Campus de São Carlos

**AGENTES-Q: UM ALGORITMO DE ROTEAMENTO ADAPTATIVO E DISTRIBUÍDO
PARA REDES DE TELECOMUNICAÇÕES**

Autora: Karla Vittori

Orientador: Prof. Dr. Aluizio Fausto Ribeiro Araújo

UNIVERSIDADE DE SÃO PAULO

**ESCOLA DE ENGENHARIA
DE SÃO CARLOS**

AGENTES-Q: UM ALGORITMO DE ROTEAMENTO ADAPTATIVO E DISTRIBUÍDO PARA REDES DE TELECOMUNICAÇÕES

Serviço de Pós-Graduação EESC/USP
EXEMPLAR REVISADO
Data de entrada no Serviço: 15/6/2000
Ass.:

Karla Vittori

Dissertação apresentada à Escola de Engenharia de São Carlos, da Universidade de São Paulo, como parte dos requisitos para a obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Aluizio Fausto Ribeiro Araújo

São Carlos

2000

Class.	TESE-EESC
Cott.	6279
Numero	0124/00

1084021

Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca – EESC/USP

V853a Vittori, Karla
Agentes-Q : um algoritmo de roteamento adaptativo e
distribuído para redes de telecomunicações / Karla
Vittori. -- São Carlos, 2000.

Dissertação (Mestrado) -- Escola de Engenharia de
São Carlos-Universidade de São Paulo, 2000.

Área: Engenharia Elétrica.

Orientador: Prof. Dr. Aluízio Fausto Ribeiro
Araújo.

1. Roteamento. 2. Redes de telecomunicações.
3. Aprendizagem-Q. 4. Aprendizagem por esforço dual.
5. Agentes baseados no comportamento coletivo de
formigas. I. Título.

FOLHA DE APROVAÇÃO

Candidata: Engenheira **KARLA VITTORI**

Dissertação defendida e aprovada em 14-04-2000
pela Comissão Julgadora:

Prof. Doutor **ALUIZIO FAUSTO RIBEIRO ARAÚJO (Orientador)**
(Escola de Engenharia de São Carlos - Universidade de São Paulo)

Prof. Associado **AMÍLCAR CARELI CÉSAR**
(Escola de Engenharia de São Carlos - Universidade de São Paulo)

Prof. Adjunto **CARLOS HENRIQUE COSTA RIBEIRO**
(Instituto Tecnológico de Aeronáutica - ITA)

Profa. Associada **VILMA ALVES DE OLIVEIRA**
Coordenadora do Programa de Pós-Graduação
em Engenharia Elétrica

JOSÉ CARLOS A. CINTRA
Presidente da Comissão de Pós-Graduação da EESC

**Dedico este trabalho aos meus pais,
Ubiratan e Meire, por sempre me
apoiarem e acreditarem em mim.**

AGRADECIMENTOS

À Deus, pois sem Ele nada seria possível.

Ao Prof. Dr. Aluizio Fausto Ribeiro Araújo, pela orientação e paciência.

Ao grande amigo Alexandre de Paula, pela amizade, paciência e grande colaboração na realização deste projeto de pesquisa.

Às minhas irmãs, Monike e Andréa, e ao meu cunhado Jordão, pelo constante apoio e incentivo.

Aos meus pais, Ubiratan e Meire, pelo exemplo de vida que me deram e que permitiu que eu não desanimasse nos momentos mais difíceis.

Aos meus amigos, Ana Cláudia, Angélica, Odwaldo, Evelina, Lucimary e Vanderly, agradeço a amizade e a consideração em todos os momentos.

Aos companheiros de laboratório, Guilherme Barreto e Arthur Plínio Braga, agradeço a paciência e amizade.

Ao Departamento de Engenharia Elétrica, pela utilização de seus laboratórios, como também aos seus professores e funcionários, que sempre se dispuseram a colaborar.

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro para a realização deste projeto de pesquisa.

SUMÁRIO

LISTA DE FIGURAS.....	i
LISTA DE TABELAS	iv
LISTA DE SIGLAS E ABREVIATURAS	vii
LISTA DE SÍMBOLOS	viii
RESUMO.....	xi
ABSTRACT.....	xii
1 – Introdução.....	1
2 - Redes de Telecomunicações.....	7
2.1 – Características Gerais	7
2.2 – Multiplexação, Comutação e Transmissão de Mensagens.....	11
2.3 –Redes de Comutação de Circuitos.....	12
2.4 – Redes de Comutação de Pacotes	13
2.5 –Redes de Alta Velocidade.....	14
2.6 – Redes Móveis	15
2.7 - Conclusão	16
3 – Roteamento em Redes de Telecomunicações	18
3.1 – Características Gerais	18
3.2 – Classificações de Sistemas de Roteamento.....	23
3.3 – Conclusão	26
4 - Aprendizagem por Reforço	27
4.1 – Definição.....	27
4.2 – Aprendizagem Baseada em Diferença Temporal	33
4.2.1 –Aprendizagem-Q.....	34
4.3 – Aprendizagem por Reforço Dual	36
4.4 – Conclusão	36

5 – Otimização Baseada no Comportamento Coletivo de Formigas.....	37
5.1 – Formigas na Natureza.....	37
5.2 – Otimização Baseada no Comportamento de Colônias de Formigas.....	43
5.3 – Conclusão	45
6- Algoritmos de Roteamento Baseados em Aprendizagem	46
6.1 – Algoritmos de Roteamento Baseados na Aprendizagem-Q	47
6.1.1 – Roteamento-Q (Q-Routing)	47
6.1.2 – Roteamento-Q Preditivo	50
6.1.3– Roteamento-Q com Reforço Dual.....	53
6.2 – Roteamento Baseado no Comportamento Coletivo de Formigas	56
6.2.1 – Controle Baseado em Formigas	56
6.2.2 – Controle Baseado em Formigas “Espertas”	60
6.3 – Conclusão	63
7 – Agentes-Q: Um Algoritmo de Roteamento Adaptativo e Distribuído	64
7.1 – Integração de Três Estratégias de Otimização para Realizar o Roteamento.....	64
7.2 – Características do Algoritmo Agentes-Q	66
7.2.1 – Componentes do Sistema.....	66
7.2.2 – Comportamento dos Agentes.....	68
7.2.3 – Roteamento de Chamadas	71
7.3 - Conclusão	74
8 – Experimentos Realizados.....	75
8.1 – Características das Redes Utilizadas	75
8.2 – Parâmetros dos Algoritmos de Roteamento Implementados	77
8.3 – Detalhes da Simulação.....	78
8.4 – Testes Realizados	79
8.5 – Resultados Obtidos.....	82
8.5.1 – Rede de Telecomunicações Britânica.....	83
8.5.1.1 - Variações dos Padrões de Tráfego da Rede	83
8.5.1.2 – Variações do Nível de Carga da Rede.....	86
8.5.1.3 – Variações da Topologia da Rede	89
8.5.1.4 – Utilização Permanente de um Fator de Ruído na Seleção das Rotas para as Chamadas	91
8.5.1.5 – Utilização Temporária de um Fator de Ruído na Seleção das Rotas para as Chamadas	94
8.5.2 – Rede de Telecomunicações Norte-Americana	97
8.5.2.1 – Variações dos Padrões de Tráfego da Rede	97
8.5.2.2 – Variações do Nível de Carga da Rede.....	100
8.5.2.3 – Variações na Topologia da Rede	103
8.5.2.4 – Utilização Permanente de Um Fator de Ruído na Seleção das Rotas para as Chamadas	105
8.5.2.5 – Utilização Temporária de Um Fator de Ruído na Seleção das Rotas para as Chamadas	108
8.6 – Discussão dos Resultados	111
8.7 – Conclusão	113

9 – Conclusões.....	115
10 – Referências Bibliográficas	119
Apêndice I.....	I
I.1 – Algoritmo TD(λ).....	I
I.2 – Algoritmo Sarsa(0).....	III
I.3 – Sarsa(λ)	IV
I.4 – Aprendizagem-Q(λ) de Watkins.....	V
I.5 – Aprendizagem-Q(λ) de Peng & Williams	VII

LISTA DE FIGURAS

Figura 1 – Topologias de rede.....	9
Figura 2 - Exemplo de rede simples de telecomunicações, com 7 nós e 9 laços.....	19
Figura 3 – Representação gráfica da aprendizagem por reforço.	29
Figura 4 – (a) Ponte sobre a qual o experimento foi realizado por Deneubourg et al. (1990), (Dorigo et al., 1999). (b) Resultados obtidos em termos da porcentagem de formigas que utilizaram cada um dos ramos da ponte ao longo do tempo (Dorigo et al., 1999).	39
Figura 5 – (a) As formigas iniciam a exploração da ponte binária (Dorigo et al., 1999). (b) Após um período de tempo, a maioria das formigas seleciona o menor caminho (Dorigo et al., 1999).	39
Figura 6 - Rede irregular 6x6.	47
Figura 7 - Rede de telecomunicações da <i>British Telecom</i> usada por Schoondervoerd et al. (1996).	57
Figura 8 – Rede de telecomunicações da <i>British Telecom</i> com números de identificação.....	76
Figura 9 – Rede de telecomunicações norte-americana.....	76
Figura 10 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede britânica.	85
Figura 11 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.	85
Figura 12 – Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede britânica.....	85
Figura 13 - Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.	85
Figura 14 – Desempenho dos três algoritmos sob o nível médio de carga da rede durante todo o período de teste para a rede britânica.	88
Figura 15 – Desempenho dos três algoritmos sob o nível alto de carga da rede durante todo o período de teste para a rede britânica.....	88
Figura 16 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível médio (do passo 8.000 ao 9.000) durante o período de teste para a rede britânica.....	88
Figura 17 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível alto (do passo 8.000 ao 9.000) durante o período de teste para a rede britânica....	88

Figura 18 – Desempenho dos algoritmos sob desconexão permanente do nó 10 no período de teste para a rede britânica.....	90
Figura 19 – Desempenho dos algoritmos sob desconexão temporária do nó 10 no período de teste (do passo 8.000 ao 9.000) para a rede britânica.	90
Figura 20 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 1% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica...	92
Figura 21 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 2% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica...	92
Figura 22 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 5% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica...	93
Figura 23 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 10% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica...	93
Figura 24 - Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 1% na seleção das rotas das chamadas no período de teste para a rede britânica.....	95
Figura 25 - Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 2% na seleção das rotas das chamadas no período de teste para a rede britânica.....	95
Figura 26 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 5% na seleção das rotas das chamadas no período de teste para a rede britânica.....	96
Figura 27 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 10% na seleção das rotas das chamadas no período de teste para a rede britânica.....	96
Figura 28 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.	99
Figura 29 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.....	99
Figura 30 – Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.	99
Figura 31 – Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.....	99
Figura 32 – Desempenho dos três algoritmos sob o nível médio de carga da rede durante todo o período de teste para a rede norte-americana.	102
Figura 33 – Desempenho dos três algoritmos sob o nível alto de carga da rede durante todo o período de teste para a rede norte-americana	102

Figura 34 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível médio (do passo 8.000 ao 9.000) durante o período de teste para a rede norte-americana.....	102
Figura 35 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível alto (do passo 8.000 ao 9.000) durante o período de teste para a rede norte-americana.....	102
Figura 36 – Desempenho dos três algoritmos sob desconexão permanente do nó 60 no período de teste para a rede norte-americana.....	104
Figura 37 – Desempenho dos três algoritmos sob desconexão temporária do nó 60 (do passo 8.000 ao 9.000) no período de teste para a rede norte-americana.....	104
Figura 38 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 1% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.....	107
Figura 39 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 2% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.....	107
Figura 40 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 5% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.....	107
Figura 41 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 10% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.....	107
Figura 42 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 1% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.....	110
Figura 43 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 2% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.....	110
Figura 44 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 5% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.....	110
Figura 45 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 10% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.....	110

LISTA DE TABELAS

Tabela 1 – Tabela de roteamento do nó i	67
Tabela 2 - Geração de sequências de chamadas (Schoonderwoerd et al., 1996).....	79
Tabela 3 – Utilização do mesmo conjunto de probabilidades de chamadas nos períodos de adaptação e teste (Schoonderwoerd et al., 1996).....	81
Tabela 4 - Utilização de conjuntos de probabilidades de chamadas diferentes nos períodos de adaptação e teste.	81
Tabela 5 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede britânica.	83
Tabela 6 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e paralização do lançamento de agentes no período de teste para a rede britânica.	83
Tabela 7 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede britânica... ..	83
Tabela 8 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.	83
Tabela 9 - Resultados obtidos sob alteração da carga da rede para o nível médio durante todo o período de teste para a rede britânica.....	86
Tabela 10 - Resultados obtidos sob alteração da carga da rede para o nível alto durante todo o período de teste para a rede britânica.....	86
Tabela 11 - Resultados obtidos sob alteração temporária da carga da rede para o nível médio durante o período de teste para a rede britânica.	86
Tabela 12 - Resultados obtidos sob alteração temporária da carga da rede para o nível alto durante o período de teste para a rede britânica.	86
Tabela 13 – Resultados obtidos sob desconexão do nó 10 durante todo o período de teste para a rede britânica.....	89
Tabela 14 - Resultados obtidos sob desconexão temporária do nó 10 durante o período de teste para a rede britânica.....	89

Tabela 15 – Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 1% durante todo o período de teste para a rede britânica.	91
Tabela 16 – Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 2% durante todo o período de teste para a rede britânica.	91
Tabela 17 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 5% durante todo o período de teste para a rede britânica.....	91
Tabela 18 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 10% durante todo o período de teste para a rede britânica.	91
Tabela 19 – Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 1% no período de teste para a rede britânica.	94
Tabela 20 – Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 2% no período de teste para a rede britânica.	94
Tabela 21 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 5% no período de teste para a rede britânica.	94
Tabela 22 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 10% no período de teste para a rede britânica.	94
Tabela 23 – Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.	97
Tabela 24 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.	97
Tabela 25 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.	97
Tabela 26 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.	97
Tabela 27 – Resultados obtidos sob alteração da carga da rede para o nível médio durante todo o período de teste para a rede norte-americana.	100
Tabela 28 – Resultados obtidos sob alteração da carga da rede para o nível alto durante todo o período de teste para a rede norte-americana.	100
Tabela 29 – Resultados obtidos sob alteração temporária da carga da rede para o nível médio no período de teste para a rede norte-americana.	101
Tabela 30 – Resultados obtidos sob alteração temporária da carga da rede para o nível alto no período de teste para a rede norte-americana.	101

Tabela 31 – Resultados obtidos sob desconexão do nó 60 durante todo o período de teste para a rede norte-americana.	103
Tabela 32 - Resultados obtidos sob desconexão temporária do nó 60 durante o período de teste para a rede norte-americana.	103
Tabela 33 – Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 1% durante todo o período de teste para a rede norte-americana.	105
Tabela 34 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 2% durante todo o período de teste para a rede norte-americana.	105
Tabela 35 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 5% durante todo o período de teste para a rede norte-americana.	105
Tabela 36 - Resultados obtidos sob utilização de um fator de na seleção das rotas das chamadas igual a 10% durante todo o período de teste para a rede norte-americana.	105
Tabela 37 – Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 1% durante todo o período de teste para a rede norte-americana.	108
Tabela 38 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 2% durante todo o período de teste para a rede norte-americana.	108
Tabela 39 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 5% durante todo o período de teste para a rede norte-americana.	108
Tabela 40 - Resultados obtidos sob utilização temporária de um fator de na seleção das rotas das chamadas igual a 10% durante todo o período de teste para a rede norte-americana.	108

LISTA DE SIGLAS E ABREVIATURAS

ABC – Ant-Based Control

ACO - Ant Colony Optimization

ATM - Asynchronous Transference Mode

B-ISDN - Broadband Integrated Services Digital Network

BT – British Telecom

CDMA – Code Division Multiplexing

CDV - Cell Delay Variation

CLR - Cell Loss Ratio

CTD - Cell Transfer Delay

FDM - Frequency Division Multiplexing

LAN - Local Area Network

MAN - Metropolitan Area Network

PDM - Processo de Decisão Markoviano

QoS - Quality of Service

SCR - Sustainable Cell Ratio

STM - Synchronous Transference Mode

TD - Temporal Difference

TDM - Time Division Multiplexing

WAN - Wide Area Network

LISTA DE SÍMBOLOS

A	- conjunto finito das ações do agente
S	- conjunto finito de estados do ambiente
$\mathbb{R}$	- conjunto dos números reais
a_t	- ação realizada no instante t
s_t	- estado do ambiente no instante t
r_t	- retorno do ambiente no instante t
$T(s_t, a_t, s_{t+1})$	- probabilidade de transição do estado s_t para o estado s_{t+1} ao se realizar a ação a_t
π	- política do agente
π^*	- política ótima do agente
a^*	- ação a ótima
$V_\pi(s)$	- avaliação do estado s ao seguir a política π
$V_\pi^*(s)$	- avaliação do estado s ao seguir a política ótima
$Q^\pi(s, a)$	- avaliação da ação a no estado s ao seguir a política π
$Q^*(s, a)$	- avaliação da ação a no estado s ao seguir a política ótima
$V, V(s)$	- estimativa da avaliação do estado s
$Q, Q(s, a)$	- estimativa da avaliação da ação a realizada no estado s
$R, R(s)$	- retorno do ambiente para o estado s
$R(s, a)$	- retorno do ambiente para ação a realizada no estado s
$\Delta V(s)$	- atualização da estimativa da avaliação do estado s

$\Delta Q(s, a)$	- atualização da estimativa da avaliação da ação a realizada no estado s
α	- taxa de aprendizagem
γ	- fator de desconto
$e_t(s)$	- traço de elegibilidade do estado s no instante t
$e_t(s, a)$	- traço de elegibilidade da ação a realizada no estado s no instante t
e_t	- erro no instante t considerando a estimativa da ação que maximiza a função de avaliação do par estado-ação
e_t'	- erro no instante t considerando a ação efetivamente realizada
δ_t	- erro de diferença temporal no instante t
q	- tempo decorrido na fila de um nó
s	- tempo decorrido na transmissão através da conexão (i, j)
$B_i(j, d)$	- melhor estimativa do tempo de transmissão entre os nós i e d utilizando a conexão (i, j)
$R_i(j, d)$	- taxa de restabelecimento para o caminho do nó i ao d utilizando a conexão (i, j)
$U_i(j, d)$	- última atualização do tempo para o caminho do nó i ao d utilizando a conexão (i, j)
$P_j(i, o)$	- probabilidade de que um agente que se destina ao nó o seja enviado do nó j ao vizinho i
ΔP_k	- incremento da probabilidade de escolha do nó k
$idade_k$	- idade absoluta do agente ao visitar o nó k
N_j	- conjunto de vizinhos do nó j
C_i	- capacidade máxima de transmissão do nó i
O_i	- capacidade de processamento ociosa do nó i
$Q_i(j, d)$	- probabilidade de escolha do nó vizinho de i igual a j com relação ao destino d
r_i	- sinal de reforço recebido pelo agente ao visitar o nó i
o	- nó origem
d	- nó destino

a	- nó atual
p	- próximo nó
n	- número de nós da rede
m_i	- número de nós vizinhos de i
i	- nó qualquer da rede
$o_p(t)$	- capacidade de processamento ociosa do nó p no instante t
c_p	- número de chamadas perdidas na rede
β, w	- constantes de proporcionalidade
$A_i(j, d)$	- avaliação de vizinho j de i em direção ao destino d
k_1, k_2	- constantes de proporcionalidade
ϕ, δ, ρ	- constantes de proporcionalidade
ω	- fator de peso

RESUMO

Vittori, K. (2000). Agentes-Q: Um Algoritmo de Roteamento Adaptativo e Distribuído para Redes de Telecomunicações. São Carlos, 2000. 125 p. Dissertação (mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

As redes de telecomunicações são responsáveis pelo envio de informação entre pontos de origem e destino. Dentre os diversos dispositivos que participam deste processo, destaca-se o sistema de roteamento, que realiza a seleção das rotas a serem percorridas pelas mensagens ao longo da rede e sua condução ao destino desejado. O avanço das tecnologias utilizadas pelas redes de telecomunicações provocou a necessidade de novos sistemas de roteamento, que sejam capazes de lidar corretamente com as diversas situações enfrentadas atualmente. Dentro deste contexto, este projeto de pesquisa desenvolveu um algoritmo de roteamento adaptativo e distribuído, resultado da integração de três estratégias de aprendizagem e da adição de alguns mecanismos extras, com o objetivo de obter um algoritmo eficiente e robusto às diversas variações das condições de operação da rede. As abordagens utilizadas foram a aprendizagem-Q, aprendizagem por reforço dual e aprendizagem baseada no comportamento coletivo de formigas. O algoritmo desenvolvido foi aplicado a duas redes de comutação de circuitos e seu desempenho foi comparado ao de dois algoritmos baseados no comportamento coletivo de formigas, que foram aplicados com sucesso ao problema de roteamento. Os experimentos conduzidos envolveram situações reais enfrentadas pelas redes, como variações dos seus padrões de tráfego, nível de carga e topologia. Além disto, foram realizados testes envolvendo a presença de ruído nas informações utilizadas para a seleção das rotas a serem percorridas pelas chamadas. O algoritmo proposto obteve melhores resultados que os demais, apresentando maior capacidade de adaptação às diversas situações consideradas. Os experimentos demonstraram que novos mecanismos de otimização devem ser anexados ao algoritmo proposto, para melhorar seu comportamento exploratório sob variações permanentes do nível de carga da rede e presença de ruído nos dados utilizados em suas tarefas.

Palavras-chave: Roteamento, redes de telecomunicações, aprendizagem-Q, aprendizagem por reforço dual, agentes baseados no comportamento coletivo de formigas.

ABSTRACT

Vittori, K. (2000). Q-Agents: An Adaptive and Distributed Routing Algorithm for Telecommunications Networks. São Carlos, 2000. 125 p. Dissertação (mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo.

