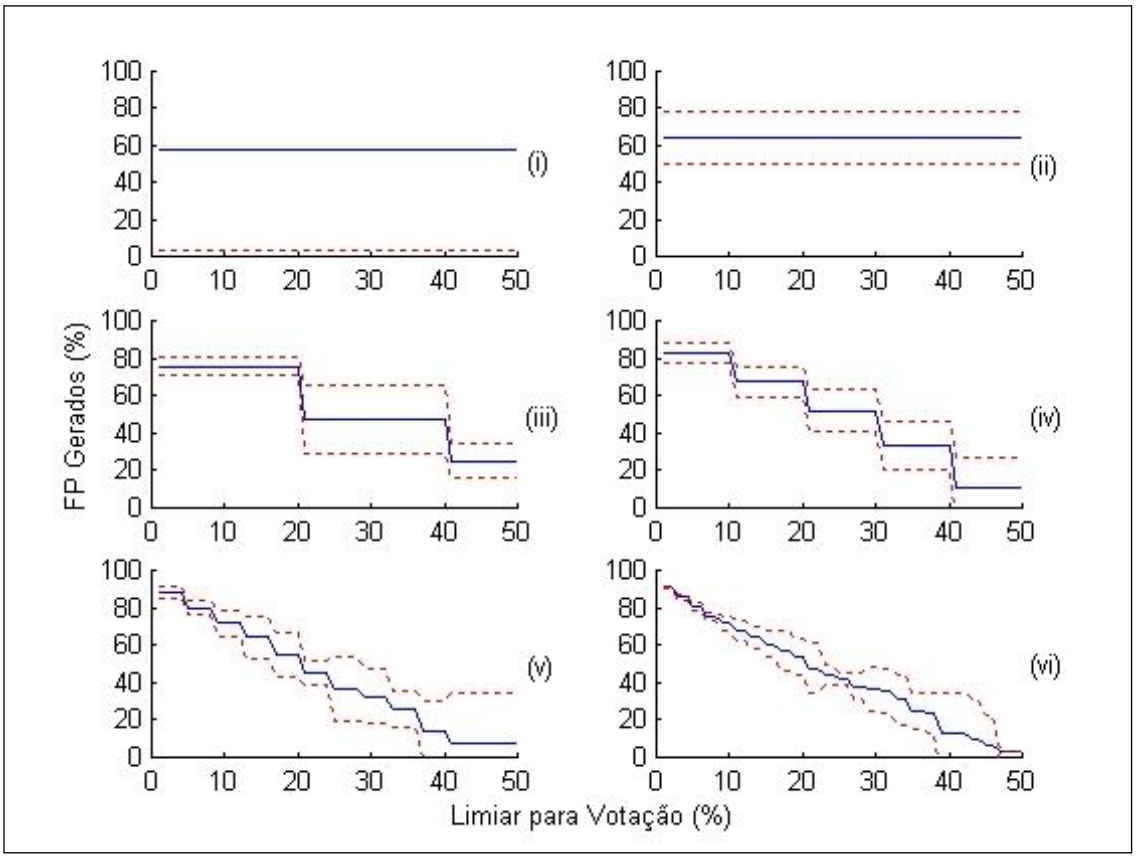
Visando se obter uma comparação qualitativa entre a Teoria do Perigo e a visão Próprio/Não-Próprio, realizou-se uma emulação utilizando os mesmos valores de parâmetros mencionados nas seções 6.1 e 6.2, com a exceção do parâmetro *td* (zona de perigo) que foi desconsiderado. Foi utilizada a mesma estrutura do algoritmo implementado para a Teoria do Perigo, porém foi considerado o sinal 2 como estando sempre ativo fazendo assim o sinal 1, ou seja, a detecção das anormalidades, como a única fonte de verificação de falhas. Com o intuito de não se criar um número muito elevado de falsos positivos, foi criada uma zona de observação temporal dentro da qual o limiar para ativação do detector (*LCaff*) deveria ser alcançado visando a ativação do detector. A partir do primeiro reconhecimento de um antígeno, um detector teria essa zona de observação temporal para conseguir alcançar o limiar *LCaff*. Caso contrário, os reconhecimentos realizados dentro dessa zona temporal não seriam considerados para sua ativação. Assim, como na Teoria do Perigo, nessa implementação Próprio/Não-Próprio tanto a ativação de um detector quanto o reconhecimento, por esse detector, de qualquer antígeno após sua ativação são considerados como falhas.

Como pode se perceber na prática, a zona de observação temporal, considerada nessa emulação Próprio/Não-Próprio, equivaleria à própria zona de perigo do algoritmo proposto no capítulo 5. Nessa zona de perigo, é necessário que um detector reconheça ao menos *LCaff* antígenos e ainda receba o sinal 2 para que possa ser ativado Porém, na emulação Próprio/Não-Próprio feita, o sinal 2 foi considerado como sendo sempre existente.