The telecommunications networks are responsible for transmitting information between source and destination points in a fast, secure and reliable way, providing low cost and high quality services. Among the several devices that takes place on this process, there is the routing system, which selects the routes to be traversed by the messages through the network and their forwarding to the destination desired. The advances in technologies used by telecommunications networks caused the necessity of new routing systems, that can work correctly with the situations faced by current telecommunications networks. Hence, this research project developed an adaptive and distributed routing algorithm, resulting of the integration of three learning strategies and addition of some extra mechanisms, with the goal of having a robust and adaptive algorithm to the several variations on operation network conditions. The approaches chosen were Q-learning, dual reinforcement learning and learning based on collective behavior of ants. The developed algorithm was applied to two circuit-switching telecommunications networks and its performance was compared to two algorithms based on ant colony behavior, which were used with success to solve the routing problem. The experiments run comprised real situations faced by telecommunications networks, like variations on the network traffic patterns, load level and topology. Moreover, we did some tests with the presence of noise in information used to select the routes to be traversed by calls. The algorithm proposed produced better results than the others, showing higher capacity of adaptation to the several situations considered. The experiments showed that new optimization mechanisms must be added to the routing algorithm developed, to improve its exploratory behavior under permanent variations on network load level and presence of noise in data used in its tasks.

Keywords: Routing, telecommunications networks, Q-learning, dual reinforcement learning, agents based on ant colonies behavior.

1 – Introdução

As redes de telecomunicações são responsáveis pelo envio de informação entre pontos de origem e destino. A informação transmitida pode ser constituída de dados, voz e/ou imagens. As redes de telecomunicações são classificadas de acordo com a natureza da informação enviada, que influencia diretamente a forma de processamento e transmissão da mesma. As principais tecnologias de redes compreendem: redes comutadas por circuitos, redes comutadas por pacotes, redes de alta velocidade e redes móveis. As redes comutadas por circuitos foram inicialmente desenvolvidas para transmissão de sinais de voz, enquanto as redes de pacotes o foram para transmissão de dados. O avanço da tecnologia e dos equipamentos utilizados pelas redes de telecomunicações possibilitou a comunicação integrada destes tipos de informação, realizada pelas redes de alta velocidade. As redes móveis constituem aquelas em que os usuários se movimentam em relação às estações de processamento das mensagens. Devido a evolução tecnológica ocorrida nos últimos tempos, atualmente as redes móveis também transmitem sinais de natureza diversa. O objetivo de todas as redes de telecomunicações é enviar a informação desejada de uma dada origem a um destino determinado, de forma a maximizar uma ou várias medidas de desempenho e minimizar seus custos de operação.

O envio da informação ao longo de uma rede de telecomunicações envolve um processo complexo de ações e tomada de decisão pelos seus diversos componentes, que as realizam de forma interdependente. Dentre os dispositivos que participam deste processo, os roteadores desempenham papel fundamental, pois são responsáveis pela seleção das rotas a serem percorridas pela informação e pelo envio desta em direção ao destino desejado (Steenstrup, 1995). Cada informação trafega pela rede movimentando-se através de uma sequência de roteadores, localizados em pontos adjacentes da mesma, até que seu destino seja alcançado. O estabelecimento das rotas a serem percorridas pelas mensagens através da rede é realizado de acordo com as condições de operação da mesma e as características do tráfego. As informações sobre o estado da rede e do tráfego são normalmente codificadas em variáveis que encontram-se armazenadas em tabelas de roteamento, presentes em cada ponto da rede. A escolha do caminho a ser utilizado por uma mensagem em direção ao destino desejado é realizada através da consulta à tabela de roteamento do roteador em que a mesma se encontra. Assim, as funções de roteamento consistem basicamente na obtenção,

atualização e utilização dos valores armazenados nas tabelas. A representação destas variáveis é dependente da tecnologia de rede utilizada. O roteamento deve realizar suas funções de forma a obter o balanço adequado entre as limitações de operação da rede e os objetivos de seus usuários.

Os algoritmos de roteamento inicialmente utilizados pelas redes de telecomunicações constituíam sistemas estáticos e centralizados. Deste modo, as rotas eram determinadas *offline*, baseadas na previsão da demanda do tráfego, e permaneciam fixas durante a operação da rede, sendo utilizadas rotas alternativas somente em casos de falha ou falta em um roteador ou linha de transmissão da rede. Assim também, o roteamento de todas as mensagens presentes na rede era realizado por um único roteador central. Estas características são decorrentes das limitações apresentadas pelos equipamentos e meios de transmissão utilizados nas redes de telecomunicações até meados da década de 80. A partir deste período, a evolução da tecnologia de chaveamento e transmissão da informação possibilitou o desenvolvimento de equipamentos com alta capacidade de transmissão, como também a obtenção de informações *online* sobre a rede e o tráfego. Assim, foi possível aumentar consideravelmente o número de usuários das redes de telecomunicações e tornou-se necessário realizar o roteamento dinâmico, que seleciona os caminhos de acordo com as condições correntes da rede e do tráfego. A utilização deste tipo de roteamento proporcionou o aumento da confiabilidade e eficiência das redes de telecomunicações, que passaram a transmitir um volume maior de informações de forma rápida e segura, além de proporcionar uma maior variedade de serviços aos usuários. Deste modo, atualmente as redes de telecomunicações oferecem serviços de videoconferência, aprendizado a distância, vídeo por encomenda, bem como o acesso a todo tipo de informação através da rede mundial de computadores, a Internet, entre outros.

Entretanto, a maioria dos algoritmos atualmente utilizados pelas redes de telecomunicações ainda realiza o roteamento centralizado e os valores armazenados nas tabelas de roteamento são atualizados a intervalos de tempo regulares. Esta situação decorre dos altos custos envolvidos na transição da tecnologia empregada em uma rede. Deste modo, os sistemas de roteamento têm apresentado dificuldades em lidar com o crescimento do número de usuários e de suas necessidades de comunicação, bem como com o comportamento dinâmico das redes de telecomunicações.

Assim, o desenvolvimento de sistemas de roteamento robustos, distribuídos e adaptativos tem se tornado uma questão de estudo e interesse para empresas de telecomunicações localizadas em todo o mundo (Ash, 1997). Deste modo, empresas provedoras de serviços e de equipamentos para redes de telecomunicações, juntamente com

instituições acadêmicas, vem realizando pesquisas intensivas no sentido do desenvolvimento desta área.

Dentro deste contexto, este projeto de pesquisa desenvolveu um algoritmo de roteamento adaptativo e distribuído, capaz de atuar de modo eficiente e eficaz face ao perfil atual das redes de telecomunicações e da demanda de seus usuários. O sistema de roteamento desenvolvido lida com diversas mudanças ocorridas nas características do ambiente de uma rede de telecomunicações através de aprendizado. Este aprendizado compreende a seleção das rotas a serem percorridas pelas mensagens e a verificação da consequência desta escolha sobre o estado da rede, com relação ao volume de chamadas processado pelos roteadores. Assim, o sistema de roteamento calcula a qualidade do caminho selecionado, o que influenciará a tomada de decisão para as próximas mensagens presentes na rede. Deste modo, o roteamento desenvolvido é capaz de se adaptar rapidamente às constantes variações no estado da rede e do tráfego. Porém, este algoritmo deve realizar o balanceamento adequado entre dinamismo e estabilidade.

A fim de otimizar o comportamento do algoritmo proposto, ele constitui um sistema resultante da combinação de propriedades pertinentes a três técnicas de aprendizagem recentemente desenvolvidas e utilizadas em sistemas de roteamento. A escolha destas estratégias encontra-se vinculada ao desempenho obtido pelos sistemas de roteamento que as utilizaram, em relação ao roteamento estático usualmente empregado pelas redes de telecomunicações. As técnicas selecionadas compreendem: aprendizagem-Q (Watkins, 1989), aprendizagem por reforço dual (Goetz et al., 1996) e um método de otimização baseado no comportamento de insetos sociais, em particular de formigas (Dorigo, 1992).

A aprendizagem-Q se baseia na interação entre um agente e seu meio ambiente, a fim de atingir uma meta previamente estabelecida. A escolha desta técnica está relacionada com sua capacidade de aprendizado e adaptação às condições do meio ambiente, sem necessitar de conhecimento prévio e completo sobre o mesmo. Desta forma, ela se ajusta ao problema de roteamento em redes de telecomunicações, onde não há possibilidade de obter exemplos representativos de todas as situações em que o ambiente pode se encontrar e do comportamento esperado do agente, devido à diversas mudanças imprevistas e inesperadas ocorridas nestas redes ao longo do tempo. Além disso, a aprendizagem-Q apresenta a comprovação de convergência para a política de decisão ótima. Os algoritmos de roteamento que utilizam esta técnica foram aplicados a redes comutadas por pacotes e apresentaram desempenho superior ao de algoritmos estáticos.

A utilização de aspectos pertinentes à segunda técnica de aprendizagem, que corresponde a aprendizagem por reforço dual, se deve a maior exploração do espaço de

busca de soluções por ela realizada. Neste caso, um roteador realiza a atualização de dois valores armazenados nas tabelas de roteamento a cada movimento realizado. Assim, o algoritmo atualiza a qualidade de caminhos percorridos em duas direções diferentes, realizando a exploração simultânea do espaço de busca de soluções para dois pares de roteadores origem-destino da rede. Os algoritmos de roteamento que combinam características desta técnica com as da aprendizagem-Q também foram aplicados a redes de comutação de pacotes e apresentaram maior velocidade de adaptação às mudanças na rede que aqueles baseados somente na aprendizagem-Q.

O método baseado no comportamento coletivo de formigas utiliza um conjunto de agentes simples, que percorrem o ambiente de forma simultânea e assíncrona em busca de boas soluções para o problema em questão. As soluções são baseadas na informação coletada pelos agentes sobre os roteadores visitados durante seu movimento ao longo do ambiente. Esta informação é utilizada para calcular a qualidade de todo o caminho selecionado, como também de cada movimento realizado. Os agentes compartilham as informações obtidas através de um método de comunicação indireta. Os algoritmos de roteamento baseados neste método foram aplicados tanto a redes comutadas por circuitos como por pacotes. Os agentes selecionam os caminhos através de uma escolha probabilística, permitindo que se obtenha informação sobre rotas alternativas. A utilização de um conjunto de agentes que se comporta desta maneira permite que cada roteador possua informação atualizada sobre as condições de operação da rede e do tráfego a todo momento, selecionando assim os melhores caminhos a cada instante de tempo. Os sistemas de roteamento que utilizam este método apresentaram desempenho superior tanto a métodos de roteamento estáticos como a sistemas dinâmicos empregados pelas empresas de telecomunicações.

Deste modo, o algoritmo Agentes-Q é composto por um conjunto de agentes simples que se movimentam através da rede de forma probabilística e se comunicam indiretamente sobre a qualidade dos caminhos percorridos. A interação entre todos os agentes permite que as melhores rotas sejam continuamente selecionadas, possibilitando a adaptação rápida e eficiente à diversas condições de operação da rede. Os valores armazenados nas tabelas de roteamento são representados por valores-Q e atualizados de forma dual através da regra utilizada pela aprendizagem-Q. Ao utilizar esta última estratégia, o algoritmo proposto possui convergência comprovada para os melhores caminhos.

O algoritmo de roteamento desenvolvido foi aplicado inicialmente a um modelo da rede de telefonia utilizada pela empresa de telecomunicações britânica (*British Telecom* - BT) e posteriormente a um modelo da rede utilizada pela empresa norte-americana AT&T. A escolha da primeira rede se deve ao fato de que os dois algoritmos que também foram

implementados neste projeto e com os quais o algoritmo proposto teve seu desempenho comparado terem sido aplicados a esta rede. Estes algoritmos são baseados no comportamento coletivo de formigas e denominados Controle Baseado em Formigas (ABC – *Ant-based Control*), (Schoonderwoerd et al., 1996; 1997) e Controle Baseado em Formigas Espertas (Bonabeau et al., 1998). Eles foram selecionados devido ao desempenho superior por eles apresentado em relação a algoritmos estáticos e dinâmicos utilizados em redes de comutação de circuitos. A escolha da segunda rede encontra-se relacionada com a necessidade de aplicar o algoritmo não somente a uma topologia de rede, para verificar sua capacidade de adaptação à diferentes tamanhos de rede e conseqüentemente a diferentes graus de complexidade do problema de roteamento.

Os resultados obtidos pelos algoritmos encontram-se relacionados com a porcentagem de chamadas que não alcançaram o destino desejado na rede, denominadas chamadas perdidas. Isto ocorre quando o roteador selecionado pelo algoritmo para transmitir a chamada não possui capacidade de processamento disponível, como também quando todos os roteadores adjacentes a um determinado ponto foram utilizados numa rota e a única solução seria utilizar novamente um destes roteadores na mesma rota.

Os experimentos realizados compreenderam situações enfrentadas pelas redes atuais. Desta forma, os testes envolveram a variação dos principais parâmetros de operação de uma rede de telecomunicações, que constituem: padrões de tráfego, nível de carga e topologia da rede. Além disso, foram utilizados diferentes níveis de ruído nas informações usadas pelo algoritmo na seleção das rotas. O nível de carga e topologia da rede sofreram mudanças contínuas e temporárias de suas características. Da mesma forma, o fator de ruído foi introduzido de forma permanente e também durante um determinado intervalo de tempo.

Os três algoritmos implementados foram capazes de aprender e se adaptar às diversas situações enfrentadas pelas duas redes consideradas. A combinação de propriedades pertencentes às três técnicas de aprendizagem acima citadas, aliada à adição de novos mecanismos de exploração do ambiente permitiu que o algoritmo proposto apresentasse melhor desempenho que os demais sob praticamente todos os experimentos realizados com relação à qualidade de adaptação às variações das redes. Os resultados produzidos pelos agentes-Q não apresentaram diferença significativa em relação aos obtidos pelos demais algoritmos somente em uma situação considerada, que constitui uma condição extrema de operação das redes. Os três algoritmos analisados também demonstraram capacidade de generalização, pois mantiveram o mesmo tipo de comportamento nos dois modelos de rede considerados. Mais uma vez, o algoritmo Agentes-Q se comportou de forma mais eficiente e eficaz que os demais, obtendo melhores resultados que seus competidores.

Este documento é iniciado com a descrição dos aspectos fundamentais e dos componentes das redes de telecomunicações e a definição das principais tecnologias de comutação e transmissão de informações. No Capítulo 3, o problema de roteamento é descrito, apresentando suas características, fatores a serem considerados na seleção das rotas e sua classificação. No Capítulo 4, as características da aprendizagem por reforço, e em particular da aprendizagem-Q, são discutidas e analisadas, seguidas, no Capítulo 5, dos aspectos do método baseado no comportamento coletivo de formigas. No Capítulo 6, são descritos e analisados alguns algoritmos de roteamento baseados na aprendizagem-Q e no comportamento coletivo de formigas, onde os primeiros foram aplicados a redes de comutação de pacotes e os últimos a redes de comutação de circuitos. Estes algoritmos foram desenvolvidos recentemente e têm apresentado um bom desempenho em relação a sistemas de roteamento tanto clássicos como inteligentes. Além disso, os algoritmos descritos utilizam a aprendizagem-Q ou o método baseado em colônias de formigas para realizar o roteamento, e por esta razão, apresentam aspectos em comum com o algoritmo desenvolvido. O algoritmo de roteamento proposto é descrito em detalhes no Capítulo 7. No Capítulo 8, são apresentadas as redes utilizadas nos experimentos realizados, os parâmetros dos algoritmos analisados e as situações consideradas nos testes. As vantagens e limitações dos algoritmos implementados são analisadas. No Capítulo 9 são apresentadas as conclusões obtidas ao longo do desenvolvimento deste algoritmo de roteamento. No Apêndice I, são apresentados os algoritmos de aprendizagem baseados em diferença temporal denominados: $TD(\lambda)$, $Sarsa(0)$, $Sarsa(\lambda)$, $Q(\lambda)$ de Watkins e $Q(\lambda)$ de Peng & Williams. Estes algoritmos são apresentados no apêndice desta dissertação para que o leitor adquira conhecimento sobre os principais métodos de aprendizagem baseados em diferença temporal e possa assim compreender as diferenças entre eles e o algoritmo de aprendizagem-Q que nos conduziram a escolha deste método para realizar o roteamento em uma rede de telecomunicações.

2 - Redes de Telecomunicações

Este capítulo é constituído inicialmente pela descrição das principais características das redes de telecomunicações, apresentando seus componentes, os tipos de ligações possíveis entre os usuários, as topologias utilizadas, seguidas pelos meios físicos de transmissão, os modos de transmitir o tráfego de informação, os tipos de sinais transmitidos e de redes empregadas e os serviços atualmente oferecidos pelas redes de telecomunicações. Em seguida, são descritas as principais tecnologias de comutação e transmissão de informações: redes de comutação de circuitos, redes de comutação de pacotes, redes de alta velocidade e redes móveis.

2.1 – Características Gerais

Uma rede de telecomunicações tem como função a transmissão de informação entre pontos de origem e destino, de acordo com os objetivos de seus usuários. Esta informação pode conter sinais de voz, dados e/ou imagens. Uma rede é formada por um conjunto de dispositivos, que trocam informações entre si e compartilham recursos, interligados através de um sistema de comunicação. Assim, uma rede consiste em um arranjo topológico composto por estações de processamento da mensagem e enlaces físicos, através dos quais ocorre a transmissão das mesmas. Dentro deste contexto, as redes de telecomunicações possuem os seguintes componentes:

- ♦ **Terminais de comunicação**

São os dispositivos que se comunicam através da troca de mensagens. Estes dispositivos são caracterizados por capacidades de processamento e armazenamento de mensagens.

- ♦ **Canais de comunicação**

São formados pelos dispositivos que interligam os terminais de comunicação, constituindo o meio físico de transmissão das mensagens, denominados conexões da rede. Estes dispositivos são caracterizados por sua capacidade de transmissão de mensagens, denominada largura de banda.

♦ Processadores

Constituem os mecanismos responsáveis pelo processamento da mensagem. São eles: modems (moduladores/demoduladores) e multiplexadores (Soares et al., 1995). Os primeiros realizam o deslocamento do sinal de sua faixa de frequência original para uma determinada faixa, a fim de que ele seja transmitido juntamente com outros sinais por um mesmo meio físico e após seu recebimento, o deslocam novamente para sua faixa de frequência original. Eles também executam a conversão do sinal analógico para digital e vice-versa. Os multiplexadores utilizam técnicas específicas para realizar a transmissão de diversos sinais por um mesmo meio físico, que serão descritas ao longo deste trabalho.

♦ Dispositivos da rede

Os dispositivos da rede são responsáveis pela comutação (ou chaveamento) e roteamento das mensagens entre as diversas conexões da rede (Soares et al., 1995). A técnica de comutação consiste em alocar os recursos da rede para que as mensagens sejam transmitidas ao longo dos diversos dispositivos que a compõem, enquanto o roteamento seleciona as rotas a serem percorridas pela informação. Além disso, existem dispositivos que asseguram a transmissão das mensagens ao longo da rede, que são: amplificadores, detectores de erro e repetidores.

♦ Sistema da rede (*software*)

Está relacionado com os sistemas de planejamento e gerenciamento empregados na rede. As redes utilizam também um conjunto de regras e procedimentos com a finalidade de dirigir a transmissão entre seus componentes, denominado protocolo.

Os terminais de comunicação podem realizar as conexões entre os usuários de diversas formas:

♦ Conexões ponto a ponto

São aquelas em que a mensagem possui somente uma origem e um destino. Este é o caso de serviços interativos como telefonia, correio eletrônico e acesso a bases de dados *online*.

♦ Conexões ponto a multiponto

Conexões em que a mensagem apresenta uma única origem e vários destinos, onde há serviços de distribuição, como sistemas de vídeo de acordo com a demanda.

♦ Conexões multiponto a ponto

Situações em que há múltiplas origens e um único destino para a mensagem, como centrais de coleta de dados.

♦ Conexões multiponto a multiponto

Conexões em que a mensagem possui várias origens e vários destinos, como videoconferências.

Figura 1 – Topologias de rede.

Os enlaces entre os terminais de comunicação são utilizados de forma diferente pela topologia de rede empregada, que pode ser dividida nas seguintes classes (Figura 1): topologia totalmente conectada, parcialmente conectada, em anel, estrela ou barra (Soares et al., 1995). Na topologia totalmente conectada, todos os terminais da rede encontram-se interconectados. Apesar de apresentar alto grau de paralelismo de comunicação entre os terminais, esta topologia se torna em geral inviável economicamente, devido ao elevado número de enlaces físicos necessários. Na topologia parcialmente conectada, também denominada topologia em grafo, não existem ligações entre todos os terminais da rede, mas sim caminhos alternativos, que podem ser utilizados em situações de falha ou congestionamentos em determinadas rotas. A topologia em anel procura diminuir o número de ligações no sistema e simplificar o tipo de ligação utilizada. Deste modo, utiliza-se em geral ligações ponto a ponto que operam em um único sentido de transmissão. As mensagens circulam pelo anel através do sentido determinado e são transmitidas de um terminal a outro, até que alcancem o destino desejado. Uma limitação desta topologia encontra-se no fato de que, em caso de falha ou falta em um terminal da rede, a transmissão de mensagens é comprometida. Entretanto, existem métodos para restabelecer a topologia em anel. Na conexão estrela, cada terminal é conectado a uma estação central que envia as mensagens

para o destino desejado. Assim, esta estação deve transmitir as mensagens de forma eficiente e confiável. A última topologia apresentada é a topologia em barra, onde os terminais são conectados em uma sequência linear. Desta forma, cada terminal recebe toda a mensagem e a transmissão da mesma não é prejudicada em caso de falha ou falta em um dos terminais.