O valor utilizado para a zona de observação temporal foi a de aproximadamente 20 chamadas, possibilitando assim uma comparação qualitativa com a Figura 6.2 da Teoria do Perigo. Os resultados desse ensaio estão mostrados na Figura 6.9. Assim como as Figuras 6.1 a 6.3, a Figura 6.9 apresenta o valor médio para 5 execuções do algoritmo (linha contínua nos gráficos) e o intervalo de confiança (ver Miller & Freund, 1985) de 95% (linhas pontilhadas nos gráficos).



(a)



(b)

Figura 6.9 Comportamento médio do algoritmo na emulação do Modelo Próprio/Não-Próprio. Zona de Observação Temporal com extensão igual a 20 chamadas e o *número de processos votantes* igual a 1, 2, 5, 10, 25 e 50. Tanto em (a) quanto em (b), os gráficos representam os seguintes números de processos votantes: (i) 1 processo; (ii) 2 processos; (iii) 5 processos; (iv) 10 processos; (v) 25 processos e (vi) 50 processos. Os gráficos de (a) apresentam o *percentual de falhas detectadas* (0 – 100%) em função do *limiar para votação* (0 – 50%). Os gráficos de (b) apresentam o *percentual de falsos positivos gerados* (0 – 100%) em função do *limiar para votação* (0 – 50). Linha contínua: valor médio obtido em 5 execuções do algoritmo. Linhas pontilhadas: intervalo de confiança de 95%.

Os resultados obtidos na Figura 6.9 indicam que, embora o Modelo Próprio/Não-Próprio possa apresentar um número elevado de detecções de falhas, o mesmo pode apresentar um número bastante elevado de falsos positivos (comparar Figuras 6.2 e 6.9). Isso se dá especialmente pelo perfil do sistema analisado, ou seja, um sistema que muda de comportamento ao longo do tempo. O número de falsos positivos elevados nesse ensaio se deu especialmente pela geração de vários falsos positivos antes da chamada de número 400 (chamada a partir da qual se iniciou a situação de falha) pelo fato de se associar uma simples mudança de comportamento a uma falha de fato (ver perfil das chamadas geradas na seção 6.1). Uma vez que o comportamento do sistema seja estável, ou ao menos se estabilize, o comportamento do Modelo Próprio/Não-Próprio tende a melhorar significativamente, diminuindo o número de falsos positivos. Deve-se, contudo, notar que a estabilização mencionada não deve implicar em uma mudança abrupta de estados, pois isso seria detectado também como uma falha, utilizando a visão Próprio/Não-Próprio em questão. Um outro ponto observado é a possibilidade de se diminuir o número de falsos positivos, na visão Próprio/Não-Próprio, utilizando o processo de votação. Contudo, em comparação com a visão do Perigo, faz-se necessária a utilização de um número de processos votantes bem maior, aumentando assim o esforço computacional despendido.

Esse ensaio pode, assim, indicar um ganho com a utilização do sinal 2 na diminuição do número de falsos positivos frente a uma mudança de comportamento do sistema analisado.

CAPÍTULO 7

Conclusões e Trabalhos Futuros

Neste capítulo, baseando-se nos resultados obtidos, são levantados os aspectos positivos do uso da Teoria do Perigo em sistemas de detecção de falhas. São listadas as principais características observadas no algoritmo e, por último, são propostas perspectivas futuras que visam dar continuidade à pesquisa.

7.1 Posicionamento da Pesquisa e Visão Geral do Capítulo

Esta dissertação teve como objetivo a implementação de um algoritmo de detecção de falhas com características de adaptabilidade e autonomia capazes de produzir bons índices de detecção e baixos índices de falsos positivos (FPs).

Buscou-se estudar o modelo imunológico apresentado por Polly Matzinger (Matzinger, 1994, 1998, 2001, 2002), a Teoria do Perigo, entendendo suas peculiaridades na interpretação de fenômenos imunológicos frente a outras iniciativas de obtenção de sistemas imunológicos artificiais. Este capítulo busca evidenciar as características do algoritmo implementado, destacando as propriedades que um sistema de detecção de falhas deve ter e que estão presentes nessa implementação. E, por último, são sugeridos estudos futuros que darão continuidade ao trabalho aqui realizado.

7.2 Características Observadas

Pelo fato da população de detectores ser constantemente atualizada pelo uso do procedimento de seleção clonal, mesmo diante de um ambiente dinâmico foram obtidos altos índices de reconhecimento e diversidade.

Devido à utilização do paradigma da Teoria do Perigo, em que um co-estímulo é necessário para confirmar o perigo causado pela presença de um antígeno detectado, anomalias externas fora do alcance do sinal 2 não foram, como esperado, consideradas como falhas. Um resultado significativo foi o percentual de cobertura na detecção de falhas com respeito às regiões que não foram cobertas pelo alarme (sinal 2). Essa cobertura se mostrou bastante significativa, pois indica a possibilidade da continuidade de uma falha ou a tendência à retomada da mesma. Sendo essa falha apresentada de forma esparsa, não permite que o sistema observado gere um alarme (é necessário alcançar uma taxa específica de não completamentos para que seja gerado o alarme – sinal 2). Assim, o uso do algoritmo nesse contexto fornece um importante complemento na detecção das falhas.

Como almejado, constatou-se que o número de FPs foi bem reduzido e o percentual de falhas detectadas se mostrou bastante significativo, embora não se tenha conseguido detectar a totalidade das falhas, em média.

Na seção 6.5, em que o sinal 2 foi desconsiderado, buscando-se assim obter uma visão próxima do Modelo Próprio/Não-Próprio, pôde-se observar um incremento no número de FPs. Isso especialmente pela mudança de comportamento do sistema sob análise. A presença do sinal 2 mostrou-se bastante significativa para a diminuição do número de FPs, apontando dessa forma para um ganho da visão proporcionada pela Teoria do Perigo sobre a visão Próprio/Não-Próprio, quando da detecção de falhas em sistemas com comportamento dinâmico.

O algoritmo implementado pode ser usado na identificação de regiões (clusters) de alta concentração de detectores ativos, os quais permitem especificar as características que resultam em falhas, i.e., configuração do conjunto de parâmetros que originam uma falha. Dessa forma, pode-se notar que o algoritmo não se limitou aos pontos de detecção de falhas almejados, mas mostrou claramente características complementares de diagnóstico de falhas típicas de um processo de FDD automático (*Automated Fault Detection and Diagnosis*).

As características próprias de um processo de FDD que podem ser destacadas como propriedades do algoritmo implementado são:

Detecção Rápida de Falhas

Uma vez que detectores já tenham sido ativados, para a detecção de uma falha basta que seja verificada a similaridade do padrão de entrada com esses detectores, gerando dessa forma uma detecção ágil de falhas.

Robustez

Devido ao fato do algoritmo trabalhar com detecção imperfeita de antígenos, ou seja, possuir uma região de reconhecimento ao redor de cada detector, não exigindo assim um casamento preciso entre detector e antígeno para iniciar o reconhecimento de uma falha, ele se torna robusto a pequenas variações comportamentais.

A rotina de votação fez com que a robustez do algoritmo aumentasse. Pequenos erros acarretados por variações na entrada se tornam mais difíceis de serem aceitos como falha quando se necessita de um consenso entre os diferentes processos que analisam a mesma entrada.

Tem-se ainda que a necessidade do sinal 2 faz com que as alterações comportamentais na entrada do sistema de análise de falhas sejam distinguidas entre anomalias ou falhas, aumentando a robustez do algoritmo.

Adaptabilidade

Uma das qualidades principais do algoritmo em questão é sua adaptabilidade a diferentes situações de normalidade e anormalidade das tentativas de chamadas analisadas. O fato da população de detectores estar em constante adaptação, combinado com a necessidade do sinal 2 para confirmar a ocorrência de uma falha, torna o algoritmo adaptável a situações comportamentais dinâmicas (ver Figura 6.4). Essa propriedade, constatada nos testes realizados, confirmou o que já se esperava do algoritmo.

Conforme mencionado em Venkatasubramanian *et al*. (2003c) e Katipamula & Brambley (2005), uma das deficiências dos métodos baseados na história do processo são suas limitações com respeito à adaptabilidade. Embora o algoritmo proposto aqui seja um método baseado na história do processo, ele ainda assim apresenta um alto nível de adaptabilidade.

Isolabilidade

De acordo com o observado nas Figuras 6.7 e 6.8, a verificação de concentração de detectores ativos em uma dada região do espaço de variáveis analisado pode indicar a presença de múltiplas falhas, ou seja, de falhas que sejam geradas por razões distintas.

Capacidade de Indicar a Qualidade da Classificação

Por meio da análise da concentração de detectores ativos (Figuras 6.7 e 6.8) é possível se obter uma comparação entre os diagnósticos fornecidos para cada falha gerada. Quanto maior a porcentagem de diagnósticos positivos, mais confiável tende a ser a indicação de falha.

7.3 Trabalhos Futuros

A seguir, são apresentados passos que visam a continuidade desse trabalho.

Identificação de processos que poderiam ajustar automaticamente os limiares mais relevantes do algoritmo

Como notado na análise de sensibilidade mostrada no Capítulo 6, os resultados alcançáveis pelo algoritmo estão intimamente relacionados a alguns parâmetros, especialmente aos que se referem à zona de perigo e ao processo de votação. A possibilidade de o algoritmo determinar automaticamente com quais limiares deve trabalhar se mostra de grande valia para o seu desempenho.

Utilização de Abordagem Fuzzy para Descrição da Afinidade

Uma abordagem *fuzzy* baseada no conceito de degeneração (Atamas, 2005; Cohen *et al.*, 2004; Hafler, 2002) poderia ser empregada para descrição da afinidade. Usando essa abordagem, a diversidade ou o número de detectores poderiam ser reduzidos e ainda assim mantida num mesmo nível a cobertura do espaço antigênico. A ativação do detector poderia ser vista como o resultado de uma análise envolvendo a combinação da intensidade do sinal 1 obtido, i.e., o nível de afinidade com cada estrutura antigênica, e um nível de sinal 2 relacionado.