Os canais de comunicação podem ser de vários tipos, sendo os principais: par trançado, cabo coaxial, fibra ótica e radiofusão (redes sem fio) que apresentam-se ordenados de forma crescente, de acordo com a velocidade de envio das mensagens (Soares et al., 1995). O sistema de radiofusão é utilizado nas redes de comunicação móvel, bem como em casos onde a instalação de cabos metálicos ou fibra ótica se torna muito difícil ou impossível. Enlaces de satélites e microondas também podem ser empregados, de acordo com as circunstâncias.

As redes de telecomunicações podem ser classificadas em três tipos, de acordo com a área coberta por cada uma delas: redes locais, redes metropolitanas e redes geograficamente distribuídas (Soares et al., 1995). Desta forma, as redes locais (*Local Area Networks* - LANs) interligam dispositivos em uma pequena região. Em geral, considera-se uma pequena região distâncias entre 100 m e 25 Km. As redes locais apresentam altas taxas de transmissão (de 0,1 a 100 Mbps – 10^6 bits por segundo) e são em geral de propriedade privada¹. As topologias mais utilizadas nestas redes são: estrela, anel e barra. A partir do momento em que a distância associada aos usuários de uma rede atinge distâncias de áreas metropolitanas, estes sistemas são denominados redes metropolitanas (*Metropolitan Area Network* - MANs). Uma rede metropolitana apresenta características semelhantes as das redes locais, porém cobrem distâncias maiores do que estas e operam em maior velocidade que as mesmas. Estas redes pertencem às empresas de telecomunicações e seu acesso é realizado através de um provedor. As redes geograficamente distribuídas (*Wide Area Networks* - WANs) foram desenvolvidas devido a necessidade de interconectar um número maior de usuários, localizados em pontos muito distantes. Estas redes são em geral públicas, por apresentarem alto custo de comunicação. As redes geograficamente distribuídas utilizam topologia totalmente ou parcialmente conectada, ou ainda em anel.

¹ É importante ressaltar que os valores associados aos termos pequena região e altas taxas de transmissão são susceptíveis a evolução tecnológica.

2.2 – Multiplexação, Comutação e Transmissão de Mensagens

A multiplexação das mensagens, que ocorre antes da transmissão das mesmas, pode ser realizada de forma síncrona ou assíncrona (Soares et al., 1995). Na primeira, empregada na comutação de circuitos, o caminho a ser percorrido por uma informação é estabelecido antes do envio da mesma e permanece dedicado até que a comunicação entre os pontos origem-destino termine. Tais mensagens são enviadas de forma ordenada, constituindo o modo de transferência síncrono (*Synchronous Transference Mode* - STM). Assim, cada uma das mensagens a serem enviadas utiliza uma capacidade fixa de transmissão da conexão e é enviada de acordo com: (i) a faixa de frequência que ocupa, multiplexação por divisão de frequência (*Frequency Division Multiplexing* - FDM), (ii) ou com um intervalo de tempo estabelecido, multiplexação por divisão de tempo (*Time Division Multiplexing* - TDM). A primeira delas constitui uma tecnologia analógica e a segunda, digital. Atualmente, uma nova tecnologia vem sendo utilizada pelas redes móveis digitais, denominada CDM, multiplexação por divisão de código (*Code Division Multiplexing* - CDM). Neste caso, ao invés de utilizar divisão de frequência ou tempo, utiliza-se códigos matemáticos para transmitir e distinguir entre diversas transmissões de informação. Como a distinção é realizada através de códigos digitais, vários usuários podem compartilhar a largura de banda de uma conexão simultaneamente. Esta evolução permitiu a melhoria da capacidade de transmissão das redes, como também o aumento da área de cobertura e da qualidade da informação transmitida.

No modo de transferência assíncrono (*Asynchronous Transference Mode* - ATM) é utilizada a multiplexação estatística, onde a mensagem é transmitida sempre que necessário, sem seguir uma ordem no domínio do tempo ou da frequência. Neste caso, a largura de banda da conexão é alocada de acordo com a demanda apresentada pelo tráfego no momento, sem haver desperdício da capacidade de transmissão de mensagens das conexões.

A comutação de mensagens pode ser realizada de duas formas: através da comutação de circuitos ou comutação de pacotes, desenvolvidas para diferentes tipos de tráfego (Steenstrup, 1995). As redes de comutação de circuitos determinam que um caminho seja estabelecido entre a origem e o destino de uma mensagem antes que sua transmissão se inicie e que tal rota permaneça dedicada até o fim da comunicação. Neste caso, a condução da mensagem é denominada orientada à conexão (*connection-oriented*). As redes de comutação de pacotes, por sua vez, constroem o caminho a ser percorrido sem a necessidade de alocação fixa dos recursos da rede. A mensagem é desmembrada em unidades menores, denominadas pacotes e cada um destes percorre uma rota de forma independente dos demais,

selecionada em cada ponto da rede visitado. Assim, a condução das mensagens é denominada não orientada a conexão (*connectionless*).

Os sinais gerados para transmissão em uma rede de telecomunicações podem ser de dois tipos: sinais analógicos e sinais digitais (Soares et al., 1995). Os sinais analógicos apresentam variações contínuas de amplitude ao longo do tempo. Os sinais digitais, por sua vez, caracterizam-se pela variação discreta de amplitude. Os equipamentos que utilizam sinais digitais representam um avanço em relação aos analógicos, permitindo o envio de um volume maior de informações e de natureza diferente, de forma rápida e precisa, reduzindo portanto o número de equipamentos utilizados pela rede. Deste modo, vem sendo utilizados nos últimos tempos um número crescente de equipamentos e tecnologias de transmissão digitais.

Devido aos avanços obtidos em tecnologia de comutação e transmissão de mensagens, as redes de telecomunicações atuais provêm vários serviços, entre eles: recursos para TV de alta definição, correio eletrônico, comunicação de voz através da Internet (*voice mail*), videoconferência, teleconferência, fax, acesso a base de dados, transferência de arquivos e acesso a rede mundial de computadores, a Internet, havendo possibilidade de expansão dos mesmos.

Após a descrição das principais características das redes de telecomunicações, são apresentados a seguir os aspectos fundamentais das principais tecnologias de comutação e transmissão de informações: redes de comutação de circuitos, redes de comutação de pacotes, redes de alta velocidade e redes móveis.

2.3 – Redes de Comutação de Circuitos

A comutação de circuitos constitui a primeira tecnologia de chaveamento utilizada pelas redes de telecomunicações (Steenstrup, 1995). Ela foi projetada originalmente para comunicação de voz, sendo portanto empregada em redes de telefonia. Inicialmente, os dispositivos das redes de comutação de circuitos eram analógicos e realizavam a multiplexação baseada na divisão da frequência (FDM). A mudança das tecnologias de chaveamento e transmissão de analógicas para digitais tem possibilitado o transporte de diversos tipos de tráfego pelas redes de comutação de circuitos, como voz, dados e vídeo. A evolução da tecnologia também permitiu que estas redes utilizassem a multiplexação por divisão de tempo (TDM).

As redes estabelecem circuitos físicos entre os pontos de origem e destino das mensagens a serem enviadas, determinando que um caminho seja estabelecido antes do

envio de uma mensagem. Esta conexão permanece dedicada até o fim da comunicação, constituindo a condução da mensagem orientada à conexão. Deste modo, os circuitos e seus recursos de chaveamento e transmissão encontram-se destinados aos usuários até o momento em que um ou ambos os extremos da conexão decidam desfazê-la. Esta estratégia acarreta que (i) a capacidade de transmissão da conexão esteja sempre disponível quando houver necessidade de comunicação; (ii) que o atraso na transferência da mensagem seja constante; (iii) que a entrega das mensagens ocorra na mesma ordem em que foram enviadas. Entretanto, a comutação de circuitos não utiliza de forma eficiente os recursos da rede, pois estes permanecem dedicados enquanto a conexão existir, independentemente da ocorrência de transmissão de mensagem. Uma outra questão encontra-se relacionada com o tempo de espera para o estabelecimento de uma conexão, que pode ser elevado, principalmente em casos de chamadas a longa distância ou internacionais e/ou em períodos de carga elevada. Assim também, quando o tráfego atinge um determinado nível considerado elevado, requisições de novas mensagens podem ser recusadas devido à falta de recursos disponíveis na rede, uma vez que todas as conexões estão alocadas, mesmo que neste momento não esteja ocorrendo transmissão de mensagens. Além disso, esta tecnologia de comutação não permite a alocação flexível da largura de banda das conexões, sendo esta fixa durante toda a transmissão e apresentando capacidade máxima de 64 Kbps.

2.4 – Redes de Comutação de Pacotes

A tecnologia de chaveamento de pacotes é mais recente que a de circuitos, tendo surgido na década de 60 para transmissão de dados (Steenstrup, 1995). A evolução desta técnica de comutação nas últimas décadas se deve aos avanços ocorridos na tecnologia de redes de computadores. Desta forma, o número de usuários das redes de comutação de pacotes sofreu um acréscimo extraordinário nos últimos tempos e as previsões são de que cada vez mais pessoas utilizem computadores. Assim, estas redes têm sido alvo de constantes pesquisas, a fim de se obterem equipamentos com menor custo e maior qualidade, promovendo o atendimento adequado ao crescente número de usuários.

A tecnologia de comutação de pacotes realiza o encapsulamento da mensagem para envio em unidades menores, denominadas pacotes. Estas unidades possuem um comprimento máximo determinado, permitindo que o atraso no envio da mensagem e o custo de transmissão da mesma sejam minimizados. Assim também, a velocidade de transmissão entre os terminais é otimizada. Os pacotes transportam, além da mensagem a ser enviada, informações sobre sua origem e seu destino.

Estas redes não são orientadas à conexão, ou seja, não há necessidade de alocar um canal por todo o caminho a ser percorrido e mantê-lo dedicado durante toda a transmissão. Logo, cada pacote utiliza somente o canal da conexão que está percorrendo no momento e torna este canal disponível após sua transmissão. Esta estratégia permite o emprego mais eficiente dos recursos da rede, pois os dados utilizam uma conexão somente quando necessário. Se a conexão encontra-se ocupada, o pacote deve ser armazenado na memória do roteador e esperar até encontrar uma conexão disponível. Desta forma, há uma espécie de fila de espera, onde as mensagens são depositadas na ordem em que foram recebidas, até que a capacidade máxima de armazenamento do roteador seja atingida.

Existem dois métodos de transmissão dos pacotes pela rede: diagramas de dados e circuitos virtuais. No primeiro método, cada roteador presente na rede é responsável pela seleção do próximo a ser visitado pelo pacote. Assim, pacotes que pertencem a uma mesma mensagem podem percorrer caminhos diferentes para alcançar o destino desejado. Logo, é possível que estes pacotes atinjam o destino em uma ordem diferente daquela em que foram gerados e/ou transmitidos. O ordenamento na recepção dos pacotes requer mecanismos de controle para garantir a entrega correta dos mesmos. A utilização de rotas diferentes pelos pacotes pode ocasionar atrasos não compatíveis com o tipo de aplicação envolvida, como tráfego de voz ou vídeo, o que constitui uma limitação apresentada por este método.

No método de transmissão baseado em circuitos virtuais, o terminal de origem da mensagem seleciona a rota a ser percorrida pela mesma ao longo da rede. Entretanto, o caminho não permanece dedicado a esta mensagem, como ocorre nas redes de comutação de circuitos. Como na transmissão através de diagramas de dados, os pacotes são armazenados em uma fila de espera do roteador, caso a conexão determinada se encontre ocupada.

As redes de comutação de pacotes utilizam protocolos com maior grau de complexidade que as redes de comutação de circuitos, o que provoca um aumento de custos em sua implementação.

2.5 – Redes de Alta Velocidade

As redes de alta velocidade foram desenvolvidas para suportar os múltiplos serviços requeridos pelo grande número de usuários das redes de telecomunicações atuais, tais como transporte de tráfego de voz, dados e vídeo. A partir de pesquisas realizadas em redes de comutação de circuitos e de pacotes, surgiu uma tecnologia de transmissão e chaveamento denominada modo de transferência assíncrono (*Asynchronous Transference Mode* - ATM), (Soares et al. 1995), que reúne as melhores características apresentadas por cada uma destas.

Esta tecnologia de transmissão foi desenvolvida para as redes digitais de serviços integrados, (*Broadband Integrated Services Digital Network* - B-ISDN) para suceder o modo de transmissão síncrono (*Synchronous Transference Mode* - STM), anteriormente empregado.

As características apresentadas por esta tecnologia de transmissão pertinentes às redes de comutação de circuitos estão relacionadas com o estabelecimento de um circuito virtual entre os pontos de origem e destino de uma mensagem antes do envio da mesma. Deste modo, o atraso no envio da informação é constante e a capacidade de transmissão encontra-se sempre disponível. Assim também, a ordem em que as células foram geradas e/ou transmitidas é mantida durante sua transmissão. Porém, o circuito virtual é utilizado somente durante a transmissão da mensagem, e não durante todo o tempo em que a conexão existir, como ocorre nas redes de comutação de circuitos. Os aspectos pertinentes às redes de comutação de pacotes utilizados pela tecnologia de transferência assíncrona são relativos ao encapsulamento da mensagem em unidades menores. Estas unidades são denominadas células e apresentam um tamanho fixo de 53 bytes, menor do que o das unidades utilizadas nas redes de comutação de pacotes. Desta forma, os atrasos no envio das células é reduzido.

Uma conexão é estabelecida nas redes de alta velocidade somente quando existem recursos suficientes para atenderem aos requisitos de qualidade de serviço (*Quality of Service Requirements* – *QoS requirements*) estabelecidos pelos usuários. Estes requisitos estão relacionados com os seguintes parâmetros das conexões: atraso na transferência de células (*Cell Transfer Delay* - CTD), sendo constituído pela soma dos atrasos de transmissão, propagação e processamento das mesmas; variação deste atraso (*Cell Delay Variation* - CDV), perda de células (*Cell Loss Ratio* - CLR) e taxa média de geração de células em uma conexão (*Sustainable Cell Ratio* - SCR).

Um dos fatores fundamentais para a utilização das redes de alta velocidade foi o desenvolvimento das fibras óticas, que são mídias de transmissão com capacidade de envio de um número elevado de dados, atingindo velocidades da ordem de Gigabits por segundo, enquanto as redes de comutação de pacotes apresentam capacidade de no máximo 45 Megabits por segundo. As características de transmissão apresentadas pelas fibras óticas possibilitou a utilização de conexões de alta velocidade, permitindo a transmissão de um grande volume de informação em um pequeno espaço de tempo.

2.6 – Redes Móveis

As redes móveis são aquelas nas quais os usuários se movimentam em relação às estações de chaveamento (Steenstrup, 1995). Durante o movimento, a comunicação entre os

usuários pode ser mantida ou suspensa, de acordo com a capacidade da rede móvel e da natureza da comunicação.

A tecnologia de redes móveis existe há mais de vinte anos, mas somente na década passada se tornou comercialmente popular. Com computadores cada vez mais portáteis e redes mais acessíveis, os usuários estão demandando das redes móveis os mesmos serviços que costumavam obter das redes com linhas físicas estacionárias. Assim como estas, as redes móveis se desenvolveram ao longo de trajetórias paralelas de chaveamento de circuitos e de pacotes, sendo telefonia celular e *packet radio* representantes de cada uma destas tecnologias de chaveamento. As redes móveis de telefonia celular se tornaram o tipo mais predominante deste tipo de tecnologia, pois provêem serviços de telefonia independente da localização dos usuários. As redes móveis de *packet radio* possuem aplicações militares, pois possibilitam comunicações de rádio robustas em topologias de rede altamente dinâmicas.

A maioria dos dispositivos móveis se baseia em conexões via rádio para se comunicarem com o restante da rede. A qualidade deste tipo de conexão pode flutuar consideravelmente sob variações das condições da rede e com relação às características do ambiente em que o usuário se encontra. Muitos destes fatores podem ser atenuados por ajustes na potência de transmissão ou por protocolos de controle de conexões, mas não podem ser inteiramente eliminados. Assim, as redes móveis devem acomodar não somente usuários móveis, como também conexões altamente variáveis.

Prover múltiplos tipos de serviços, garantindo a qualidade dos mesmos, constitui um desafio e o principal tópico da pesquisa atual tanto em redes estacionárias como móveis.

2.7 - Conclusão

Este capítulo descreveu as características gerais das redes de telecomunicações, bem como os aspectos fundamentais das principais tecnologias de chaveamento e transmissão da informação utilizadas: redes de comutação de circuitos, redes de comutação de pacotes, redes de alta velocidade e redes móveis. As características das redes de telecomunicações descritas compreendem: componentes destas redes e suas funções realizadas, meios físicos de transmissão, modos de transmitir o tráfego de informação, tipos de sinais transmitidos e de redes empregadas, além dos serviços atualmente oferecidos pelas redes de telecomunicações. Os aspectos fundamentais das tecnologias de rede descritos estão relacionados com a evolução do modo de chaveamento e transmissão da informação utilizados por estas redes, que são dependentes do tipo de tráfego enviado, bem como dos

serviços oferecidos pelas mesmas. As vantagens e limitações apresentadas por cada uma destas tecnologias foram discutidas e analisadas.

Os aspectos das redes de telecomunicações pertinentes às suas características e funções desempenhadas pelos seus componentes serão úteis no entendimento do problema do roteamento e de como os roteadores se relacionam com os demais componentes de uma rede. Assim também, os aspectos das tecnologias de chaveamento e transmissão da informação relacionados com as características de cada rede e do tráfego transmitido pela mesma, serão incorporados como restrições à operação do sistema de roteamento desenvolvido.

A partir das características das redes de telecomunicações e das principais tecnologias apresentadas, o próximo capítulo irá descrever o problema de roteamento neste ambiente.

3 – Roteamento em Redes de Telecomunicações

Neste capítulo é descrito o problema do roteamento em redes de telecomunicações. São apresentados as principais propriedades deste problema, as funções do roteamento, os fatores utilizados na determinação das rotas, as características a serem apresentadas pelos sistemas de roteamento e suas classificações.

3.1 – Características Gerais

O roteamento em uma rede de telecomunicações é o mecanismo responsável pela escolha da rota a ser percorrida por uma mensagem e por sua transmissão ao longo da rede (Bellman, 1958; Steenstrup, 1995). O roteamento realiza um conjunto complexo de procedimentos interdependentes, determinados por objetivos e restrições conflitantes. Seu objetivo é encontrar o melhor caminho com relação a uma métrica previamente estabelecida, no sentido de promover o equilíbrio entre as necessidades de serviço do tráfego e as restrições de operação da rede. Dentro deste contexto, o roteamento deve maximizar o desempenho da rede e minimizar os custos envolvidos no processo de transferência da mensagem. As restrições são impostas pelo tipo de tecnologia de comutação e transmissão da informação utilizadas, pelas características dinâmicas da rede e do tráfego, como também pelos serviços oferecidos pela rede e aqueles requeridos pelos usuários. Deste modo, Bertsekas & Gallager (1992) argumentam que implementar um algoritmo de roteamento que encontre os melhores caminhos em uma rede de telecomunicações constitui um problema de otimização de múltiplos objetivos e múltiplas restrições.

Nesta pesquisa, o problema de roteamento em uma rede de telecomunicações é representado por um grafo não orientado $G = (R, T)$, onde os nós pertencentes ao conjunto R representam os roteadores e os laços pertencentes ao conjunto T representam as linhas de transmissão das mensagens (Bertsekas & Gallager, 1992). A Figura 2 apresenta um exemplo de uma rede de telecomunicações simples, composta por sete roteadores, identificados por letras de A-G, e 9 linhas de transmissão, rotuladas por números de 2-5, que representam a distância geográfica existente entre os roteadores. Estes últimos são caracterizados por sua capacidade de processamento e armazenamento de mensagens, enquanto cada linha que os interconecta se caracteriza por sua capacidade de transmissão. O sistema de transmissão é

bidirecional e os nós aos quais um nó se encontra diretamente conectado são considerados seus vizinhos (Schoonderwoerd et al., 1996).

Figura 2 - Exemplo de rede simples de telecomunicações, com 7 nós e 9 laços.

As redes de telecomunicações constituem ambientes dinâmicos e estocásticos. Isto significa que as condições de operação da rede e as características do tráfego podem sofrer variações ininterruptas ao longo do tempo, em muitos momentos de forma imprevista e inesperada. Deste modo, uma mesma ação realizada em instantes de tempo diferentes pode conduzir a rede a situações diversas. Assim também, o grau de eficácia de uma determinada ação é variável ao longo do tempo.

Dentro deste contexto, o problema de roteamento em uma rede de telecomunicações apresenta como principais características (Di Caro & Dorigo, 1998):

♦ Funções distribuídas e limitações de tempo

Na verdade, a estrutura de dados sobre a rede e as decisões de roteamento são distribuídas entre todos os roteadores da rede, que realizam suas funções de forma interdependente. As decisões devem ser realizadas rapidamente, pois os atrasos ocorridos na propagação de informação sobre a rede podem afetar os padrões de tráfego da mesma. Além disso, não é possível obter informação global completa e atualizada sobre o estado da rede e do tráfego a cada instante. Assim, os roteadores selecionam as rotas baseados em informação local atualizada obtida pelos mesmos, combinada àquela atrasada recebida de seus vizinhos.

♦ Condições estocásticas e não-estacionárias

A quantidade de mensagens geradas, sua duração ou tamanho (no caso de redes de circuitos e de pacotes, respectivamente), juntamente com os efeitos provocados nas condições de operação da rede, são estocásticos e não-estacionários. Deste modo, não se dispõe de um modelo prévio e completo sobre o comportamento da rede.

♦ Múltiplos objetivos

As redes de telecomunicações apresentam como objetivos maximizar o desempenho da rede e minimizar seus custos de operação. Além disso, diversas medidas de desempenho devem ser consideradas de acordo com a tecnologia de rede utilizada. Como exemplo, as redes de comutação de circuitos em geral possuem como medidas de desempenho a porcentagem de chamadas aceitas na rede e o tempo médio de espera para aceitar ou recusar uma chamada. Já as redes de comutação de pacotes apresentam a quantidade de bits entregue por unidade de tempo, em conjunto com os atrasos ocorridos na transmissão dos pacotes. Além disso, existem outras medidas que levam em consideração os efeitos causados pelo sistema de roteamento nos recursos da rede com relação a memória requerida para seu processamento.

♦ Múltiplas restrições

As restrições típicas de um sistema de roteamento estão relacionadas com a tecnologia de rede utilizada, os serviços por ela oferecidos e aqueles requeridos pelos usuários. Estes requerem atendimento rápido, seguro, diversificado, de baixo custo e alta qualidade. Os provedores de serviços, por sua vez, procuram atender as necessidades dos usuários de acordo com fatores econômicos e tecnológicos, maximizando uma determinada medida de desempenho.

Os vários algoritmos de roteamento existentes possuem determinadas funções em comum (Steenstrup, 1995), que são: (i) a reunião e distribuição da informação com respeito ao estado da rede e às características do tráfego; (ii) a seleção e geração das rotas a serem percorridas pela mensagem e (iii) a condução da mensagem através dos caminhos selecionados, otimizando uma dada medida de desempenho.

O estado da rede está relacionado com os recursos utilizados, os serviços oferecidos e as restrições ao uso destes. As características do tráfego compreendem os serviços requeridos e a localização dos usuários. Estas informações podem ser constituídas a partir de valores medidos e/ou previstos pela própria rede, como também por fontes externas. As informações são codificadas e armazenadas em tabelas existentes em cada nó da rede. Os valores armazenados nas tabelas orientam o sistema de roteamento na seleção das rotas. Os sistemas podem utilizar apenas uma métrica para selecionar a melhor rota a ser percorrida, assim como uma combinação de várias métricas expressa por uma variável escalar. As métricas mais comuns são as seguintes (Downes et. al., 1998):

♦ Comprimento do caminho

O comprimento do caminho pode ser determinado pelo número de roteadores intermediários a serem visitados pela mensagem e/ou pela distância geográfica entre seus pontos de origem e destino.

♦ Atraso

O envio de uma mensagem envolve um determinado período de tempo entre sua emissão do nó origem e recepção no nó destino, denominado atraso. O atraso ocorre nas seguintes fases: espera na fila dos roteadores existentes ao longo do caminho, processamento e transmissão da mensagem através da mídia utilizada e sua propagação ao longo de uma conexão.

♦ Largura de banda

As linhas de transmissão possuem uma capacidade máxima de transferência de dados por período de tempo, caracterizada pelo número de bits de informação transmitidos por segundo (bps ou b/s). A seleção das rotas é realizada de forma que as linhas que apresentem maior largura de banda sejam preferencialmente utilizadas, para que o atraso no envio da mensagem seja menor.

♦ Carga

A carga se refere ao grau de ocupação dos recursos da rede, ou seja, o número de mensagens processadas por um roteador ou conexão ao longo do tempo. Deste modo, as linhas de transmissão e roteadores que se encontram processando menor número de mensagens devem ser mais utilizados que os demais.

A partir da informação sobre a rede, o algoritmo de roteamento seleciona as possíveis rotas, que podem ser ótimas ou apenas viáveis. As rotas viáveis são aquelas que atendem os requisitos de serviço dos usuários e os da rede, enquanto as rotas ótimas são as melhores rotas viáveis de acordo com um ou mais critérios de desempenho determinados. É necessário salientar o balanço a ser realizado entre a seleção de conexões que possuam maior largura de banda e aquelas que se encontram processando um número baixo de mensagens. Isto decorre do fato de que ao utilizar somente as conexões que apresentam maior capacidade de transmissão, o sistema de roteamento pode conduzi-las ao congestionamento. Em muitas ocasiões, o aumento no atraso resultante do envio da mensagem por uma conexão com menor capacidade de transmissão, mas com menor número de mensagens em processamento,

causa menores prejuízos ao desempenho global da rede que a utilização de uma conexão com maior largura de banda, mas que se encontra altamente carregada no momento. Após selecionar a rota a ser percorrida, o algoritmo de roteamento realiza a condução do tráfego ao longo do caminho selecionado. Esta condução pode ser realizada através de dois paradigmas: comutação de circuitos e comutação de pacotes.

A fim de lidar de forma eficiente com as características dinâmicas e estocásticas das redes de telecomunicações, os sistemas de roteamento necessitam possuir uma ou mais das características descritas a seguir (Downes et. al., 1998):

♦ Robustez

A capacidade de executar corretamente as funções de roteamento sob condições imprevistas e inesperadas, como condições de falhas ou faltas em equipamentos, é primordial para garantir o atendimento rápido e preciso aos usuários de uma rede de telecomunicações.

♦ Simplicidade

A utilização dos recursos computacionais deve ser mínima e a operação do sistema de roteamento deve ser simples, de forma que o administrador da rede o compreenda facilmente.

♦ Otimização da rota

O algoritmo deve ser capaz de selecionar a melhor rota, de acordo com a métrica considerada, sob quaisquer condições da rede e do tráfego.

♦ Convergência Rápida

Os algoritmos de roteamento devem convergir rapidamente para as rotas ótimas, principalmente diante de variações da rede. A convergência lenta dos roteadores pode causar *loops* de roteamento e afetar seriamente o desempenho da rede.

♦ Flexibilidade

Os algoritmos de roteamento também devem ser flexíveis, adaptando-se corretamente e rapidamente a uma variedade de circunstâncias, como variações da largura de banda das conexões e da capacidade de processamento disponível dos nós.

3.2 – Classificações de Sistemas de Roteamento

Os sistemas de roteamento podem ser classificados de acordo com os seguintes aspectos:

- ♦ O dispositivo responsável pela escolha das rotas: centralizados ou descentralizados;
- ♦ O tipo de mudança ocorrida no estado da rede e/ou nas características do tráfego a que respondem e a velocidade de resposta, bem como a informação utilizada para gerar e selecionar as rotas: estáticos ou dinâmicos;
- ♦ O número de caminhos selecionados para um mesmo par origem-destino da mensagem: único caminho ou múltiplos caminhos;
- ♦ A localização do roteador que irá determinar a rota a ser percorrida: roteamento determinado pela fonte ou por cada roteador pertencente a rota;
- ♦ A existência ou não de uma hierarquia de roteamento: hierárquicos ou não-hierárquicos.

A partir deste ponto, os sistemas de roteamento serão, portanto, discutidos de acordo com a classificação acima.

As funções de roteamento podem ser implementadas de forma centralizada ou descentralizada (Steenstrup, 1995). O nível de descentralização está intimamente ligado ao dinamismo, robustez e gerenciamento desejados para o sistema de roteamento. Em um algoritmo de roteamento centralizado, a determinação das rotas é responsabilidade de um único roteador, denominado Centro de Roteamento, conectado a todos os outros roteadores da rede, como a topologia em estrela apresentada na Figura 1(d). Assim, este dispositivo obtém de cada um dos nós as informações a respeito do estado de suas conexões e carga, determina as rotas a serem utilizadas e as envia aos roteadores, para que suas tabelas de roteamento sejam atualizadas. Este procedimento simplifica as funções de gerenciamento do tráfego da rede e reduz custos, pois as funções de roteamento são realizadas em apenas um dispositivo, além de possibilitar decisões precisas sobre o cálculo das rotas.

Entretanto, se a rede for constituída por um grande número de roteadores, a seleção das rotas sob condições de tráfego variável deve ser realizada de forma rápida e coerente, exigindo alta carga de processamento do roteador central (Soares et al., 1995). Deste modo, a implementação centralizada produz melhores resultados em redes que não apresentam mudanças frequentes em suas condições de operação. Uma outra limitação apresentada por esta classe de algoritmos se reflete na ocorrência de uma falha ou falta no roteador central, o

que impossibilita a realização das funções de roteamento. Além disso, deve-se estar atento à geração de rotas inconsistentes, uma vez que cada roteador recebe a informação a respeito das rotas em momentos diferentes, devido ao atraso sofrido em cada uma das transmissões realizadas.

Na alternativa descentralizada, vários roteadores executam os cálculos das rotas, de forma reproduzida ou distribuída. Na primeira situação, as funções de roteamento são realizadas por cada dispositivo de modo independente, sem transmitir os resultados obtidos aos outros roteadores, enquanto na segunda cada um deles realiza suas funções e transmite os resultados a todos os outros. Esta transmissão pode ser síncrona ou assíncrona. Apesar de aumentar a complexidade do gerenciamento do sistema de roteamento, a implementação descentralizada apresenta algumas vantagens. Ao reproduzir as funções de roteamento, ela possibilita maior tolerância a faltas ou falhas em nós ou conexões, além da redução do atraso relativo ao processamento da mensagem. Por outro lado, ao distribuir as funções de roteamento, ela reduz os recursos necessários a cada um dos roteadores para o estabelecimento das rotas em relação à implementação centralizada e permite que o sistema acompanhe rapidamente mudanças na rede.

Os algoritmos de roteamento estáticos são caracterizados pela seleção de rotas fixas baseadas na previsão do tráfego e em uma topologia fixa de rede, sem levar em consideração o estado atual da mesma e do tráfego. As rotas são selecionadas de forma que constituam o menor caminho entre a origem e o destino da mensagem, de acordo com um determinado critério de custo. Novas rotas são utilizadas em casos de falta ou falha em um nó ou conexão pertencentes a rota previamente definida. Esta classe de algoritmos de roteamento apresenta facilidade de implementação em redes onde o tráfego é relativamente previsível e estacionário. Porém, ela apresenta perda de eficiência sob condições imprevistas de variação do tráfego, devido à impossibilidade de adaptação rápida a estas situações. Este tipo de roteamento era amplamente utilizado até meados da década de 80, devido às limitações dos equipamentos e tecnologia de redes disponíveis neste período.

O roteamento dinâmico calcula as rotas de acordo com as condições atuais da rede e do tráfego. Ele pode ser realizado de três formas: dependente do tempo, dependente do evento e adaptativo (DaSilva, 1999). A seleção dos caminhos pode ser realizada em determinados horários e permanecer fixa durante intervalos de tempo, o que ocorre nos algoritmos dependentes do tempo. Os algoritmos dependentes do evento atualizam a informação armazenada nas tabelas de roteamento de acordo com o resultado produzido pelo uso de um caminho com relação à medida de desempenho utilizada. Os algoritmos adaptativos, por sua vez, selecionam as rotas de acordo com as características correntes da

rede e do tráfego. O nível adequado de dinamismo de um algoritmo de roteamento está relacionado com o tipo e a frequência das mudanças ocorridas na rede e no tráfego, assim também com os recursos de processamento, memória e transmissão disponíveis na rede, entre outros fatores. Este tipo de roteamento vem sendo amplamente desenvolvido e utilizado por diversas empresas de telecomunicações localizadas na Europa, Estados Unidos e Japão, entre outras (Ash, 1995, 1997).

Alguns algoritmos de roteamento são capazes de enviar diversas mensagens entre um mesmo par origem-destino por caminhos distintos, enquanto outros selecionam apenas um caminho disponível. A multiplexação estatística de mensagens permite que várias informações compartilhem a capacidade de transmissão de uma mesma conexão, favorecendo a transmissão da informação por caminhos diferentes. Em contraposição, as técnicas de multiplexação tradicionais alocam previamente um espaço fixo para cada uma das mensagens a serem enviadas, independente da existência de informação a ser transmitida. Nestes casos, realiza-se a transmissão de mensagens pertencentes a um mesmo par de nós origem-destino através de uma única rota.

Existem algoritmos de roteamento em que o roteador que constitui a origem da mensagem é o responsável pela seleção da rota ao longo da rede; eles são denominados algoritmos de roteamento determinado pela origem (*source-routing*). Neste caso, os outros roteadores pertencentes à rota somente armazenam a informação e a enviam ao vizinho previamente estabelecido. Este processo previne o surgimento de *loops* de roteamento e em geral encontra as melhores rotas, pois possui informação sobre todos os caminhos possíveis que conduzem ao destino em questão. Em contrapartida, esta estratégia requer grande quantidade de tempo e memória de processamento. Ela é em geral utilizada por redes de comutação de circuitos.

Em outros algoritmos, os cálculos das rotas são realizados por cada roteador em particular, baseados nas informações armazenadas em suas tabelas (*hop-by-hop routing*). Esta classe de algoritmos deve dispor de um método padrão de seleção das rotas em cada roteador, a fim de se evitar *loops* de roteamento. Este procedimento é empregado em redes de comutação de pacotes, uma vez que possibilitam um baixo nível de processamento em cada roteador pertencente à rota.

Em um algoritmo hierárquico, a rede é dividida em regiões e os roteadores transmitem informações entre si sobre o estado das conexões e a topologia da região a que pertencem (Akinpelu, 1984; Ash, 1995, 1997). Isto pode ser exemplificado pelo uso dos códigos de áreas utilizados nas redes telefônicas. Em cada uma das regiões é estabelecida uma hierarquia, onde um roteador é selecionado e torna-se o líder do grupo, sendo responsável

pela reunião e distribuição das informações sobre sua região aos outros roteadores que a compõem, como também aos líderes dos demais grupos existentes na rede. Este processo se repete em todos os níveis hierárquicos estabelecidos; desta forma, à medida que os roteadores encontram-se em níveis superiores, eles selecionam as rotas de uma quantidade maior de tráfego, sendo o roteamento realizado de forma progressiva em direção ao destino. Em geral, as redes são divididas em dois níveis hierárquicos, mas podem apresentar também múltiplos níveis. Os algoritmos hierárquicos são utilizados em grandes redes de telecomunicações, onde a quantidade de informação necessária ao roteamento é muito grande, o que torna o processo de seleção das rotas extremamente oneroso. Nestas situações, ele proporciona a redução do volume de informação necessária ao processamento dos roteadores; entretanto, não há garantia de que a rota selecionada seja ótima.

Em uma estrutura não-hierárquica, todos os roteadores encontram-se em um único nível e desempenham as mesmas funções, transmitindo as mensagens sobre suas conexões entre si, calculando as rotas baseados na informação recebida e atualizando suas tabelas de roteamento (Akinpelu, 1984; Ash, 1995, 1997).

3.3 – Conclusão

Neste capítulo foram descritas as principais características do problema de roteamento em redes de telecomunicações, envolvendo sua representação, funções desempenhadas por um sistema de roteamento, fatores utilizados na determinação das rotas, características a serem apresentadas pelos sistemas de roteamento e suas classificações.

O conhecimento das principais propriedades do problema de roteamento apresentadas neste capítulo será fundamental para a implementação do algoritmo a ser proposto, possibilitando a familiarização com o ambiente das redes de telecomunicações e com as características a serem apresentadas por este sistema, no sentido de lidar corretamente com as condições atuais das redes.

Assim, após a descrição dos princípios básicos do roteamento em redes de telecomunicações, serão apresentados no próximo capítulo os aspectos fundamentais do método de aprendizagem por reforço, que vem sendo aplicado com sucesso ao problema de roteamento, seguidos pela principal técnica pertencente a esta estratégia.

4 - Aprendizagem por Reforço

Neste capítulo são apresentados inicialmente os princípios básicos de uma abordagem de aprendizagem de máquinas baseada na interação entre um agente e um meio ambiente: aprendizagem por reforço. A seguir, são descritas as características principais da aprendizagem baseada em diferença temporal, entre os quais se destaca a aprendizagem-Q e da aprendizagem por reforço dual, que vem sendo aplicadas com sucesso a sistemas de roteamento. As demais estratégias que compõem os métodos de diferença temporal, TD(λ), Sarsa(0), Sarsa(λ), Q(λ) de Watkins e Q(λ) de Peng & Williams, são descritos no Apêndice A, uma vez que eles não são utilizados diretamente neste projeto de pesquisa.

4.1 – Definição

Na aprendizagem por reforço, o aprendizado ocorre através da interação entre um elemento aprendiz, denominado agente, que seleciona as ações a serem realizadas, e o meio ambiente em que ele se encontra (Kaelbling & Littman, 1996; Sutton & Barto, 1998). Russel & Norvig (1995) definem um agente como uma entidade capaz de perceber seu meio ambiente e agir sobre ele. Um agente é denominado autônomo quando a escolha das ações a serem realizadas é dependente de sua própria experiência. Um agente pode ser: (i) baseado em reflexos, responde de imediato a uma nova percepção; (ii) baseado em metas, age para atingir uma meta e (iii) baseado em uma utilidade, age guiado por uma função de utilidade. Singh (1994) argumenta que para que o agente realize suas ações de forma autônoma, ele necessita ser capaz de aprender. Na aprendizagem por reforço, a interação do agente com seu meio ambiente visa alcançar um determinado objetivo.

A estrutura do método de aprendizagem por reforço apresenta três conjuntos (Kaelbling et al., 1996):

- ♦ Um conjunto discreto de estados do ambiente, S ;
- ♦ Um conjunto discreto de ações a serem selecionadas e realizadas pelo agente, A ;
- ♦ Um conjunto de valores escalares que representam os sinais de reforço do ambiente, $\mathcal{R}$;

Estes três conjuntos são utilizados pelos quatro componentes principais da aprendizagem por reforço:

- ♦ Uma política $\pi: S \rightarrow A$, que mapeia um determinado estado ($s \in S$) do ambiente em uma ação ($a \in A$). A política define o comportamento do agente a cada momento.
- ♦ Uma função de reforço $r: S, A \rightarrow \mathbb{R}$, que mapeia estados e ações do ambiente em um valor escalar real, denominado sinal de reforço. O reforço determina a resposta real do ambiente para cada ação do agente.
- ♦ Uma função de avaliação que estima uma função da soma dos sinais de reforço obtidos ao longo de todos os estados que o agente deve visitar a partir do estado atual até atingir a meta. Esta função pode mapear estados em valores, $V: S \rightarrow \mathbb{R}$ ou pares estado-ação em valores $Q: S, A \rightarrow \mathbb{R}$. A função de avaliação especifica quais ações são mais adequadas a longo prazo.
- ♦ Um modelo do ambiente que consiste em uma função de probabilidade de transição de estado $T(s_t, a_t, s_{t+1})$, que representa a probabilidade do agente se mover de um estado s_t para um estado s_{t+1} através de uma ação a_t . O modelo procura simular o comportamento do ambiente. O modelo do ambiente não é obrigatório, ao contrário dos outros três componentes.

Na classe de modelos de aprendizagem por reforço considerados nesta dissertação, a interação entre o agente e o ambiente é definida como um processo de decisão Markoviano (PDM). Em um PDM, a probabilidade que o sistema esteja em um dado estado após n passos é determinada conhecendo-se apenas o estado e a última ação do passo imediatamente anterior (Clarke & Disney, 1979). Deste modo, o conhecimento do estado em que o ambiente se encontra no momento e da última ação realizada será suficiente para prever as probabilidades de transição para o estado futuro do mesmo.

Em cada passo da interação t entre o agente e o meio, ocorrem os seguintes eventos:

1. O agente recebe como entrada uma indicação do estado atual do ambiente, s_t e o sinal de reforço r_t ;
2. O agente seleciona e realiza uma ação a_t , que é gerada como saída;
3. Um passo de tempo após a realização da ação, o agente recebe um sinal de reforço escalar r_{t+1} , em consequência da ação realizada e da modificação do estado do ambiente para s_{t+1} .

4. Baseado neste sinal, o agente atualiza a estimativa do valor do estado s_t ou do par estado-ação (s_t, a_t) selecionado.

Figura 3 – Representação gráfica da aprendizagem por reforço.

O agente busca obter informação sobre qual ação maximiza a soma dos sinais de reforço obtidos a longo prazo. Isto geralmente é realizado através de tentativa e erro, podendo ser utilizada uma série de algoritmos. No caso geral do problema de aprendizagem por reforço, as ações selecionadas não afetam somente o sinal de reforço imediato, mas também os estados futuros que o ambiente visitará. Desta forma, o agente pode realizar uma sequência de ações sem receber nenhum sinal de reforço e finalmente atingir um estado que apresente este sinal (reforço atrasado). Portanto, o agente deve aprender quais ações são desejáveis no presente de maneira a otimizar o reforço a ser obtido no futuro.

Os métodos de aprendizagem por reforço podem ser divididos em duas classes principais com relação a utilização de um modelo do meio ambiente:

- a) Baseada em um modelo do ambiente: O agente possui conhecimento prévio sobre o comportamento do ambiente. Utiliza-se este modelo em casos de planejamento, em que a ação é escolhida considerando situações futuras que nunca ocorreram antes.
- b) Não utiliza um modelo do ambiente: O agente aprende a atuar no meio sem possuir conhecimento prévio sobre o mesmo, através de tentativa e erro.

Um agente de aprendizagem por reforço é caracterizado principalmente pela exploração do meio ambiente em que atua. Esta necessidade gera uma dicotomia entre exploração para obtenção de possíveis soluções (*exploration*) e exploração para melhoria de uma solução promissora (*explotation*). Isto é, quanto um ambiente deverá ser investigado em contraposição a quanto uma alternativa promissora de solução deverá ser mais utilizada e

aperfeiçoada. Algumas das questões que determinam o compromisso entre estes dois tipos de exploração são: o tempo de resposta requerido para o agente, o tempo em que o agente deverá concluir sua tarefa, as consequências de convergências prematuras para um sub-ótimo e a capacidade de exploração do agente.