Utilização das Superfícies de Detectores Ativos no Isolamento, Busca da Origem e da Propagação de Falhas e na Verificação da Confiabilidade do Diagnóstico

Pode-se buscar, por meio da observação dinâmica das superfícies de detectores ativos, isolar as falhas obtendo-se os padrões que caracterizem suas origens e evoluções, bem como o grau de confiabilidade dos diagnósticos com base no nível de concentração dos detectores ativos no momento (as superfícies geradas, e.g. Figuras 6.7 e 6.8, modificam-se dinamicamente).

Utilização de Múltiplos Alarmes

Visando ampliar a eficiência e aplicabilidade do algoritmo proposto, pode-se buscar a utilização de alarmes combinados. Não somente a utilização de sinais estimuladores, mas também sinais com efeito supressor. A ponderação dos múltiplos sinais é que serviria como o sinal 2, e não somente a utilização de um único sinal. O emprego desses múltiplos sinais baseia-se no modelo das células dendríticas proposto por Greensmith *et al.* (2005).

Utilização de Endereços Hierarquizados

Nos experimentos propostos, as dimensões da rede telefônica, o número de chamadas, e o número de variáveis capazes de caracterizar o cenário completo são reduzidas, não correspondendo efetivamente à caracterização de uma aplicação real. Como o principal propósito aqui foi o de validar a proposta e verificar suas habilidades em expressar

adaptabilidade e operação adequada em termos de evitar falsos positivos, então os experimentos satisfazem plenamente os objetivos.

Um dos pontos, no entanto, que deve ser abordado na continuidade desse projeto é a tentativa de tornar a análise realizada mais próxima de uma situação real. Para tanto, uma dimensão maior da rede se torna necessária, aumentando assim o esforço computacional para a detecção das falhas. Como forma de contornar essa situação, pode-se fazer uma análise hierarquizada dos endereços, contanto que os endereços utilizados também sejam hierarquizados.

Como exemplo, suponha que uma rede possua 30 endereços de origem e 30 de destino, sendo esses endereços dados por 2 dígitos, e.g., 00, 01, 02, 10, 11 , 12, ..., 90, 91, 92. O primeiro dígito corresponderia a macro-regiões da rede, regiões com uma hierarquia maior. O segundo dígito corresponderia a sub-regiões das macro-regiões mencionadas, ou seja, regiões com uma hierarquia menor. Por exemplo, os endereços 10, 11 e 12 indicariam as sub-regiões 0, 1 e 2 da macro região 1. O algoritmo deve ser capaz de iniciar a busca pelos endereços de maior hierarquia na rede. Uma vez detectada a região de falha, ou seja, a região indicada pela maior concentração de detectores ativos, passa-se então a uma segmentação da região de análise, sendo então feita a continuação da análise em um nível de endereçamento hierarquicamente inferior, conforme indicado na Figura 7.1.