A avaliação da qualidade do aprendizado obtido por um método de aprendizagem por reforço pode ser realizada considerando (Kaelbling et al., 1996):

- a) Convergência para o ótimo: De acordo com Watkins & Dayan (1992), alguns algoritmos possuem garantia de convergência assintótica para o comportamento ótimo. Porém, aqueles que atingem rapidamente um nível elevado de otimização da solução, como 99%, serão mais utilizados do que outros que também possuem a garantia de convergência, entretanto inicialmente apresentam uma taxa de aprendizado baixa.
- b) Velocidade de convergência para o comportamento ótimo: Como o comportamento ótimo é geralmente um resultado assintótico, mede-se a velocidade de convergência para um comportamento próximo do ótimo, de acordo com o problema em questão. Uma forma de verificar este fator é medir o nível de desempenho do algoritmo após um determinado período de tempo.
- c) Arrependimento: É o decréscimo esperado no retorno obtido, através da execução do algoritmo de aprendizagem desde o início do processo de exploração, ao invés de seguir a estratégia ótima, supondo-a já conhecida (Berry & Fristedt, 1985). Esta medida penaliza erros ocorridos ao longo do percurso do agente.

Para que o agente aprenda uma política ótima, é necessário decidir qual será o modelo de otimização utilizado, ou seja, de que forma o agente levará o futuro em consideração na seleção das ações no presente. Existem três métodos para considerar o retorno do meio ambiente que são usualmente empregados em problemas de aprendizagem por reforço:

1. Modelo de horizonte finito: Considera que o agente otimiza o reforço total esperado para os próximos h passos:

$$E\left(\sum_{t=0}^h r_t\right) \quad (1)$$

no qual $E(.)$ representa o valor esperado do reforço r_t considerando t passos no futuro.

2. Modelo de horizonte infinito descontado: Calcula-se o retorno de longo prazo, onde os reforços recebidos no futuro são ponderados utilizando-se um fator de desconto γ :

$$E\left(\sum_{t=0}^{\infty} \gamma^t r_t\right) \quad (2)$$

no qual $0 \leq \gamma < 1$.

3. Modelo de recompensa média: O agente procura otimizar o reforço médio de longo prazo:

$$\lim_{h \rightarrow \infty} E\left(\frac{1}{h} \sum_{t=0}^h r_t\right) \quad (3)$$

Ao empregar o modelo de horizonte infinito descontado, o valor de um estado s_t sob uma política π é denotado por $V_{\pi}(s_t)$. Este é definido como o reforço esperado ao se iniciar o processo no estado s_t e seguir a política π (Sutton & Barto, 1998). Assim, a função de avaliação $V_{\pi}(s_t)$ do estado s_t é o valor instantâneo do reforço esperado somado ao valor de reforço esperado do próximo estado descontado, ao executar a política π .

$$V_{\pi}(s_t) = [r(s_t) + \gamma \sum_{s_{t+1} \in \mathcal{S}} T(s_t, s_{t+1}) V_{\pi}(s_{t+1})] \quad (4)$$

na qual $T(s_t, s_{t+1})$ = probabilidade de transição do estado s_t para o estado s_{t+1}

Do mesmo modo, a função de avaliação do par estado-ação (s_t, a_t) para a política π é o valor de uma ação a_t realizada em um estado s_t sob uma política π , denotada por $Q_{\pi}(s, a)$ (Sutton & Barto, 1998). Este valor é definido como o retorno esperado ao se iniciar a trajetória do agente em um estado s_t , realizar uma ação a_t e a partir desta seguir a política π (Kaelbling et al., 1996):

$$Q_{\pi}(s_t, a_t) = [r(s_t, a_t) + \gamma \sum_{s_{t+1} \in \mathcal{S}} T(s_t, a_t, s_{t+1}) Q_{\pi}(s_{t+1}, a_{t+1})] \quad (5)$$

na qual $T(s_t, a_t, s_{t+1})$ = probabilidade de transição do estado s_t para o estado s_{t+1} através da ação a_t .

Uma política π é melhor que, ou igual a outra política π' se a expectativa de retorno sob π é maior ou igual à expectativa de retorno sob π' para todos os estados. Assim:

$$\pi(s_t) \geq \pi'(s_t) \text{ para qualquer } s_t \in S \quad (6)$$

Sempre há ao menos uma política melhor que ou igual a todas as outras. Esta é denominada então a política ótima e denotada por π^* . A função de avaliação sob a política ótima é definida como:

$$V^*(s_t) = V_{\pi^*}(s_t) = \max_{\pi} V_{\pi}(s_t), \text{ para qualquer } s_t \in S \quad (7)$$

A função de avaliação do par estado-ação sob a política ótima é dada por:

$$Q^*(s_t, a_t) = Q_{\pi^*}(s_t, a_t) = \max_{\pi} Q_{\pi}(s_t, a_t), \text{ para qualquer } s_t \in S \text{ e } a_t \in A(s_t). \quad (8)$$

Finalmente, a política ótima pode ser definida através da seguinte equação:

$$\pi^*(s_t) = \arg \max_a [r(s_t, a_t) + \gamma \sum_{s_{t+1} \in S} T(s_t, a_t, s_{t+1}) V_{\pi^*}(s_{t+1})] \quad (9)$$

Apesar de grande parte dos métodos de aprendizagem realizarem estimativas de funções de avaliação de estado ou de pares estado-ação, há uma outra forma de resolver problemas de aprendizagem por reforço. Existem algumas estratégias, como algoritmos genéticos e métodos de programação genética, que realizam diretamente a busca no espaço de políticas a fim de descobrir aquela que se adequa ao ambiente, sem necessariamente estimar funções de avaliação (Schmidhuber, 1996). Nos casos em que o espaço de políticas é relativamente pequeno, ou em que ele é estruturado de forma que boas políticas sejam facilmente encontradas, os métodos denominados evolucionários podem ser eficazes. Estes métodos também são adequados para problemas nos quais o agente não é capaz de sensoriar de forma precisa o estado em que o ambiente se encontra. Este tipo de ambiente é denominado ambiente parcialmente observável.

Existem três abordagens para resolução de problemas de aprendizagem por reforço (Sutton & Barto, 1998): programação dinâmica, métodos de Monte Carlo e aprendizagem baseada em diferença temporal. Cada uma destas técnicas apresenta vantagens e limitações e

se adequam a problemas específicos de aprendizagem por reforço. A estratégia de programação dinâmica requer um modelo completo e preciso do ambiente. Os métodos de Monte Carlo não necessitam de um modelo do ambiente e são conceitualmente simples, porém não são adequados a problemas em que as avaliações são realizadas passo a passo. A aprendizagem baseada em diferença temporal também não requer um modelo do ambiente e é incremental, entretanto apresenta maior dificuldade de análise. Estas classes também diferem com relação a eficiência e velocidade de convergência, entre outros aspectos.

A seguir são descritos os princípios básicos da aprendizagem baseada em diferença temporal, seguidos pelas principais características da aprendizagem-Q. Esta técnica foi selecionada devido ao fato de estar sendo utilizada de forma satisfatória no roteamento em redes de telecomunicações.

4.2 –Aprendizagem Baseada em Diferença Temporal

A aprendizagem baseada em diferença temporal (*Temporal Difference* - TD) possui características comuns aos demais métodos de aprendizagem por reforço, denominados métodos de Monte Carlo e de programação dinâmica. Com relação aos métodos de Monte Carlo, TDs também aprendem diretamente através da experiência, sem necessitar de um modelo do meio ambiente. Como ocorre nos métodos de programação dinâmica, TDs atualizam suas estimativas utilizando estimativas anteriormente aprendidas, sem esperar por um resultado ao final da sequência de ações realizadas. Logo, o valor estimado de um estado é ajustado baseado no retorno imediato obtido e no valor estimado do próximo estado.

Deste modo, considerando o método simples de diferença temporal, denominado TD(0), prever $V_{\pi}(s_t)$ corresponde a calcular o retorno esperado após visitar o estado s_t , seguindo uma política π . Assim, se o estado s_t é visitado no instante de tempo t , a atualização da estimativa $V(s_t)$ é baseada no que acontece após esta visita. Os métodos de diferença temporal esperam somente até o próximo passo de tempo para atualizar $V(s_t)$:

$$\Delta V(s_t) = \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] \quad (10)$$

Assim, no instante de tempo $t+1$, estes métodos utilizam o retorno obtido r_{t+1} e a estimativa $V(s_{t+1})$ do estado s_{t+1} . Deste modo, o valor de referência para os métodos de diferença temporal é $r_{t+1} + \gamma V(s_{t+1})$, que constitui o alvo da aprendizagem desta estratégia.

Este alvo é uma estimativa por duas razões: o valor esperado é amostrado e a estimativa atual de $V(s_t)$ é utilizada ao invés do valor verdadeiro $V_\pi(s_t)$.

Existem dois tipos de frequência para atualização das estimativas, que caracterizam os sistemas como problemas de um passo e problemas de múltiplos passos (Sutton, 1988). No primeiro caso, toda a informação sobre a qualidade da predição é obtida a cada passo de tempo, enquanto no segundo somente informação parcial a respeito da predição realizada é obtida a cada passo, sendo sua totalidade revelada após vários instantes de tempo. Assim, os métodos de um passo envolvem a utilização do valor de estado (ou par estado-ação) de um único sucessor, enquanto os de múltiplos passos se baseiam também na distribuição completa dos valores de estado de todos os possíveis sucessores. Os problemas de múltiplos passos são predominantes em aplicações ao mundo real.

Observa-se que, enquanto os métodos supervisionados de aprendizagem para predição realizam suas ações baseados no erro entre resultados obtidos e desejados, as estratégias de diferença temporal utilizam o erro entre predições sucessivas no tempo (Sutton, 1988). Deste modo, o aprendizado ocorre sempre que há uma mudança na predição ao longo do tempo.

Os métodos de aprendizagem baseados em diferença temporal se dividem em duas classes: dependentes da política (*on-policy*) e não dependentes da política (*off-policy*). Os métodos dependentes da política são aqueles em que a atualização dos valores de estado $V(s_t)$ ou do par estado-ação $Q(s_t, a_t)$ é baseada na ação selecionada, de acordo com a política π utilizada pelo algoritmo. Os algoritmos dependentes da política compreendem: algoritmos TD(λ), TD(0), Sarsa(0) e Sarsa(λ) (Sutton & Barto, 1998). Os métodos não dependentes da política, por sua vez, atualizam o valor do par estado-ação $Q(s_t, a_t)$ baseados na ação que representa a melhor escolha para o próximo estado s_{t+1} , independentemente da política π utilizada. Os algoritmos não dependentes da política compreendem: algoritmo de aprendizagem-Q e algoritmo Q(λ) de Watkins (Sutton & Barto, 1998). O algoritmo Q(λ) de Peng & Williams (Peng & Williams, 1994, 1996) realiza atualizações dos valores-Q dependentes e não-dependentes da política. O Apêndice I apresenta os métodos TD(λ), Sarsa(0), Sarsa(λ), Q(λ) de Watkins e Q(λ) de Peng & Williams, que não são descritos neste capítulo, uma vez que não são diretamente utilizados neste projeto de pesquisa.

4.2.1 –Aprendizagem-Q

Um dos mais importantes avanços em aprendizagem por reforço foi o desenvolvimento de um algoritmo de controle independente da política denominado aprendizagem-Q, realizado por Watkins (1989). A aprendizagem-Q é definida como um

método de aprendizagem por reforço que não se baseia em um modelo do ambiente e que pode ser visto como uma estratégia assíncrona de programação dinâmica (Watkins & Dayan, 1992). Este método é uma forma primitiva de aprendizado, que pode ser utilizada como base para estratégias mais elaboradas. Exemplos de sua utilização encontram-se em (Barto & Singh, 1990; Sutton, 1990; Chapman & Kaelbling, 1991; Mahadevan & Connell, 1991 e Lin, 1992), que desenvolveram este método de forma independente.

Em sua forma mais simples, a regra de atualização da aprendizagem-Q de um passo é definida como:

$$\Delta Q(s_t, a_t) = \alpha[r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (11)$$

A forma geral deste algoritmo é mostrada a seguir:

```

Inicialize  $Q(s, a)$  arbitrariamente;
Repita (para cada trajetória):
    Perceba o estado atual do ambiente  $s_t$ ;
    Repita (para cada passo da trajetória):
        Escolha uma ação  $a_t$ ;
        Realize a ação  $a_t$ , observe o retorno  $r_{t+1}$  e o próximo estado  $s_{t+1}$ ;
        Atualize o valor  $Q(s_t, a_t)$ :
             $Q(s_t, a_t) = Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$ 
        Faça  $s_t = s_{t+1}$ ;
    Até que  $s_t$  seja o estado final.
Até que a trajetória seja a última.

```

Neste caso, a_t = ação dada pela política π para o estado s_t e $t = 0, 1, 2, 3, \dots$

Ao atualizar a função de avaliação Q utilizando o máximo valor de $Q(s_{t+1}, a_{t+1})$, a função de avaliação aprendida se aproxima diretamente da função da avaliação ótima Q^* , independente da política que esteja sendo seguida (Watkins & Dayan, 1992). A política ainda possui um efeito na aprendizagem, no sentido que define quais pares estado-ação serão visitados e atualizados. Entretanto, para que ocorra convergência, é necessário que todos os pares estado-ação continuem a ser visitados e atualizados.

O algoritmo de aprendizagem-Q foi combinado com traços de elegibilidade, representados por λ , para que a atualização dos valores-Q levasse em consideração os valores dos pares estado-ação de todos os possíveis sucessores ou de um conjunto deles, caracterizando-se como um problema de múltiplos passos. Esta integração procurou aumentar a eficiência destes algoritmos, para que eles se adaptem melhor à problemas do mundo real. Os algoritmos são descritos no Apêndice I.

4.3 – Aprendizagem por Reforço Dual

Na aprendizagem por reforço dual, o agente utiliza o sinal de reforço recebido do ambiente em cada passo de sua trajetória para aprender uma política tanto na direção do seu movimento como na direção oposta. Deste modo, o espaço de busca do ambiente é explorado de forma mais eficiente que na aprendizagem-Q, permitindo que o sistema aumente a velocidade e qualidade de adaptação à diversas situações. Esta estratégia foi inicialmente aplicada ao problema de comunicação de satélites (Goetz et al., 1996), possibilitando que os dois pontos terminais do sistema se adaptem em conjunto às condições do ambiente ao compartilharem o sinal de reforço recebido por cada um deles.

Uma questão importante reside no fato de que a aprendizagem por reforço dual deve ser aplicada em problemas de custo simétrico, isto é, quando o custo de se movimentar do estado s_t para o estado s_{t+1} é equivalente ao movimento no sentido inverso.

4.4 – Conclusão

Neste capítulo foram apresentados os principais aspectos da aprendizagem por reforço, dando ênfase aos métodos baseados em diferença temporal. Estes métodos foram selecionados para este projeto devido à capacidade de aprendizado por eles apresentada, sem necessidade de um modelo completo do comportamento do ambiente. Assim, estas estratégias se ajustam ao problema de roteamento, onde não é possível conhecer previamente todas as condições em que o ambiente se encontrará. Dentre elas, destacou-se a aprendizagem-Q, que apresenta a capacidade de convergência para a política ótima, independente da política seguida, desde que todos os pares estado-ação do ambiente sejam continuamente visitados. Assim também, foi apresentada a aprendizagem por reforço dual, que realiza a exploração mais eficaz do espaço de busca de soluções e permite que o sistema aumente sua velocidade e qualidade de adaptação. Desta forma, estas duas estratégias são utilizadas no algoritmo de roteamento proposto, o que contribuiu para o desempenho satisfatório por ele apresentado.

O algoritmo Agentes-Q também utiliza propriedades pertinentes a outra abordagem de aprendizagem, a qual é baseada no comportamento de colônias de formigas e que tem sido recentemente aplicada com sucesso ao problema de roteamento em redes de telecomunicações. Deste modo, no próximo capítulo serão apresentados os aspectos fundamentais deste método.

5 – Otimização Baseada no Comportamento Coletivo de Formigas

Neste capítulo é discutido inicialmente o comportamento de colônias de formigas reais, que apresentam elevado nível de estruturação e conseqüente capacidade de realização de tarefas complexas. A observação destas características conduziu alguns pesquisadores a desenvolverem um método de otimização baseado em agentes múltiplos (*Ant Colony Optimization* - ACO) para resolução de problemas de otimização combinatorial complexos (Colormi et al., 1991; Dorigo, 1992; Dorigo et al, 1991, 1996, 1999). Em seguida, são apresentados os principais aspectos desta técnica de otimização, assinalando suas semelhanças e diferenças em relação as colônias de formigas reais.

5.1 – Formigas na Natureza

As formigas são denominadas insetos sociais, devido ao fato de viverem em colônias e se comportarem de modo a garantir a sobrevivência de toda a colônia em detrimento de um único indivíduo pertencente a esta. O alto nível de estruturação apresentado por colônias de insetos sociais, em particular de formigas, tem levado a comunidade científica a estudar com bastante interesse a vida destes insetos e os fatores que influenciam seu comportamento (Deneubourg & Goss, 1989; Franks, 1989; Deneubourg et al., 1990; Goss et al., 1989; Beckers et al., 1992; Stickland et al., 1992 e Theraulaz & Bonabeau, 1995). As formigas apresentam individualmente memória limitada e comportamento simples, com alto componente aleatório, mas em conjunto são capazes de realizar tarefas complexas, como construção e proteção do ninho, regulagem da temperatura do mesmo, emigração de uma colônia e exploração da melhor fonte de alimento disponível (Franks, 1989; Holldobler & Wilson, 1994 e Schoonderwoerd et al., 1996).

A pesquisa científica tem demonstrado que a auto-organização constitui um dos principais componentes do comportamento coletivo de insetos sociais (Deneubourg & Goss, 1989), pois muitas atividades realizadas coletivamente por eles resultam em estruturas complexas. Esta é uma característica do processo de auto-organização, que foi originalmente introduzido nas áreas de física e química para descrever como processos microscópicos

originam estruturas macroscópicas em sistemas fora do equilíbrio. As teorias de auto-organização podem ser extendidas a área de etologia, em especial a insetos sociais, para demonstrar que a interação entre indivíduos simples pode produzir um comportamento altamente estruturado (Haken, 1977; Nicolis & Prigogine, 1977).

Um aspecto importante das colônias de formigas está relacionado com o seu movimento em busca de alimento. As formigas são capazes de descobrir os menores caminhos entre o ninho e a fonte de alimento, realizando esta tarefa de forma distribuída e baseadas somente em informação local. Ao caminharem em direção à fonte de alimento e retornarem ao ninho, as formigas depositam no solo uma substância química, denominada feromônio, formando uma trilha (Beckers et al., 1992). Esta substância pode ser sentida através do olfato pelas demais formigas da colônia, que tomam suas decisões sobre a rota a ser percorrida de forma probabilística, baseadas na concentração de feromônio sobre os possíveis caminhos. Esta concentração é função do número de formigas que percorreram aquele caminho e da quantidade de feromônio que cada uma delas depositou. Da mesma forma, ao retornarem da fonte de alimento ao ninho, as formigas encontram o caminho através da presença desta substância.

A fim de verificar o comportamento de formigas em busca de alimento sob condições controladas, Deneubourg et al. (1990) realizaram o experimento da ponte binária (Figura 4a). Assim, o ninho e a fonte de alimento destes insetos foram separados por uma ponte que possui dois ramos de mesmo comprimento. Inicialmente, não havia feromônio em nenhum dos ramos e as formigas se movimentavam aleatoriamente sobre ambos (Figura 4b). Após um determinado período de tempo, um dos ramos se tornou aleatoriamente mais utilizado, sendo neste caso o superior. Deste modo, houve uma maior concentração de feromônio sobre este ramo, o que conduziu um número maior de formigas a percorrerem este caminho, que a partir deste momento possuiu maior probabilidade de ser selecionado.

Figura 4 – (a) Ponte sobre a qual o experimento foi realizado por Deneubourg et al. (1990), (Dorigo et al., 1999). (b) Resultados obtidos em termos da porcentagem de formigas que utilizaram cada um dos ramos da ponte ao longo do tempo (Dorigo et al., 1999).

Goss et al. (1989) realizaram um experimento em que os dois ramos de uma ponte dupla que interliga o ninho das formigas à fonte de alimento possuem comprimentos diferentes (Figura 5a). Neste caso também não havia inicialmente feromônio depositado nos possíveis caminhos no início do experimento. Assim, inicialmente a mesma proporção de formigas percorreu os dois ramos da ponte. Após um determinado período, as formigas que selecionaram o menor caminho alcançaram seu destino em um menor intervalo de tempo, enquanto outras ainda se encontravam percorrendo o caminho mais longo (Figura 5b). Deste modo, este caminho possuiu uma maior concentração de feromônio do que o outro, o que estimulou as próximas formigas a selecionarem o menor caminho.

Figura 5 – (a) As formigas iniciam a exploração da ponte binária (Dorigo et al., 1999). (b) Após um período de tempo, a maioria das formigas seleciona o menor caminho (Dorigo et al., 1999).

Porém, se as formigas são apresentadas inicialmente a um caminho longo e somente após um intervalo de tempo ocorre a apresentação a um caminho mais curto, elas podem não necessariamente selecionar a melhor opção. Em algumas espécies de formigas, como a *Lasius Niger*, é possível descobrir o menor caminho nesta situação através da memória de cada formiga sobre a direção em que se encontra o ninho ou a fonte de alimento, combinada com a trilha de feromônios construída pelas formigas que aleatoriamente selecionarem o menor caminho (Bonabeau et al., 1997).

Portanto, Grassé (1946) observou que os insetos são capazes de responder a um estímulo, neste caso a presença de feromônio, que ativa uma reação codificada geneticamente. No caso dos insetos sociais, os efeitos destas reações podem agir como um novo estímulo tanto para o inseto que as produziu como para os demais indivíduos da colônia. A produção de um novo estímulo como reação a um já existente determina uma forma de coordenação de atividades e pode ser interpretada como um modo de comunicação indireta. O mecanismo de comunicação indireta entre as formigas que permite a emergência deste processo é denominado *stigmergy* (Grassé, 1946; Theraulaz & Bonabeau, 1999). Assim, apesar de cada formiga ser capaz de encontrar o menor caminho entre o ninho e a fonte de alimento, somente a interação de toda a colônia torna possível a obtenção dos melhores caminhos. *Stigmergy* se diferencia de outras formas de comunicação através das seguintes propriedades: (1) a natureza física da informação liberada pelos insetos, que corresponde a mudanças nos estados físicos do ambiente visitados pelos mesmos e (2) a natureza local da informação, que somente pode ser acessada pelos insetos que visitam o estado no qual elas foram liberadas ou alguma vizinhança deste.

A mudança de estados do ambiente através do processo de *stigmergy* pode ser realizada de duas formas diferentes (Schoonderwoerd et al., 1996). Na primeira delas, as características físicas do ambiente são modificadas através de ações realizadas pelas formigas relacionadas diretamente com a tarefa em questão, como cavar um buraco no solo. Neste caso, a mudança do estado do ambiente pode estimular as demais formigas da colônia a aumentarem o buraco cavado inicialmente por alguns indivíduos. Assim, os efeitos cumulativos das modificações locais ocorridas no ambiente podem conduzir ao crescimento de uma estrutura complexa. Este tipo de influência exercida pelas formigas é denominado *sematectonic* (Wilson, 1975). A segunda forma de *stigmergy* ocorre através do depósito de alguma substância no solo que não se encontra diretamente relacionada com a tarefa em questão realizada, mas que influencia o comportamento dos demais indivíduos que percorrem os estados do ambiente modificados. A substância que constitui o sistema de

sinalização entre as formigas é o feromônio, como descrito acima. Este tipo de processo é denominado *stigmergy* baseado em sinal.

Uma questão importante relacionada com o mecanismo de *stigmergy* diz respeito ao entendimento de como os estímulos aos quais os insetos respondem, no caso a presença de feromônio, são organizados no tempo e no espaço para assegurar a realização correta da tarefa em questão. Deneubourg (1977) verificou que a concentração de feromônios no ambiente pode diferir tanto quantitativa quanto qualitativamente.

Verifica-se que a descoberta dos menores caminhos pelas formigas emerge através da interação e cooperação entre todos os indivíduos da colônia, caracterizando-se, ao menos em parte, como um processo de auto-organização. As interações entre os indivíduos são baseadas somente em informação local, sem referir-se a uma visão global do ambiente, a qual surge como uma propriedade emergente do sistema. No caso das formigas, as estruturas emergentes do sistema são as trilhas de feromônios, que constituem padrões organizados no tempo e no espaço.

As características do processo de auto-organização que podem ser evidenciadas no comportamento de colônias de formigas são (Bonabeau et al., 1997): (1) a criação de estruturas espaço-temporais em um meio inicialmente homogêneo; (2) a possível convergência para vários estados estáveis, denominada multi-estabilidade: os movimentos aleatórios dos componentes do sistema podem conduzir à convergência para diversos estados, de acordo com as condições iniciais; (3) a existência de bifurcações, quando alguns parâmetros do sistema são modificados. Como exemplo desta última característica, podemos citar o comportamento de colônias de cupins: a construção de colunas para sustentarem os ninhos requer a presença de um determinado número de indivíduos. Assim, o sistema apresenta uma bifurcação neste ponto determinado: se existir um número de cupins inferior ao requerido, não é possível a emergência de colunas, mas acima deste valor esta tarefa se torna viável.

Os componentes do processo de auto-organização que podem ser evidenciados no comportamento das colônias de formigas são (Bonabeau et al., 1997): (1) retorno positivo, que se caracteriza pela maior concentração de feromônios sobre os melhores caminhos; (2) retorno negativo, que está relacionado com o número limitado de formigas da colônia, com a existência de diversas fontes de alimento e com o esgotamento da fonte considerada; (3) aleatoriedade, que permite que os indivíduos descubram novas soluções e (4) múltiplas interações, que se caracterizam pela interação entre todos os integrantes do sistema e pela sua capacidade de utilizar os resultados de suas próprias atividades, como também dos demais.

A emergência de estruturas complexas como consequência de um comportamento coletivo pode ter sido favorecida pela evolução das espécies, que se baseia na teoria de seleção natural. A seleção dos indivíduos pode agir sobre fatores ou parâmetros que influenciam a estrutura de suas colônias, sejam estas auto-organizadas ou não. Estes fatores incluem o limiar de resposta (*threshold*) aos estímulos gerados pelos indivíduos, as reações desencadeadas pela resposta, como também as propriedades específicas do estímulo gerado e/ou recebido. Mudanças nestes parâmetros podem afetar a estrutura do sistema e as condições sob as quais elas emergem e são mantidas. Como exemplo, citamos a volatilidade do feromônio depositado pelas formigas ao longo do caminho percorrido: um feromônio altamente volátil requer uma grande quantidade de indivíduos para manter uma trilha estável desta substância, o que influencia a eficiência da colônia na realização de uma determinada tarefa.

Sistemas auto-organizados não são necessariamente adaptativos ou cooperativos. Porém, no caso dos insetos sociais, ao atuar sobre os aspectos comportamentais dos indivíduos, permitiu que eles apresentassem tais características. As estratégias de auto-organização podem ter surgido nos insetos sociais inicialmente devido ao seu comportamento individual simples e à facilidade das condições necessárias à emergência de um comportamento coletivo complexo. A seleção dos mecanismos individuais destes insetos foi função da eficiência de suas atitudes diante do ambiente em que eles se encontravam.

Os mecanismos biológicos adquiridos pelos insetos sociais com a evolução das espécies também podem ter contribuído para a utilização de estratégias de auto-organização por estes indivíduos. As formigas podem realizar tarefas diferentes a partir dos mesmos sinais em algumas espécies, como também através de sinais diversos. Isto decorre do fato de que o mecanismo de reconhecimento do sinal e posterior realização da tarefa se mantém em todas as situações.

Uma outra razão para a relação entre a seleção natural de insetos sociais e a existência de processos auto-organizados entre os mesmos, encontra-se no fato de que indivíduos de espécies diferentes que possuam comportamento semelhante, podem provocar respostas coletivas diferentes em ambientes diversos. Esta situação sugere que as regras de comportamento podem ser qualitativamente similares em todas as espécies, porém os aspectos evolutivos proporcionaram variação quantitativa, devido possivelmente a diferentes limiares de resposta.

Apesar de todas as evidências sobre a ligação entre a evolução dos insetos sociais e seu comportamento auto-organizado, a questão sobre como estes aspectos interagem ainda é amplamente discutida na comunidade científica.

5.2 –Otimização Baseda no Comportamento de Colônias de Formigas

O método de otimização baseado no comportamento de colônias de formigas, (*Ant Colony Optimization – ACO*), (Dorigo, 1992; Dorigo et al., 1991, 1996, 1999) é constituído por um conjunto de agentes simples, que realizam suas tarefas de forma cooperativa, com o objetivo de encontrar boas soluções para problemas de otimização discretos complexos. As características apresentadas pelas formigas reais podem ser facilmente extendidas a agentes artificiais através de: (1) definição de variáveis de estado apropriadas aos estados do problema e (2) acesso local aos valores destas variáveis pelos agentes artificiais. Os agentes artificiais irão simular a construção da trilha de feromônios realizada pelas formigas através da modificação das variáveis associadas aos estados do problema visitados durante a busca de soluções para o mesmo. As formigas artificiais apresentam algumas características semelhantes àquelas das formigas reais, adicionadas a algumas capacidades necessárias a resolução de problemas de otimização combinatorial complexos.

Os aspectos fundamentais deste método de otimização originários do comportamento de insetos sociais em busca dos melhores caminhos são os seguintes: (1) existência de uma colônia de agentes que cooperam mutuamente, (2) modificação das variáveis de estado do problema como forma de comunicação indireta, que simula a construção de trilhas de feromônios, (3) realização de uma sequência de movimentos locais entre os estados do ambiente e (4) utilização de uma política estocástica de decisão sobre o caminho a ser percorrido utilizando apenas informação local.

Assim como as colônias de formigas existentes na natureza, os algoritmos baseados no comportamento destes insetos são compostos por uma população de agentes móveis que realizam tarefas de forma assíncrona e simultânea, cooperando entre si para obter boas soluções. A cooperação entre os agentes ocorre através da atualização de uma variável numérica associada a cada estado do ambiente e do acesso a este novo valor pelos demais agentes que percorrem estes estados. A atualização desta variável está relacionada com a qualidade dos movimentos realizados pelas formigas em relação ao objetivo do problema.

Em geral, os algoritmos baseados no comportamento de formigas apresentam um mecanismo de evaporação do feromônio, que procura simular o fenômeno que ocorre na natureza. Este mecanismo permite que os agentes esqueçam lentamente seus movimentos passados, de forma que novas soluções possam ser obtidas, o que permite um comportamento dinâmico diante das condições do ambiente.

As formigas existentes na natureza e as artificiais procuram encontrar o menor caminho, ou seja, aquele que apresenta o menor custo, entre um determinado ponto de origem e um ponto de destino. As formigas reais se movimentam através de estados próximos do terreno e assim também, as formigas artificiais percorrem passo a passo estados adjacentes de um determinado problema. Os estados e a adjacência são determinados de acordo com o problema em questão.

Ambos os insetos sociais e os agentes artificiais constroem suas soluções para o problema utilizando uma política de transição de estados probabilística, baseada apenas na informação local sobre o ambiente. A política é função da informação prévia representada pelas características do ambiente, como a estrutura do terreno em que as formigas reais se encontram, juntamente com as modificações locais do mesmo realizada pelas formigas que o percorreram.

As formigas artificiais diferem em alguns aspectos das reais, sendo estes descritos a seguir.

- ♦ As formigas artificiais vivem em um mundo discreto e seus movimentos consistem em transições através dos estados do mesmo.
- ♦ Os agentes possuem um estado interno, que constitui sua memória, onde são armazenadas informações sobre as ações realizadas.
- ♦ As formigas artificiais simulam o depósito de feromônio no ambiente através da atualização das variáveis de estado, que é proporcional a qualidade da solução encontrada.
- ♦ O período de tempo em que as formigas modificam as informações armazenadas nos estados do ambiente é dependente do problema. A atualização destes valores pode ser realizada após a geração de uma solução completa, em seguida a cada visita do agente a um estado, ou através da combinação das duas.
- ♦ A fim de melhorar o desempenho do sistema, os algoritmos de otimização baseados no comportamento coletivo de formigas podem apresentar algumas capacidades extras, como a previsão da qualidade de estados futuros (Michel & Middendorf, 1998), otimização local (Dorigo & Gambardella, 1997; Stützle & Hoos, 1997) e movimentação baseada em estados passados, *backtracking* (Di Caro & Dorigo, 1997, 1998).

A estratégia de aprendizagem baseada no comportamento coletivo de formigas foi aplicada inicialmente ao problema de otimização combinatorial do caixeiro viajante, em que um indivíduo deve percorrer o menor caminho entre duas cidades visitando somente uma vez

cada cidade da região em questão (Dorigo et al., 1991). Recentemente, esta abordagem tem sido utilizada para solucionar diversos tipos de problemas de otimização, como escalonamento de operações realizadas por uma série de máquinas da área industrial (Colorni et al., 1994), roteamento de um conjunto de veículos que devem partir e retornar ao mesmo ponto, levando encomendas a uma série de clientes (Bullnheimer et al., 1998) e ordenamento sequencial (Gambardella & Dorigo, 1997), que consiste em encontrar o menor caminho entre duas cidades, onde estas não se encontram diretamente conectadas.

5.3 – Conclusão

Neste capítulo foram descritos inicialmente os principais aspectos de colônias de formigas, que são constituídas por indivíduos simples que interagem e cooperam entre si e se comunicam indiretamente. Estas características permitem que estes insetos sociais realizem coletivamente tarefas altamente complexas de forma eficiente e eficaz

Em seguida, foi descrita uma abordagem de otimização baseada neste comportamento e desenvolvida para solucionar problemas complexos. Esta estratégia vem sendo aplicada a diversos problemas de otimização e os resultados obtidos demonstram sua adequação aos mesmos. Dentre os sistemas de otimização desenvolvidos baseados neste método, encontram-se os que realizam o roteamento em uma rede de telecomunicações. Estes sistemas têm apresentado a capacidade de se adaptar às variações nas condições de operação da rede, obtendo desempenho superior a sistemas de roteamento tanto clássicos como baseados em agentes móveis, usualmente utilizados pelas empresas de telecomunicações.

Dentro deste contexto, o sistema de roteamento desenvolvido neste projeto de pesquisa, que será apresentado no Capítulo 7, possui algumas das principais características descritas neste capítulo. O conhecimento destas características foi de suma importância na implementação do algoritmo proposto, pois possibilitou a compreensão das propriedades fundamentais desta abordagem, que a tornaram amplamente utilizado nesta década.

Serão apresentados e analisados no próximo capítulo alguns algoritmos de roteamento aplicados a redes de comutação de circuitos e a redes de comutação de pacotes baseados em ACO. Assim também, serão descritos e discutidos alguns algoritmos desenvolvidos recentemente baseados em técnicas de aprendizagem. Os algoritmos utilizados nas redes de alta velocidade e redes móveis não serão apresentados, uma vez que o sistema de roteamento desenvolvido neste projeto de pesquisa foi aplicado somente a redes de comutação de circuitos.

6- Algoritmos de Roteamento Baseados em Aprendizagem

Neste capítulo são apresentados os principais algoritmos de roteamento baseados em técnicas computacionais inteligentes desenvolvidos recentemente e aplicados a redes de comutação de circuitos e de pacotes. Os algoritmos recebem esta denominação devido a sua capacidade de aprender a resolver o problema do roteamento sem um conhecimento prévio sobre as condições de operação da rede e por se adaptarem continuamente às situações enfrentadas pela mesma, baseados somente na experiência e sem a necessidade de intervenção humana.

Dentro deste contexto, na primeira parte deste capítulo são descritos e analisados três algoritmos de roteamento baseados em aprendizagem-Q, denominados Roteamento-Q (Littman & Boyan, 1993; Boyan & Littman, 1994), Roteamento-Q Preditivo (Choi & Yeung, 1996) e Roteamento-Q com Reforço Dual (Kumar & Miikkulainen, 1997), que foram aplicados a redes de comutação de pacotes. O primeiro deles constitui a aplicação pioneira desta técnica de aprendizagem ao problema do roteamento em redes de telecomunicações e os outros dois constituem evoluções do primeiro. O desempenho destes algoritmos foi comparado ao de um sistema clássico estático sob diversas situações. Em todas as situações consideradas, os algoritmos baseados na aprendizagem-Q obtiveram melhores resultados que o sistema estático.

Na segunda parte deste capítulo, são descritos dois algoritmos de roteamento baseados no comportamento coletivo de formigas, denominados Controle Baseado em Formigas (Schoonderwoerd et al., 1996, 1997) e Controle Baseado em Formigas Espertas (Bonabeau et al., 1998), os quais foram aplicados a redes de comutação de circuitos. Os resultados produzidos por estes algoritmos foram comparados ao de um algoritmo de roteamento clássico e ao de algoritmos recentemente desenvolvidos baseados em agentes móveis. Os algoritmos baseados em colônias de formigas apresentaram melhor desempenho que os demais sob uma variedade de situações consideradas.

6.1 – Algoritmos de Roteamento Baseados na Aprendizagem-Q

Nos últimos anos, tem havido um grande interesse da comunidade científica na aplicação da aprendizagem por reforço às redes de telecomunicações e em especial ao problema de roteamento, que constitui um problema de otimização combinatória dinâmico e complexo.

Isto se deve ao fato da aprendizagem por reforço apresentar características que se ajustam ao problema de roteamento, que possui aspectos dinâmicos e estocásticos. Estas características compreendem a capacidade de aprender a atuar em um meio sem a necessidade de possuir um modelo prévio sobre o mesmo e a capacidade de adaptação a situações imprevistas nas condições de operação do meio.

Este método de aprendizagem vem apresentando desempenho superior às técnicas convencionais utilizadas, pois possibilita que o agente se adapte mais rápida e precisamente a novas situações ocorridas, característica fundamental de um roteador.

A seguir, são descritos três dos principais algoritmos que aplicam a aprendizagem-Q a redes de comutação de pacotes.

6.1.1 – Roteamento-Q (*Q-Routing*)

O Roteamento-Q (Littman & Boyan, 1993; Boyan & Littman, 1994) constitui um algoritmo dinâmico e distribuído, que se adapta a mudanças na rede em tempo real e onde cada nó seleciona as rotas de forma independente. A métrica para escolha das rotas está relacionada com o atraso decorrido na entrega do pacote. As rotas são estabelecidas a partir de informação local. O Roteamento-Q foi aplicado em uma rede de 36 nós irregularmente conectada, apresentada na Figura 6.

Figura 6 - Rede irregular 6x6.

O algoritmo do Roteamento-Q é descrito a seguir:

1 - Cada nó i da rede armazena em sua tabela de roteamento valores $Q_i(j, d)$, que representam a estimativa do tempo necessário para que um pacote P seja transmitido a partir de i através de seu nó vizinho j para o destino d , incluindo o atraso sofrido na fila de i . As tabelas são inicializadas de forma aleatória e apresentam baixos valores- Q . Quando o nó i necessita tomar uma decisão de roteamento, ele escolhe o nó vizinho j para o qual o valor $Q_i(j, d)$ é mínimo:

$$j = \arg \min_j Q_i(j, d) \quad (12)$$

Desta forma, verifica-se que o algoritmo procura minimizar o tempo total dispendido na transmissão de um pacote ao longo do caminho estabelecido.

2 - Ao enviar o pacote P para j , i imediatamente recebe a estimativa vinda de j sobre o tempo t remanescente na rota para alcançar d :

$$t = \arg \min_k Q_j(k, d) \quad (13)$$

na qual $k \in$ conjunto de vizinhos de j .

3 - Representando o tempo decorrido na fila do nó i por q e o tempo decorrido na transmissão através da conexão entre os nós i e j por s , i atualiza sua estimativa da seguinte forma:

$$\Delta Q_i(j, d) = \eta[(q + s + t) - Q_i(j, d)] \quad (14)$$

na qual η = taxa de aprendizagem; os autores utilizaram 0.5.

Assim, à medida que os pacotes são transmitidos, seus valores- Q gradualmente incorporam informação global sobre a rede.

Os testes realizados com o Roteamento-Q consistiram em variações nos níveis de carga da rede, que constitui o número de pacotes introduzidos na rede a cada passo de tempo

da simulação. Os experimentos compreenderam dois níveis de carga da rede: baixo e alto. Os pacotes foram introduzidos periodicamente em nós aleatoriamente escolhidos e seus destinos também foram selecionados de forma aleatória. Os nós foram considerados com capacidade de armazenamento de pacotes infinita.

O desempenho do algoritmo foi medido através do tempo médio de entrega dos pacotes após o aprendizado de uma política de roteamento e comparado com os tempos de entrega de um algoritmo de rotas fixas baseado nos menores caminhos, que atua *offline*. Verificou-se que sob condições de nível de carga baixo, após um período de aprendizado, o Roteamento-Q realiza a transmissão dos pacotes em tempo médio similar ao roteamento de menor caminho, sendo capaz de selecionar rotas ótimas. Entretanto, no período em que a carga da rede é acrescida, o roteamento de menor caminho não se adapta à nova condição de operação e congestionamento o tráfego de pacotes na região central da rede. O Roteamento-Q aprende uma política eficiente e distribui o acréscimo de carga pela região superior da mesma, aumentando o tempo decorrido na transmissão dos pacotes, porém evitando o congestionamento do tráfego na região central da rede.

Os experimentos realizados demonstraram que ambos os algoritmos apresentam uma queda significativa em seu desempenho sob condições de variação de carga: o roteamento de menor caminho não se adapta à esta variação e logo produz o congestionamento da rede, enquanto o Roteamento-Q permanece atuando de forma eficiente no ambiente até que a carga atinja um determinado nível, a partir do qual ele não é mais capaz de se adaptar às variações deste parâmetro.

Um outro aspecto importante deste algoritmo é verificado quando o nível de carga decresce, situação em que o algoritmo se adapta lentamente e não é capaz de atingir os melhores caminhos. Isto se deve ao fato da política de atualização por ele utilizada ser aplicada somente aos melhores vizinhos. Desta forma, se um nó possuir uma estimativa do tempo de transmissão de pacotes superior ao valor real, uma rota sub-ótima será selecionada enquanto o tempo de transmissão do pacote sobre esta rota for menor que aquele estimado de forma incorreta para a rota ótima. Assim, a velocidade do aprendizado é comprometida. Alguns algoritmos que foram propostos para solucionar este problema são apresentados a seguir.

É importante ressaltar que, apesar das limitações descritas, o Roteamento-Q ainda apresenta melhor desempenho que aquele baseado em rotas fixas sob condições de variação do nível de carga na rede.

6.1.2 – Roteamento-Q Preditivo

O Roteamento-Q Preditivo (Choi & Yeung, 1996) é um algoritmo de aprendizagem-Q adaptativo baseado em memória. Neste sistema, as melhores experiências aprendidas são armazenadas e utilizadas para prever a tendência dos padrões de tráfego da rede. Ele foi desenvolvido com o objetivo de lidar com dois problemas enfrentados pelo algoritmo de Roteamento-Q: a incapacidade de descobrir atalhos sob condições de nível baixo de carga e a inabilidade para aprender novas políticas de roteamento ótimas sob condições de decréscimo do nível de carga.

A fim de lidar com o compromisso entre a utilização dos melhores caminhos (*exploitation*) e a exploração de caminhos alternativos (*exploration*) inerente aos algoritmos de aprendizagem por reforço, o Roteamento-Q Preditivo propõe a realização de ambas as fases de forma conjunta.