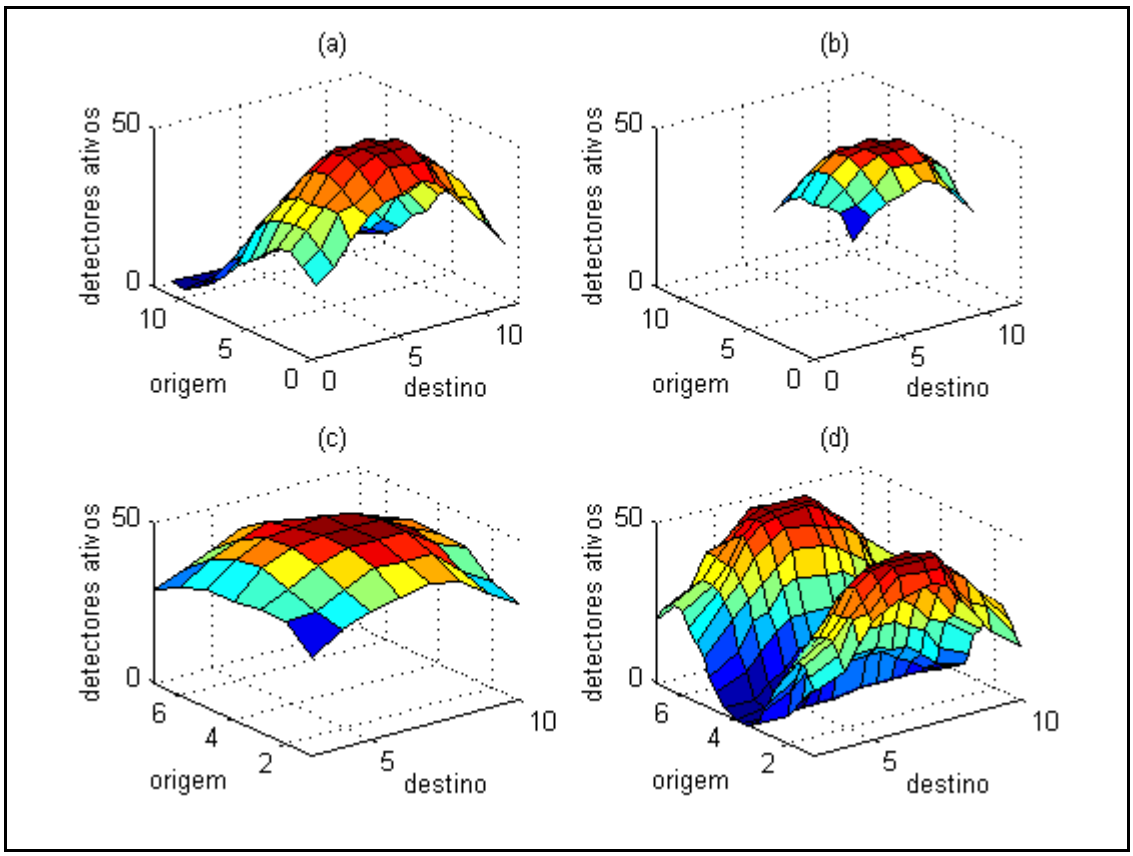


Figura 7.1 Utilização de endereços hierarquizados na detecção de falhas. (a) detecção de região de falhas em um plano de hierarquia superior; (b) segmentação da região com maior incidência de detectores ativos, no plano de hierarquia superior; (c) região segmentada, visualizada dentro do plano de análise com hierarquia inferior, antes da análise nesse plano; (d) região segmentada, visualizada dentro do plano de análise com hierarquia inferior, depois da análise nesse plano.

Utilização da Teoria do Perigo em Análise de Difusão de Inovações

Por meio de uma nova metáfora, a da contaminação de células (células atingidas por um patógeno poderiam ser necrosadas e esse dano propagado), ou ainda a do nível de detectores ativados, pode-se obter um modelo mercadológico para o estudo de demanda.

Atualmente, por meio da abordagem ABMS (*Agent-Based Modelling Simulation*) é possível simular processos de difusão dentro de um mercado específico (Ávila *et al.*, 2006). Para tanto, são utilizadas curvas comportamentais, como as sugeridas por Rogers (1983), e outras características consideradas como sendo próprias do perfil de consumidores dentro de suas respectivas classes econômicas. São gerados diversos agentes que interagem e

fornecem o comportamento de difusão analisado. Essa análise, diferentemente do uso de dinâmica de sistemas, possui um perfil *bottom-up*.

A utilização de SIAs pode trazer a análise de difusão para mais próximo da realidade. Os dados referentes ao perfil de consumo de uma sociedade, ou grupo social, obtidos experimentalmente, podem ser mapeados dentro de um sistema imunológico. A reação desse sistema frente a inovações, seja por meio de uma metáfora de contaminação, ou de nível de detectores ativados, pode ilustrar a demanda que se deseja analisar.

A presença de sinais de perigo, emitidos por células contaminadas, pode ser entendida como a influência que cada indivíduo exerce sobre o outro em uma sociedade. Para esse tipo de análise, pode ser necessário trabalhar com diferentes tipos de células artificiais análogas às constituintes do sistema imunológico, coerente com o paradigma que fundamentou a proposta deste trabalho.

Bibliografia

Aickelin, U., Bentley, P., Cayzer, S., Kim, J. & McLeod, J. (2003). "Danger Theory: The Link between AIS and IDS?", 2nd International Conference on AIS (ICARIS 2003), pp. 147-155.

Aickelin, U., Greensmith, J., Twycross, J. (2004). "Immune System Approaches to Intrusion Detection – A Review", 3rd International Conference on AIS (ICARIS 2004), pp. 316-329.

Anchor, K. P., Zydallis, J. B., Gunsch, G. H., Lamont, G. B. (2002). "Extending the Computer Defence Immune System: Network Intrusion Detection with a Multiobjective Evolutionary Programming Approach", 1st International Conference on AIS (ICARIS 2002).

Atamas, S.P. (2005). "Les affinities electives", Dossier Pour La Science, 46.

Ávila, I. M. A., Pinto, J. C. L., Lemos, L. M., Holanda, G. M. (2006). "Redes sociais e disseminação de inovações tecnológicas: uma modelagem por agentes aplicada ao caso da TV digital brasileira", Revista CPqD (a ser publicado).

Ayara, M., Timmis, J., De Lemos, R., de Castro, L.N. & Duncan, R. (2002). "Negative Selection: How to Generate Detectors", 1st International Conference on AIS (ICARIS 2002), pp 89-98.

Balthrop, J.,Esponda, F., Forrest, S., Glickman, M. (2002a). "Coverage and generalization in an artificial immune system", Proceedings of GECCO, pp. 3–10.

Balthrop, J., Forrest, S., Glickman, M. (2002b). "Revisiting lisys: Parameters and normal behaviour", Proceedings of the Congress on Evolutionary Computation, pp. 1045–1050.

Banchereau, J., Briere, F., Caux, C., Davoust, J., Lebecque, S., Liu, Y.-J., Pulendran, B., Palucka, K. (2000). "Immunobiology of dendritic cells", *Annu. Rev. Immunol.*, 18, pp. 767–811.

Bauer, E. & Kohavi, R. (1999). "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants", Machine Learning, vol.36, no.1-2, pp.105-139, July-August 1999.