Assim, sob condições de nível de carga baixo, procura-se aprender a política de roteamento ótima, em que os menores caminhos são selecionados. À medida que o nível de carga sofre um acréscimo, a política dos menores caminhos não se mantém viável, pois estas rotas tornam-se congestionadas. Se estes caminhos não forem utilizados durante um determinado período de tempo, eles serão restabelecidos e tornar-se-ão novamente aptos a enviar os pacotes com um pequeno valor de atraso. Após este período, eles poderão ser utilizados ocasionalmente, de forma que o espaço de busca seja explorado e ao mesmo tempo estes caminhos não se tornem novamente congestionados. A frequência de utilização deste mecanismo de exploração constitui um aspecto fundamental deste algoritmo e portanto deve ser dependente do nível de congestionamento da rede e da velocidade de processamento do caminho em questão. O primeiro pode ser obtido através dos valores-Q das tabelas de roteamento e a última constitui uma medida para reutilização da rota, denominada taxa de restabelecimento e deve ser estimada como parte do processo de aprendizado.

A taxa de restabelecimento está vinculada ao volume de tráfego corrente da rede e a possibilidade de falha ou falta em nós ou conexões. Como o tempo decorrido na transmissão do pacote não reflete realmente o tempo de processamento atual de um caminho, o valor obtido da taxa deve ser ajustado para cada pacote enviado. É importante ressaltar que este valor deve ser negativo, pois do contrário o valor-Q previsto seria acrescido e o caminho por ele representado não seria novamente utilizado.

A política de escolha do próximo nó a ser visitado e de atualização dos valores-Q ao transmitir um pacote do nó i com destino ao nó d utiliza as seguintes variáveis:

$Q_i(j,d)$ = estimativa do tempo de transmissão de um pacote do nó i ao destino d passando pelo nó vizinho j .

$B_i(j,d)$ = melhor estimativa do tempo de transmissão do nó i ao d utilizando o vizinho j .

$R_i(j,d)$ = taxa de restabelecimento para o caminho do nó i ao d utilizando o vizinho j .

$U_i(j,d)$ = último passo de tempo em que $Q_i(j,d)$ foi atualizada.

Assim, o algoritmo do Roteamento-Q Preditivo é descrito a seguir:

1 – Ao enviar um pacote P do nó i ao nó destino d , a escolha do vizinho j a ser visitado é realizada a partir dos seguintes cálculos:

$$\Delta t = \text{passo de tempo atual} - U_i(j,d) \quad (15)$$

$$Q'_i(j,d) = \max(Q_i(j,d) + \Delta t R_i(j,d), B_i(j,d)) \quad (16)$$

$$j = \arg \min_j Q'_i(j,d) \quad (17)$$

Deste modo, verifica-se que o algoritmo seleciona o próximo nó j a ser visitado baseado na taxa de restabelecimento de todos os nós vizinhos de i , a fim de que sejam obtidas estimativas atualizadas dos valores-Q correspondentes a cada vizinho. Isto ocorre de forma a levar em consideração a característica dinâmica do ambiente, aproximando o algoritmo da realidade das redes de telecomunicações. Após os cálculos destes novos valores, seleciona-se o nó que apresenta o menor valor-Q.

2 – Ao receber o pacote P , o nó selecionado j envia ao nó i a menor estimativa do valor-Q presente em sua tabela com relação ao destino d , que corresponde ao tempo t remanescente na rota para alcançar d , representada pela eq.(13).

3 - Representando o tempo decorrido na fila do nó i por q e o tempo decorrido na transmissão através da conexão entre os nós i e j por s , i atualiza sua estimativa do valor-Q com relação ao destino d passando por j da mesma forma que o Roteamento-Q.

$$\Delta Q_i(j,d) = \eta[(q + s + t) - Q_i(j,d)] \quad (18)$$

na qual α é a taxa de aprendizagem.

4 – Os valores das demais variáveis utilizadas na seleção dos nós a serem visitados pelos pacotes são atualizados da seguinte forma:

$$B_i(j, d) = \min(B_i(j, d), Q_i(j, d)) \quad (19)$$

$$\text{Se } (\Delta Q < 0)$$

$$\Delta R = \frac{\Delta Q}{(\text{passo de tempo atual} - U_i(j, d))} \quad (20)$$

$$R_i(j, d) = R_i(j, d) + \beta \Delta R \quad (21)$$

Senão

$$R_i(j, d) = \gamma R_i(j, d) \quad (22)$$

Fim se

$$U_i(j, d) = \text{passo de tempo atual} \quad (23)$$

nas quais: α = taxa de aprendizado, igual a 1.

β = parâmetro utilizado para aprender a taxa de restabelecimento, igual a 0.7;

γ = parâmetro utilizado para controlar o decaimento da taxa de restabelecimento, igual a 0.9.

O Roteamento-Q Preditivo também foi aplicado a rede de 36 nós apresentada na Figura 6. Seu desempenho foi comparado ao do Roteamento-Q, sob variações dos padrões de tráfego e nível de carga da rede, com relação ao tempo médio decorrido na transmissão dos pacotes através da rede.

Os experimentos realizados demonstraram que o algoritmo Roteamento-Q Preditivo produz melhores resultados que o Roteamento-Q com relação à velocidade de aprendizado e à adaptação sob condições de nível de carga baixo e variações da carga. Os algoritmos obtiveram comportamento similar sob condições de nível de carga alto, porém o Roteamento-Q Preditivo produziu menores resultados que seu competidor. Isto se deve ao fato de que o Roteamento-Q utiliza em geral políticas estáveis e envia os pacotes sobre caminhos com maior capacidade de processamento, independente do tempo de transmissão

atual. Enquanto isto, o Roteamento-Q Preditivo procura minimizar o tempo de transmissão utilizando ocasionalmente os menores caminhos. A abordagem do Roteamento-Q produz bons resultados para níveis de carga extremamente elevados, porém isto não se mantém para os demais níveis. Mesmo no caso de carga muito alta, o Roteamento-Q apresenta resultados inferiores ao algoritmo preditivo.

Quando o nível de carga permanece extremamente alto e os padrões de tráfego não variam, o Roteamento-Q Preditivo gradualmente se comporta como o Roteamento-Q, até que estes padrões variem novamente. A grande vantagem da utilização deste algoritmo encontra-se no menor tempo necessário para adaptação às novas condições de operação da rede por ele apresentado. Isto ocorre devido ao fato de que a política por ele utilizada leva em consideração a taxa de restabelecimento dos caminhos, além dos valores-Q.

Com relação aos requisitos computacionais necessários à execução dos algoritmos, o Roteamento-Q Preditivo requer maior capacidade de memória para estimar a taxa de restabelecimento dos caminhos, porém somente com relação aos nós visitados. No pior caso, o nível de memória extra requerido se assemelha ao do Roteamento-Q. Com relação ao custo computacional, o cálculo para estimar a taxa de restabelecimento é bastante simples, tornando a implementação do algoritmo de Roteamento-Q Preditivo viável.

6.1.3– Roteamento-Q com Reforço Dual

O Roteamento-Q com Reforço Dual (Kumar & Miikkulainen, 1997) é um algoritmo adaptativo e distribuído baseado na estratégia de aprendizagem por reforço dual, inicialmente desenvolvida para sistemas de comunicação via satélite (Goetz et al., 1996). Dentro deste contexto, o agente aprende políticas de roteamento ao realizar “exploração dupla” do espaço de busca, através da atualização de dois valores-Q a cada movimento do pacote ao longo da rede. A primeira delas se assemelha à realizada no Roteamento-Q, que se baseia na informação relativa à menor estimativa do roteador vizinho sobre o tempo remanescente para alcançar o destino, denominada exploração *forward*. A segunda se baseia no tempo decorrido desde a geração do pacote na rede, que é armazenado na memória do mesmo, denominada exploração *backward*.

Desta forma, o algoritmo do Roteamento-Q com Reforço Dual é composto das seguintes etapas:

- 1 - Seja o nó origem o de um pacote P que encontra-se atualmente no nó i , com destino ao nó d . O nó i seleciona o nó vizinho j para o qual o pacote será enviado a

partir do menor valor-Q presente em sua tabela referente ao destino d , representado pela eq.(12).

2 - Ao receber o pacote P , o nó selecionado j envia ao nó i a menor estimativa do valor-Q presente em sua tabela com relação ao destino d , que corresponde ao tempo t remanescente na rota para alcançar d , representado pela eq.(13).

Ao ser enviado de i a j , o pacote P carrega a informação sobre o tempo decorrido desde a sua geração no nó o até alcançar o nó i . Esta informação corresponde ao menor valor de tempo estimado por i para alcançar o nó o e é representada por n :

$$n = \arg \min_l Q_i(l, o) \quad (24)$$

na qual $l \in$ conjunto de nós vizinhos de i .

3 - A partir destas informações, o nó i atualiza sua estimativa com relação a escolha do nó vizinho j em direção ao destino d , de acordo com a eq.(14), como no Roteamento-Q. A taxa de aprendizado utilizada nesta equação, denominada η_b , foi igual a 0,7. Assim também, o nó j atualiza sua estimativa sobre o tempo para alcançar o nó o utilizando o nó vizinho i da seguinte forma:

$$\Delta Q_j(i, o) = \eta_b [(n + q) - Q_j(i, o)] \quad (25)$$

na qual: η_b = taxa de aprendizado utilizada na exploração *backward*; os autores utilizaram 0,9.

q = tempo de espera do pacote P na fila do nó j .

Observa-se, assim, que o algoritmo de Roteamento-Q com Reforço Dual é semelhante ao do Roteamento-Q. As diferenças encontram-se somente na forma de atualização realizada pelo algoritmo. Ao invés de realizar somente a atualização do valor-Q presente na tabela do nó i com relação ao destino d do pacote, também é atualizado o valor-Q armazenado na tabela de seu vizinho j com relação ao nó origem o do mesmo. Além disso, a exploração *backward* produz resultados mais precisos que a *forward*, uma vez que utiliza valores reais

do tempo decorrido no transporte do pacote pela rede, ao invés de estimativas sobre o tempo remanescente para alcançar o destino.

O algoritmo também foi empregado na rede de 36 nós apresentada na Figura 6. Os experimentos foram realizados sob variações dos níveis de carga da rede. Foram utilizadas três classes de valores: nível baixo (de 0.5 a 1.5 pacotes gerados por tempo de simulação), nível médio (de 1.75 a 2.25 pacotes gerados por tempo de simulação) e nível alto (acima de 2.5 pacotes gerados por tempo de simulação). Seu desempenho foi comparado ao do algoritmo Roteamento-Q e do algoritmo de menor caminho.

Sob condições de nível de carga baixo, o roteamento de menor caminho produz os melhores resultados. O Roteamento-Q aprende uma política próxima da ótima, porém ele necessita de um tempo aproximadamente três vezes maior que o Roteamento-Q com Reforço Dual para atingir este comportamento. Ainda assim, este último obtém uma política levemente superior. Sob condições de nível de carga médio e alto, o Roteamento-Q com Reforço Dual também aprende uma política de roteamento eficiente em um período de tempo superior ao Roteamento-Q original, em aproximadamente duas vezes.

Verifica-se, portanto, que o Roteamento-Q com Reforço Dual apresenta melhor desempenho que o algoritmo original em todos os experimentos realizados. Uma possível limitação a sua utilização seria o acréscimo produzido na carga computacional do algoritmo relativo ao maior número de atualizações realizadas. Porém, o acréscimo observado foi de apenas 0.1%, o que torna a implementação do algoritmo uma alternativa viável e eficiente para realizar o roteamento de uma rede de telecomunicações.

A capacidade de armazenamento de pacotes pelos roteadores utilizada nos experimentos apresentados foi considerada ilimitada, o que constitui uma simplificação das características reais destes equipamentos. Assim, a implementação do algoritmo de Roteamento-Q com Reforço Dual levando em consideração a capacidade de armazenamento de pacotes limitada dos roteadores seria de grande relevância para a verificação do seu desempenho sob condições de variação do nível de carga da rede, aproximando-o da realidade das redes de telecomunicações atuais.

Os experimentos realizados com os algoritmos de roteamento baseados na aprendizagem-Q apresentados levaram em consideração dois dos parâmetros principais das redes de telecomunicações, que constituem o nível de carga e os padrões de tráfego. Desta forma, não é possível obter conhecimento sobre o seu comportamento em relação a variações na topologia da rede, que também constitui um dos parâmetros fundamentais de operação das redes de telecomunicações.

Além disto, verificou-se que o algoritmo de Roteamento-Q não é capaz de aprender uma política que descubra continuamente os menores caminhos, uma vez que somente as melhores estimativas dos valores-Q armazenados pelos vizinhos são sempre atualizadas. Se um destes vizinhos subestimar o tempo de entrega de um pacote para uma rota que não é a ótima, esta será selecionada até o momento em que outro caminho apresente um tempo menor, quando então os nós pertencentes a esta já poderão se encontrar congestionados.

6.2 - Roteamento Baseado no Comportamento Coletivo de Formigas

O método de otimização baseado no comportamento coletivo de formigas tem sido aplicado com sucesso ao problema de roteamento em redes de telecomunicações devido a utilização de um conjunto de agentes móveis simples, que realizam suas tarefas de forma distribuída e se comunicam indiretamente. A interação entre todos os agentes do sistema permite que as melhores soluções sejam encontradas, possibilitando também a adaptação rápida e eficaz dos agentes às diversas situações enfrentadas pelas redes.

A seguir, são descritos e analisados dois algoritmos de roteamento baseados no comportamento de colônias de formigas aplicados a redes de comutação de circuitos. O primeiro algoritmo pertencente a esta classe, desenvolvido por Schoonderwoerd et al. (1996, 1997), obteve desempenho superior tanto a métodos convencionais de roteamento como a outras técnicas baseadas em agentes móveis. A partir destes resultados promissores obtidos, outras versões deste sistema foram desenvolvidas por Bonabeau et al. (1998), Di Caro & Dorigo (1998) e White et al. (1998).

6.2.1 – Controle Baseado em Formigas

Dorigo et al. (1999) argumentam que o trabalho desenvolvido por Schoonderwoerd et al. (1996, 1997) constitui a primeira aplicação de um algoritmo de otimização baseado em colônias de formigas (*Ant Colony Optimization* - ACO) a um problema de roteamento em redes de telecomunicações. O algoritmo por eles desenvolvido, denominado Controle Baseado em Formigas, (*Ant-Based Control* - ABC) procura realizar o balanceamento da carga em uma rede de telecomunicações, que consiste em distribuir o volume de chamadas presentes na rede ou um possível aumento deste em regiões que possuam capacidade de processamento disponível, evitando assim o congestionamento da mesma.

Este sistema foi aplicado a um modelo da rede de telefonia da empresa de telecomunicações britânica (*British Telecom* - BT), apresentado na Figura 7. A rede foi modelada por um grafo $G = (R, T)$, onde cada nó i pertencente ao conjunto R representa um roteador com capacidade de processamento de chamadas limitada. Os laços pertencentes ao conjunto T representam os meios físicos de transmissão entre os roteadores; tais laços apresentam capacidade ilimitada de processamento.

Figura 7 - Rede de telecomunicações da *British Telecom* usada por Schoonderwoerd et al. (1996).

Desta forma, a capacidade de processamento dos nós é o único fator limitante na operação da rede. Cada nó i possui um número que o identifica e um conjunto de vizinhos denominado N_i . Cada nó é caracterizado por sua capacidade total de processamento C_i e por sua capacidade ociosa S_i . C_i representa o número máximo de conexões que o roteador i pode estabelecer, enquanto S_i representa a porcentagem da capacidade de processamento que ainda encontra-se disponível para novas conexões. Todos os nós possuem uma tabela de roteamento com $(n - 1)$ linhas e m_i colunas. Neste caso, n é o número de nós da rede e m_i é o número de vizinhos de cada nó i . Cada laço (i, j) que conecta o nó i ao nó j possui a ele associado um vetor com valores relacionados com as trilhas de feromônio construídas pelas formigas. Os valores armazenados na tabela de roteamento correspondem a estes vetores, representados por $P_i(j, d)$, onde $d = 1, \dots, i - 1, i + 1, \dots, n$. O valor de $P_i(j, d)$ representa uma medida da qualidade da escolha da linha (i, j) , quando o destino é o nó d . Além disso, cada nó possui uma probabilidade de ser um ponto terminal de uma chamada (origem ou destino da mesma). Esta distribuição de probabilidades constitui o padrão de tráfego aplicado à rede.

O algoritmo do sistema de controle baseado em formigas é descrito a seguir:

1 - As formigas são lançadas de cada nó i da rede em direção ao destino d , determinado de forma aleatória, em determinados intervalos regulares de tempo.

2 - Ao se movimentarem ao longo da rede, as formigas constróem os caminhos entre seus nós origem e destino aplicando uma regra de transição probabilística que utiliza a informação armazenada nas tabelas de roteamento. Esta informação corresponde aos valores associados às trilhas de feromônio $P_i(j, d)$. A política de decisão estocástica utiliza estes valores da seguinte forma: $P_i(j, d)$ representa a probabilidade de que uma formiga que possua como destino o nó d , seja enviada no instante de tempo considerado do nó i ao seu vizinho j . Um mecanismo de exploração, denominado fator de ruído, é adicionado à política de decisão da formiga: com uma pequena probabilidade (5%), as formigas podem escolher de modo uniformemente aleatório o vizinho a ser visitado. Este mecanismo procura lidar com o problema de estagnação das rotas. Através de sua utilização, assegura-se que rotas aparentemente ruins sejam percorridas ocasionalmente, de forma que se obtenha alguma informação sobre elas, a fim de que possam ser utilizadas em caso de bloqueio das rotas usualmente percorridas.

3 - Ao se movimentarem ao longo do ambiente, os agentes armazenam em seu estado interno os valores de seu nós origem e destino e o número de passos de tempo decorridos desde o seu lançamento. Este período de tempo é denominado idade da formiga, sendo determinado pela soma do número de nós visitados com o atraso sofrido em cada um deles. Este atraso é determinado de acordo com o volume de chamadas processadas pelo nó no momento da visita. Nenhuma memória sobre os nós visitados é mantida; desta forma, as formigas podem visitar o mesmo nó mais de uma vez, ou seja, podem ocorrer ciclos na rota percorrida.

no qual: idade = número de nós visitados + atraso sofrido em cada um destes. (26)

$$Atraso = 80 e^{-0.075 S_i} \quad (27)$$

no qual S_i é a porcentagem da capacidade ociosa do nó i no momento em que é visitado pela formiga ($0 < S_i < 100$).

4 - Ao visitar um nó, a formiga incrementa o valor armazenado na sua tabela correspondente ao nó de onde ela foi originada, com relação ao nó vizinho visitado no

instante anterior. Assim, uma formiga que foi gerada no nó o e encontra-se no nó j , vindo de seu nó vizinho i , atualiza o seguinte valor de probabilidade:

$$P_j(i, o) = \frac{P_j(i, o) + \Delta P}{1 + \Delta P} \quad (28)$$

na qual: $P_j(i, o)$ = nova probabilidade de que uma formiga que possua como destino o nó o seja enviada do nó j ao vizinho i ;

$$\Delta p = \frac{a}{idade} + b, \text{ é o incremento na probabilidade;} \quad (29)$$

a e b = constantes de proporcionalidade obtidas empiricamente.

Assim também, ela decrementa os valores dos demais vizinhos correspondentes ao seu nó de origem.

$$P_j(v, o) = \frac{P_j(v, o)}{1 + \Delta P} \quad (30)$$

na qual $v \in N_j$, que é o conjunto de nós vizinhos de j .

O valor $(1 + \Delta P)$ utilizado no denominador é um termo de normalização. Deste modo, a soma das probabilidades correspondentes a todos os vizinhos de j com relação ao nó o é sempre igual a um. O mecanismo de decremento de probabilidades corresponde à evaporação do feromônio depositado pelas formigas na natureza. Esta estratégia de atualização só pode ser aplicada quando a rede é simétrica com relação ao custo, ou seja, quando é possível utilizar a estimativa do custo de se movimentar do nó i ao nó j como uma estimativa do movimento no sentido inverso.

Deste modo, a quantidade de feromônio depositada pelas formigas é inversamente proporcional ao comprimento e ao nível de congestionamento do caminho selecionado. Assim, o efeito produzido pelas formigas nos valores das tabelas é tal que as rotas frequentemente visitadas e percorridas por formigas “jovens” são favorecidas na escolha dos caminhos que serão percorridos por novas chamadas.

5 - As chamadas são geradas a cada passo de tempo, seguindo a média de uma chamada. A origem e o destino da chamada são selecionados aleatoriamente de acordo

com a probabilidade de cada nó da rede constituir um ponto terminal. A duração da chamada é de 170 passos de tempo em média.

6 - Ao estabelecer os nós origem o e destino d , é selecionado de forma sequencial, a partir do nó o , o vizinho que possui o maior valor de probabilidade até que o nó destino d seja alcançado. Neste momento, a capacidade ociosa de cada nó que compõe a rota estabelecida, S_i , é decrementada por um valor fixo, correspondente ao inverso de sua capacidade máxima de processamento de chamadas. Se durante esta fase, qualquer um dos nós pertencentes a rota selecionada possui capacidade ociosa nula, a chamada é perdida.

7 - Quando a duração de uma chamada chega ao fim, a capacidade de processamento dos nós inclusos na rota é incrementada.

Este algoritmo foi testado na rede de telecomunicações britânica (*British Telecom-BT*) e seu desempenho com relação a porcentagem de chamadas perdidas foi comparado com outros três algoritmos: o primeiro deles também utiliza agentes móveis e foi desenvolvido por Appleby & Steward (1994) para a empresa BT; o segundo constitui uma modificação deste algoritmo realizada por Schoonderwoerd et al. (1996, 1997) e o terceiro constitui um algoritmo fixo baseado nos menores caminhos, que não realiza o balanceamento de carga. Os experimentos realizados consistiram na variação dos padrões de tráfego da rede. O algoritmo baseado no comportamento coletivo de formigas obteve melhores resultados que os demais sob todas as condições.