Bentley, P. J., Greensmith, J., Ujjin, S. (2005). "Two Ways to Grow Tissue for Artificial Immune Systems", 4th International Conference on AIS (ICARIS 2005), pp. 139-152.

Billingham, R. E., Brent, L., Medawar P. B. (1953). "Actively Acquired Tolerance of Foreign Cells", Nature, vol. 172, pp. 603-6.

Boudec, J. & Sarafijanovic, S. (2003). "An artificial immune system approach to misbehavior detection in mobile ad-hoc networks", Technical Report IC/2003/59, Ecole Polytechnique Federale de Lausanne.

Bretscher P., Cohn M. (1970). "A Theory of Self-Nonself Discrimination", Science, vol. 169, pp. 1042-9.

Burnet, F. M. (1959). "The Clonal Selection Theory of Acquired Immunity", Vanderbilt University Press, Nashville, TN.

Canham, R. O., Tyrrell, A. M. (2002). "A Multilayered Immune System for Hardware Faut Tolerance Within an Embryonic Array", 1st International Conference on AIS (ICARIS 2002).

Cohen, I. R., Hershberg, U., Solomon, S. (2004). "Antigen-Receptor Degeneracy and Immunological Paradigms", Molecular Immunology, vol. 40, pp. 993-6.

Costa Branco, P. J., Dente, A., Viela Mendes, R. (2003). "Using Immunology Principles fo Fault Detection", IEEE Transactions on Industrial Eletronics, 50, pp. 362-373.

Dasgupta, D. & Balachandran, S. (2006). "Artificial Immune Systems: A Bibliography", Technical Report CS-04-003, University of Memphis, Computer Science Division.

Dasgupta, D., KrishnaKumar, K., Wong, D., Berry, M. (2004). "Negative Selection Algorithm for Aircraft Fault Detection", 3rd International Conference on AIS (ICARIS 2004), pp. 1-13.

Dasgupta, D. & González, F. A. (2002). "An immunity-based technique to characterize intrusions in computer networks", IEEE Transactions on Evolutionary Computation, 6(3), pp. 281–291.

Dasgupta, D. (Ed.) (1998). *Artificial Immune Systems and Their Applications,* Springer-Verlag.

Dasgupta, D. & Forrest, S. (1996). "Novelty detection in time series data using ideas from immunology", 5th Int. Conf. Intelligent Systems, Reno, NV, June 19–21.

Davidson, J. & Peters, J. (2000). "Voice over IP Fundamentals". Cisco Press.

De Boer, R. J., Segel, L. A. & Perelson, A. S. (1992). "Pattern Formation in One- and Two-Dimensional Shape-Space Models of the Immune System", J. theor. Biol., 155, pp. 295-333.

De Castro, L. N (2001). "Engenharia Imunológica: Desenvolvimento e Aplicação de Ferramentas Computacionais Inspiradas em Sistemas Imunológicos Artificiais", Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, Unicamp.

De Castro, L.N & Timmis, J.I. (2002). Artificial Immune Systems: A New Computational Intelligence Approach, Springer-Verlag.

De Castro, L.N & Von Zuben, F.J. (2002). "Learning and Optimization Using the Clonal Selection Principle", IEEE Transactions on Evolutionary Computation, vol. 6, no. 3, pp. 239-251.

Doherty, P. C. & Christensen, J. P. (2000). "Accessing complexity: The dynamics of virus-specific T cell responses," *Annu. Rev. Immunol.*, 18, pp. 561–592.

Dutton, R. W. , Bradley, L. M., Swain, S. L. (1998). "T cell memory", *Annu. Rev. Immunol.*, 16, pp. 201–223.

Faynberg, I., Lawrence, G., Lu, H.-L. (2000). "Converged networks and services: Internetworking IP and the PSTN". New York: John Wiley & Sons.

Forrest, S. & A. Perelson (1992). "Computation and the Immune System", SIGBIO Newsletter, Association for Computing Machinery, 12(2), pp. 52-57.

Forrest, S., A. Perelson, Allen, L. & Cherukuri, R. (1994). "Self-Nonself Discrimination in a Computer", Proc. do IEEE Symposium on Research in Security and Privacy, pp. 202-212.

Freitas, A. A. & Rocha, B (2000). "Population biology of lymphocites: The flight for survival", *Annu. Rev. Immunol.*, 18, pp. 83–111.

Fuchs E. (1993). "Reply from Ephraim Fuchs", Immunology Today, vol. 14, pp. 236-237.

Gallucci, S. & Matzinger, P. (2001). "Danger Signals: SOS to the Immune System", Current Opinion in Immunology, vol. 13, pp. 114-119.

Gangardiwala, A. & Polikar, R. (2005). "Dynamically Weighted Majority Voting for Incremental Learning and Comparison of Three Boosting Based Approaches", Proceedings of the International Joint Conference on Neural Networks, pp. 1131-1136.

González, F. A. & Dasgupta, D. (2003). "Anomaly detection using real-valued negative selection", Journal of Genetic Programming and Evolvable Machines, 4, pp. 383–403.

González, F. A. & Dasgupta, D. (2002). "An Immunogenetic Technique to Detect Anomalies in Network Traffic", Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pp. 1081-1088.

Greensmith, J., Aickelin, U., Cayzer, S. (2005). "Introducing Dendritic Cells as a Novel Immune-Inspired Algorithm for Anomaly Detection", 4th International Conference on AIS (ICARIS 2005), pp. 153-167.

Hafler, D. A. (2002). "Degeneracy, as Opposed to Specificity, in Immunotherapy", The *Journal of Clinical Investigation*, vol. 109, pp. 581-584.

Hajela, P., & Lee, J. (1996). "Constrained Genetic Search via Schema Adaptation: An Immune Network Solution", Structural Optimization, 12(1), pp. 11-15.

Hersent, O., Guide, D., Petit, J.P. (2002). "Telefonia IP: Comunicação Multimídia Baseada em Pacotes", Addison-Wesley.

Hightower, R. R., Forrest S. & Perelson, A. S. (1996). "The Baldwin Effect in the Immune System: Learning by Somatic Hypermutation", R. K. Belew and M. Mitchel (Eds.), Adaptive Individuals in Evolving Populations, Addison-Wesley, Reading, MA, pp. 159-167.

Hofmeyr S. A. & Forrest, S. (2000). "Architecture for an Artificial Immune System", *Evolutionary Computation*, 7(1), pp. 45-68.

Hofmeyr S. A. & Forrest, S. (1999). "Immunity by Design: An Artificial Immune System", Proc. of GECCO'99, pp. 1289-1296.

Hunt, J. E. & Cooke, D. E. (1996). "Learning Using an Artificial Immune System", *Journal of Network and Computer Applications*, 19, pp. 189-212.

Janeway, C. A., Travers P., Walport M., Shlomchik M. (2002). "Imunobiologia: O Sistema Imune na Saúde e na Doença", Artmed, 5ª Ed.

Janeway, C. A. (1992). "The Immune System Evolved to Discriminate Infectious Nonself from Noinfectious Self". Immunology Today, vol. 13, pp. 11-6.

Janeway, C. A. (1989). "Approaching the Asymptote? Evolution and Revolution in Immunology". Cold Spring Harbour Symposium, vol. 54, pp. 1-13.

Katipamula, S. & Brambley, M. R. (2005). "Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems – A Review, Part I", HVAC&R Research, 11, pp. 3-25.

Kephart, J. (1994). "A biologically inspired immune system for computers", In Proceedings of the Fourth International Workshop on Synthesis and Simulatoin of Living Systems, Artificial Life IV, pp. 130–139.

Kim, J., Wilson, W., O., Aickelin, U., McLeod J. (2005). "Cooperative Automated Worm Response and Detection ImmuNe Algorithm (CARDINAL) Inspired by T-Cell Immunity and Tolerance", 4th International Conference on AIS (ICARIS 2005), pp. 168-181.

Lafferty, K. J., Cunningham, A. (1975). "A New Analysis of Allogeneic Interections", Australian Journal of Experimental Biology and Medical Sciences, vol. 53, pp. 27-42.

Lederberg, J. (1959). "Genes and Antibodies", Science, vol. 129, pp. 1649-53.

Liu L., Rich B. E., Inobe J., Chen W., Weiner H. L. (1997). "A Potential Pathway of Th2 Development During the Primary Immune Response: IL-10 Pretreated Dendritic Cells Prime Naive CD4 T Cells to Secrete IL-4", Adv. Exp. Med. Biol., vol. 417, pp. 375-381.

Matan, O. (1996). "On Voting Ensembles of Classifiers", In Proceedings of AAAI-96, Workshop on Integrating Multiple Learned Models, pp. 84-88.

Matzinger, P. (2002). "The Danger Model: A Renewed Sense of Self", Science, vol. 296, pp. 301-305.

Matzinger, P. (2001). "The Danger Model in Its Historical Context", Scandinavian Journal of Immunology, vol. 54, pp. 4-9.

Matzinger, P. (1998). "An Innate Sense of Danger", Seminars in Immunology, vol. 10, pp. 399-415.

Matzinger, P. & Fuchs, E.J. (1996). "Beyond 'self' and 'non-self': Immunity is a conversation not a war", Journal of NIH Research, vol. 8, pp. 35-39.

Matzinger, P. (1994). "Tolerance, Danger and the Extended Family", Annual Review of Immunology, vol. 12, pp. 991-1045.

Miller, I. & Freund, J. E. (1985). "Probability and Statistics for Engineers", Prentice-Hall, 3rd Ed.

Perelson, A. S. & Oster, G. F. (1979). "Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliability of Self-Nonself Discrimination", J. theor.Biol., 81, pp. 645-670.

Rogers, E. (1983). "Diffusion of Innovations". The Free Press, Third Edition.

Sarafijanovic, S. & Boudec, J. (2004). "An Artificial Immune System for Misbehavior Detection in Mobile Ad-Hoc Networks with Virtual Thymus, Clustering, Danger Signal, and Memory Detectors", 3rd International Conference on AIS (ICARIS 2004), pp. 316-329.

Sarafijanovic, S. & Boudec, J. (2003). "An artificial immune system approach with secondary response for misbehavior detection in mobile ad-hoc networks", Technical Report IC/2003/65, Ecole Polytechnique Federale de Lausanne.

Segel, L. & Perelson, A. S. (1988). "Computations in Shape Space: A New Approach to Immune Network Theory", Theoretical Immunology, Parte II, A. S. Perelson (Ed.), pp. 321-343.

Seong, SY. & Matzinger, P. (2004). "Hydrophobicity: An Anciente Damage-Associated Molecular Pattern that Ininitiate Immune Responses", Nature Reviews Immunology, vol. 4, pp. 469-478.

Smith, D. J., Forrest, S., Hightower, R. R. & Perelson, A. (1997). "Deriving Shape Space Parameters from Immunological Data", J. theor. Biol., 189, pp. 141-150.

Somayaji, A., Forrest, S., Hofmeyr, S., Longstaff, T. (1996). "A sense of self for unix processes", IEEE Symposium on Security and Privacy, pp. 120–128.

Starlab, URL: http://www.starlab.org/genes/ais/

Timmis, J. (2000). "Artificial Immune Systems: A Novel Data Analysis Technique Inspired by the Immune Network Theory", Tese de Doutorado, Department of Computer Science, University of Whales, September.

Timmis, J., Neal, M., Hunt, J. (2000). "An artificial immune system for data analysis". *Biosystems*, 55, pp. 143–150.

Vargas, P. A. (2005). "Sistemas Computacionais Bio-Inspirados: Síntese e Aplicação em Inteligência Computacional e Homeostase Artificial", Tese de Doutorado, Faculdade de Engenharia Elétrica e de Computação, Unicamp.

Venkatasubramanian, V., R. Rengaswamy, K. Yin, and S. N. Kavuri (2003a). "A review of process fault detection and diagnosis, Part I: Quantitative model-based methods", Computers in Chemical Engineering, 27, pp. 293-311.

Venkatasubramanian, V., R. Rengaswamy, K. Yin, and S. N. Kavuri. (2003b). "A review of process fault detection and diagnosis, Part II: Qualitative models and search strategies", Computers in Chemical Engineering, 27, pp. 313-326.

Venkatasubramanian, V., R. Rengaswamy, K. Yin, and S. N. Kavuri. (2003c). "A review of process fault detection and diagnosis, Part III: Process history based methods", Computers in Chemical Engineering, 27, pp. 327-346.

Wilbanks G. A. & Streilein J. W. (1997). "Fluids from Immune Privileged Sites Endow Macrophages with the Capacity to Induce Antigen-Specific Immune Deviation Via a Mechanism Involving Transformin Growth Factor Beta", Eur J Immunol, vol. 22, pp. 1031-1036.

Willsky (1976), "A Survey of Design Methods for Failure Detection in Dynamic Systems", Automatica, 29, pp. 601-611.

Wilson, R., Martinez, T. R., (1997). "Improved Heterogeneous Distance Functions", Journal of Artificial Intelligence Research, no 6, pp 1-34.

Zinkernagel, R. M. (2000). "On immunological memory", *Philos. Trans. R. Soc. London*, 355, pp. 369–371.

Siglário

ABMS – *Agent-Based Modelling Simulation*

ACM – *Address Complete Message*

ANM – *Answer Message*

APC – *Antigen Presenting Cell*

CLONALG – *CLONal selection ALGorithm*

EI – Engenharia Imunológica

FDD – *Automated Fault Detection and Diagnosis*

FP – Falso Positivo

HEOM – *Heterogeneous Euclidean-Overlap Metric*

HSP – *Heat-Shock Protein*

HTTP – *Hypertext Transfer Protocol*

IAM – *Initial Address Message*

IBMS – *Immunity-Based Systems*

IETF – *Internet Engineering Task Force*

INS – *Infectious-Nonself Model*

IP – *Internet Protocol*

ISUP – *Integrated Services Digital Network User Part*

ITU-T – *International Telecommunications Union – Telecommunications Standardization Sector*

MHC – *Major Histocompatibility Complex*

MMUSIC – *Multiparty Multimedia Session Control*

MTP – *Message Transfer Part*

PAMP – *Pathogen-Associated Molecular Pattern*

PRR – *Pattern Recognition Receptor*

REL – *Release Message*

RFC – *Request for Comments*

RLC – *Release Complete Message*

RTCP – *Real-Time Transport Control Protocol*

RTP – *Real-Time Transport Protocol*

RTPC – Rede Telefônica Pública Comutada

SDP – *Session Description Protocol*

SIA – Sistema Imunológico Artificial

SIH – Sistema Imunológico Humano

SIP – *Session Initiation Protocol*

SNS – *Self-Nonself*

SNSD – *Self-Nonself Discrimination*

SS7 – *Signaling System Number 7*

TUP – *Telephone User Part*

UDP – *User Datagram Protocol*

URI – *Uniform Resource Identifier*

URL – *Universal Resource Locator*

Índice Remissivo de Autores