6.2.2 – Controle Baseado em Formigas “Espertas”

O algoritmo desenvolvido por Bonabeau et al. (1998) estende o algoritmo de Schoonderwoerd et al (1996,1997), introduzindo um mecanismo relacionado com os métodos de programação dinâmica anteriormente utilizado por Guérin (1997), os quais foram descritos por Bellman (1957), Bertsekas (1995) e Sutton e Barto (1998). Neste tipo de estratégia, a probabilidade de escolha de um nó vizinho considera um fator dependente da idade relativa do agente ao visitá-lo. Apesar do método de Guérin (1997) ter sido aplicado a agentes que se movimentam do nó origem ao nó destino e também no caminho inverso, é possível utilizá-lo em agentes que se movimentam em uma única direção. A fim de distinguir estes agentes daqueles desenvolvidos por Schoonderwoerd et al. (1996, 1997), estes foram denominados formigas “espertas”.

O algoritmo do sistema de roteamento baseado em formigas “espertas” é semelhante àquele do sistema de Controle Baseado em Formigas. A única diferença encontra-se na informação armazenada no estado interno das formigas e na forma de atualização das tabelas dos nós visitados pelos agentes.

O estado interno das formigas espertas armazena os valores dos nós visitados na rota percorrida, juntamente com a idade dos agentes ao alcançar cada um destes nós. Assim, ao visitar um nó em uma dada rota, a formiga atualiza todas as linhas da tabela que correspondem a nós visitados anteriormente ao longo da rota estabelecida. Cada um destes nós visitados será considerado um destino para a atualização da tabela.

Assim, uma formiga que foi gerada no nó o e encontra-se no nó j , vindo de seu nó vizinho i , atualiza os seguintes valores de probabilidade:

$$P_j(i, k) = \frac{P_j(i, k) + \Delta P_k}{1 + \Delta P_k} \quad (31)$$

no qual: $P_j(i, k)$ = probabilidade de que uma formiga que se destina ao nó k seja enviada do nó j ao vizinho i .

k = nós anteriormente visitados (de o a i).

Os incrementos são realizados utilizando a idade relativa dos agentes, ao invés da idade absoluta utilizada pelo sistema ABC :

$$\Delta P_k = \frac{a}{idade_j - idade_k} + b \quad (32)$$

no qual: ΔP_k = incremento da probabilidade de atingir o nó k ;

$idade_j$ = a idade absoluta da formiga ao alcançar o nó j ;

$idade_k$ = a idade absoluta da formiga ao visitar o nó k ;

k = nós anteriormente visitados (de o a i).

a e b = constantes de proporcionalidade obtidas empiricamente; os valores são iguais aos utilizados por Schoonderwoerd et al. (1996, 1997).

Assim, o incremento utilizado para atualizar o valor da probabilidade de j com relação a escolha do vizinho i em direção ao destino o será igual a idade absoluta da formiga ao atingir o nó j . Este comportamento do agente se assemelha àquele dos agentes do sistema de controle baseado em formigas.

As demais probabilidades armazenadas em cada uma destas linhas são decrementadas, de forma que a soma das mesmas continue a ser igual a 1:

$$P_j(v, k) = \frac{P_j(v, k)}{1 + \Delta P_k} \quad (33)$$

na qual $v \in N_j$, que é o conjunto de vizinhos de j .

Este algoritmo também foi aplicado ao modelo da rede de telefonia britânica apresentado na Figura 7. Os experimentos realizados e parâmetros utilizados são aqueles empregados por Schoonderwoerd (1996, 1997), para comparação de desempenho dos dois algoritmos sob diferentes padrões de tráfego da rede. O algoritmo baseado em formigas “espertas” obteve melhores resultados que o algoritmo baseado em colônias de formigas convencional em todos os experimentos.

Os algoritmos baseados no comportamento de colônias de formigas produziram melhores resultados tanto em relação aos algoritmos estáticos, utilizados de modo convencional pelas redes de comutação de circuitos, como a um algoritmo baseado em agentes móveis utilizado pela empresa de telecomunicações britânica. Entretanto, o desempenho destes sistemas foi avaliado apenas com relação à variação nos padrões de tráfego da rede. Deste modo, não se verifica o desempenho destes algoritmos sob condições variáveis dos demais parâmetros de operação da rede, como seu nível de carga e sua topologia.

O sistema de controle baseado em formigas foi aplicado a uma rede de tamanho médio. Os autores deste sistema afirmam que a estrutura do algoritmo desenvolvido pode ser aplicada a redes maiores de telecomunicações, tanto de comutação de circuitos como de comutação de pacotes. Eles não detêm conhecimento sobre o quanto as propriedades topológicas e estatísticas da rede influenciam a determinação ideal dos parâmetros do sistema.

Schoonderwoerd et al. (1996) afirmam que é fundamental a investigação dos efeitos causados no sistema em consequência dos parâmetros utilizados, como o número de formigas empregadas nos experimentos, a quantidade de tempo relacionada com o atraso imposto as mesmas, bem como a influência destes agentes sobre os valores de feromônios armazenados nas tabelas de roteamento dos nós.

Os autores afirmam que seria interessante verificar o desempenho do sistema de roteamento sob a utilização de uma política de decisão probabilística para as chamadas, em

todas as rotas selecionadas ou em apenas uma proporção destas. Outra alternativa seria a realização de decrementos das probabilidades armazenadas nas tabelas de feromônios em circunstâncias particulares, ao invés de seu incremento.

6.3 – Conclusão

Neste capítulo foram descritos e analisados os principais algoritmos inteligentes de roteamento desenvolvidos nesta década. Os algoritmos selecionados se baseiam na aprendizagem-Q e no método de otimização baseado no comportamento coletivo de formigas. Eles foram selecionados devido ao bom desempenho por eles apresentado sob diversas situações enfrentadas pelas redes de telecomunicações em relação tanto a algoritmos de roteamento clássicos como a sistemas propostos e utilizados recentemente por grandes empresas desta área. O comportamento satisfatório dos algoritmos apresentados nos levou a utilizar os princípios básicos de alguns deles no algoritmo que propusemos. Desta forma, procuramos combinar os melhores aspectos dos principais algoritmos de roteamento inteligentes e anexar alguns mecanismos de exploração, a fim de obter um sistema eficiente e eficaz face ao novo perfil das redes de telecomunicações.

Dentro deste contexto, no próximo capítulo será descrito em detalhes o algoritmo proposto, apresentando seus aspectos em comum com alguns dos algoritmos aqui descritos e aqueles anexados ao sistema, com o intuito de promover o equilíbrio adequado entre dinamismo e estabilidade diante de mudanças na rede.

7 – Agentes-Q: Um Algoritmo de Roteamento Adaptativo e Distribuído

Neste capítulo é descrito em detalhes o algoritmo de roteamento desenvolvido e aplicado a redes de comutação de circuitos. Este algoritmo combina características pertinentes a três algoritmos de roteamento baseados em aprendizagem-Q (Watkins, 1989), aprendizagem por reforço dual (Goetz et al., 1996) e o método de otimização baseado em colônias de formigas (*Ant Colony Optimization* - ACO), (Dorigo, 1992). Inicialmente é apresentada a justificativa para a combinação das técnicas de otimização utilizadas, seguida pela descrição dos componentes do sistema, do comportamento dos agentes e do roteamento de chamadas.

7.1 – Integração de Três Estratégias de Otimização para Realizar o Roteamento

O algoritmo de roteamento desenvolvido é resultante da integração de características de três algoritmos: (1) Roteamento-Q, baseado em aprendizagem-Q; (2) Roteamento-Q com Reforço Dual, baseado em aprendizagem por reforço dual e (3) Controle Baseado em Formigas, que utiliza o método de otimização baseado no comportamento de colônias de formigas. Estes algoritmos de roteamento foram desenvolvidos de forma sucessiva por diversos pesquisadores e aplicados em diferentes tipos de redes. As duas primeiras técnicas foram utilizadas em redes de comutação de pacotes, enquanto a última foi empregada tanto em redes de comutação de circuitos como de pacotes. Inicialmente, a escolha destes três algoritmos se deve ao bom desempenho apresentado por cada um deles em resolver o problema do roteamento. A combinação de suas principais características em um único algoritmo procurou reunir os melhores aspectos de cada um deles e obter assim, um sistema de roteamento adaptativo e distribuído que se adapte melhor e mais rapidamente às diversas situações enfrentadas pelas redes de telecomunicações atuais do que os algoritmos que o originaram. O algoritmo desenvolvido é resultante da integração das principais propriedades

dos algoritmos de roteamento descritos aliada a alguns mecanismos extras de otimização que foram anexados ao sistema.

A escolha de um algoritmo de roteamento baseado na aprendizagem-Q se deve a capacidade desta abordagem de aprender um comportamento adequado às condições de um meio ambiente sem possuir conhecimento prévio sobre o mesmo, como também a capacidade de adaptação contínua e em tempo real às situações diversas e inesperadas enfrentadas pelo meio. Estas características vão de encontro às aquelas apresentadas por um sistema de roteamento em uma rede de telecomunicações, onde não é possível ou é muito custoso se obter informação global completa e atualizada sobre as variações no estado da rede e do tráfego a cada instante. Além disto, a aprendizagem-Q tem sua convergência comprovada (Watkins & Dayan, 1992). A utilização de aspectos de um algoritmo de roteamento que adiciona a aprendizagem por reforço dual à aprendizagem-Q está relacionada com sua capacidade de atualizar informação sobre o ambiente no sentido direto e inverso ao movimento do agente, o que permite que o sistema aprenda uma política melhor e em um menor período de tempo do que aquele que utiliza somente a primeira estratégia. O bom desempenho destas duas técnicas em realizar o roteamento em redes de comutação de pacotes motivou o seu emprego no roteamento em uma rede de comutação de circuitos.

A escolha de um algoritmo baseado no comportamento coletivo de formigas se deve à utilização de um conjunto de agentes móveis simples, que realizam o roteamento de forma distribuída e concorrente. Os agentes cooperam entre si na busca pelos menores caminhos, trocando informações sobre as características da rede e a qualidade da rota por eles percorrida de forma indireta. Assim, os agentes aprendem a se comportar de modo incremental e contínuo, a partir de informação atualizada sobre as condições da rede. A utilização de uma política de decisão probabilística pelos agentes, baseada somente em informação local, possibilita a obtenção de rotas alternativas, que podem ser utilizadas sob variações nas condições da rede e do tráfego. A interação entre todos os indivíduos do sistema permite que os melhores caminhos sejam encontrados.

Os agentes baseados no comportamento de colônias de formigas acrescentam à aprendizagem-Q a capacidade de realizar tarefas de forma distribuída e assíncrona e de simultaneamente trocar informações sobre a qualidade das ações realizadas, através de um processo de comunicação indireta, o que aumenta a velocidade e qualidade de adaptação do sistema a mudanças no ambiente. Assim também, a seleção das ações realizadas pelos agentes é probabilística, permitindo que soluções alternativas sejam encontradas e aumentando consequentemente a sua capacidade de resposta às mudanças nas condições do ambiente.

Deste modo, foi realizada a integração das três estratégias de otimização acima mencionadas, obtendo-se um algoritmo de roteamento distribuído e adaptativo, capaz de acompanhar de forma mais eficiente e eficaz um número de situações enfrentadas pelas redes de telecomunicações superior àquele considerado pelos algoritmos anteriormente desenvolvidos.

Assim também, foram realizadas algumas modificações no algoritmo proposto com relação a simples junção dos três métodos anteriores, para aumentar a robustez e flexibilidade do roteamento às diversas variações dos parâmetros da rede, tanto temporárias como permanentes. Deste modo, foi introduzido um mecanismo na seleção das rotas a serem percorridas pelas chamadas que procura balancear a informação armazenada nas tabelas e a situação corrente dos roteadores com relação à carga processada, a fim de diminuir a diferença entre os dados coletados pelos agentes e as condições atuais dos nós, que são altamente dinâmicas e estocásticas. Assim também, um mecanismo de reforço negativo foi anexado na atualização dos nós em direção ao destino desejado, com o objetivo de evitar que nós altamente carregados sejam constantemente utilizados pelas chamadas e se tornem congestionados, enquanto outros nós se encontram com grande parte de sua capacidade de processamento ociosa.

7.2 – Características do Algoritmo Agentes-Q

Nesta sub-seção serão descritos inicialmente os componentes do algoritmo proposto, seguidos pelo comportamento dos agentes-Q e pelo roteamento das chamadas.

7.2.1 – Componentes do Sistema

O sistema de roteamento desenvolvido apresenta os seguintes componentes:

- ♦ Um conjunto de estados envolvendo todos os roteadores da rede, os quais podem ser visitados pelos agentes em seu movimento em direção ao destino desejado.
- ♦ Um conjunto de ações que estabelece as conexões utilizadas pelos agentes ao longo de sua trajetória.
- ♦ Um conjunto de agentes que se movimentam entre estados adjacentes do sistema, selecionando suas ações através de uma política de decisão probabilística, baseada nos caminhos que apresentam menores comprimento e nível de congestionamento.

- ♦ Dois sinais locais de retorno: um relacionado com o período de tempo dispendido pelo agente para percorrer a rota selecionada e outro com o valor da capacidade de processamento ociosa apresentada pelo nó atual.
- ♦ Uma função de avaliação dos pares estado-ação, $Q_i(j,d)$, que representa a probabilidade de um agente que se encontra em i escolher o vizinho j para alcançar o destino d . Esta é uma estimativa, que representa o valor esperado dos retornos recebidos ao longo do caminho estabelecido a partir do estado atual do sistema e da ação selecionada. Os valores-Q são armazenados na tabela de roteamento de cada nó i da rede e utilizados pelos agentes para se movimentarem ao longo da rede.
- ♦ Uma função de avaliação dos nós vizinhos de i com relação a um determinado destino, que determina a escolha do nó a ser percorrido por uma chamada. Esta função combina os valores-Q dos nós candidatos com sua capacidade de processamento disponível no momento da construção da rota. Assim, $A_i(j,d)$ representa a avaliação do vizinho j de i em direção ao destino d .

Os valores $Q(s,a)$, que correspondem a qualidade de realizar a ação a a partir do estado s em que o agente se encontra, são armazenados nas tabelas de roteamento da seguinte forma:

Tabela 1 – Tabela de roteamento do nó i .

Nós destinos	Nós vizinhos			
	v_1	v_2	...	v_k
d_1	$Q_i(v_1, d_1)$	$Q_i(v_2, d_1)$	...	$Q_i(v_k, d_1)$
d_2	$Q_i(v_1, d_2)$	$Q_i(v_2, d_2)$	...	$Q_i(v_k, d_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
d_j	$Q_i(v_1, d_j)$	$Q_i(v_2, d_j)$	...	$Q_i(v_k, d_j)$

na qual: d_j = nó destino j , $j = 1, 2, \dots, n-1$;

v_k = nó vizinho k , $k = 1, 2, \dots, m_i$;

m_i = número de vizinhos de i ;

n = número de nós da rede;

$Q_i(v_k, d_j)$ = probabilidade estimada pelo nó i com relação a escolha de seu vizinho v_k em direção ao destino d_j .

7.2.2 – Comportamento dos Agentes

Do mesmo modo que as formigas artificiais discutidas no Capítulo 5, os agentes empregados neste algoritmo selecionam as rotas de forma simultânea e assíncrona. O algoritmo seguido pelos agentes-Q é descrito a seguir:

1 - A cada passo de tempo, um agente é lançado de cada nó da rede em direção a um destino determinado de forma aleatória.

2 - Os agentes se movimentam através de uma sequência de estados adjacentes do ambiente. A seleção do próximo nó a ser visitado é realizada de forma probabilística, baseada nos valores-Q armazenados na tabela de roteamento do nó em que o agente se encontra. Um mecanismo de exploração, denominado fator de ruído, é adicionado ao comportamento dos agentes. De acordo com este mecanismo, o próximo nó a ser percorrido possui 5% de probabilidade de ser selecionado através de uma distribuição uniformemente aleatória.

3 - Ao se movimentarem ao longo do ambiente, os agentes armazenam em sua memória os nós origem e destino e o período de tempo decorrido desde o seu lançamento. Este período de tempo é denominado idade do agente e constitui um dos sinais de reforço utilizados pelo algoritmo. Como no sistema ABC, a idade do agente é determinada pela soma do número de nós visitados com o atraso sofrido em cada um deles. Este atraso é determinado de acordo com o volume de chamadas processadas pelo nó no momento da visita. O cálculo do atraso é realizado através da eq.(27) empregada pelo sistema ABC, que foi determinada empiricamente. O segundo sinal de reforço utilizado pelos agentes em sua trajetória é função da porcentagem da capacidade máxima de processamento do nó atual que está sendo utilizada no momento da visita, denominada carga do nó. Assim, um agente originado em o que utiliza a conexão (i,j) em direção ao destino d . Ao visitar o nó j , o agente recebe os seguintes sinais de reforço do ambiente:

$$r_j = \frac{\beta}{idade\ do\ agente} + \rho \quad (34)$$

$$r_i = w \left(\frac{\phi}{C_j + \delta} \right) \quad (35)$$

nas quais: r_i = reforço obtido do ambiente em decorrência da ação realizada a partir do

nó i ;

C_j = carga do nó j ;

$\beta, w, \phi, \delta, \rho$ = constantes de proporcionalidade.

4 – A partir destes sinais, o agente atualiza dois valores-Q associados à conexão percorrida. Um destes valores encontra-se armazenado na tabela de roteamento do nó atual e se refere a escolha do nó anteriormente visitado com relação a origem do agente. O outro valor-Q encontra-se armazenado na tabela do último nó visitado e refere-se à escolha do nó em que a formiga se encontra com relação ao destino desejado.

Considere um agente originado no nó o que utiliza a conexão (i,j) em direção ao destino d . Ao visitar o nó j , o agente atualiza os seguintes valores-Q: (1) o valor $Q_i(j,d)$ correspondente a escolha de j para alcançar o destino d e (2) o valor $Q_j(i,o)$ correspondente a escolha de i como próximo nó em direção ao destino o . As atualizações são realizadas da seguinte forma:

Atualização no sentido direto:

$$Q_i(j,d) = \frac{Q_i(j,d) + \Delta Q_i(j,d)}{\sum_v Q_i(v,d)} \quad (36)$$

Atualização no sentido inverso:

$$Q_j(i,o) = \frac{Q_j(i,o) + \Delta Q_j(i,o)}{\sum_z Q_j(z,o)} \quad (37)$$

$$\text{nas quais: } \Delta Q_i(j,d) = \alpha_i [r_i + \gamma_i \max_z Q_j(z,d) - Q_i(j,d)] \quad (38)$$

$$\Delta Q_j(i,o) = \alpha_j [r_j + \gamma_j \max_v Q_i(v,o) - Q_j(i,o)] \quad (39)$$

γ_i, γ_j = fatores de desconto

$$v = \text{nó} \in N_i.$$

$$z = \text{nó} \in N_j.$$

Os denominadores das eqs.(36) e (37) constituem termos de normalização dos valores-Q, a fim de que a soma dos valores de todos os vizinhos de um nó com relação a um determinado destino permaneça igual a 1.

Quando a carga do nó j é igual ou superior a um determinado valor obtido empiricamente, no caso 60%, a probabilidade $Q_i(j,d)$ é decrementada por ΔQ , ao invés de ser incrementada. Esta estratégia procura penalizar os nós que se encontram processando um grande número de chamadas no momento da visita do agente. Este mecanismo é utilizado em problemas de aprendizagem por reforço para representar uma ação que deve ser evitada. Neste caso, queremos reduzir a probabilidade de escolha de nós que encontram-se utilizando grande parte de sua capacidade de processamento e favorecer os nós que possuam maior capacidade disponível, realizando o balanceamento de carga sobre a rede e consequentemente reduzindo o número de chamadas perdidas.

Os valores-Q relativos aos demais nós vizinhos de j com relação ao destino o e aqueles referentes aos demais vizinhos de i com relação ao destino d também são atualizados, de forma que a soma dos novos valores seja mantida igual a 1:

Atualização no sentido direto:

$$Q_i(v,d) = \frac{Q_i(v,d)}{\sum_v Q_i(v,d)} \quad (40)$$

Atualização no sentido inverso:

$$Q_j(z,o) = \frac{Q_j(z,o)}{\sum_z Q_j(z,o)} \quad (41)$$

Observa-se que os sinais de reforço recebidos pelos agentes-Q favorecem as rotas que realizam o balanço entre os menores caminhos e os nós menos carregados.

A estratégia de atualização utilizada pode ser aplicada quando a rede de telecomunicações é simétrica com relação aos custos de utilização das conexões, ou seja, quando é possível utilizar uma mesma estimativa de custo para se movimentar do nó i ao nó j em relação a um determinado destino ou para o movimento no sentido inverso. No modelo de rede utilizado, esta é uma consequência direta das

características atribuídas aos roteadores e conexões empregados, que serão apresentadas na próxima seção.

5 – Quando um agente-Q atinge o nó destino, ele é eliminado do algoritmo.

7.2.3 – Roteamento de Chamadas

As chamadas são introduzidas na rede após um determinado período de tempo, denominado inicialização, no qual existem somente agentes na rede. O algoritmo seguido pelas chamadas é realizado da seguinte forma:

1 – Antes que a chamada seja enviada, toda a rota é construída, selecionando-se os nós a serem utilizados em direção ao destino de forma sequencial e determinística. O nó vizinho escolhido é aquele que apresenta a melhor avaliação. Assim, a avaliação do vizinho j do nó i com relação ao destino d é calculada da seguinte forma:

$$A_i(j, d) = (1 - \omega)Q_i(j, d) + \omega O_j \quad (42)$$

na qual: ω = fator de peso, $0 < \omega < 1$;

$$\omega = \begin{cases} \omega_1 & \text{se a carga de } j \text{ é crescente} \\ \omega_2 & \text{se a carga de } j \text{ é decrescente} \end{cases}$$

$$\omega_1 = \frac{1}{1 + \exp[(-k_1 C_j) + k_2]}; \quad (43)$$

$$\omega_2 = \frac{1}{1 + \exp[(-k_3 C_j) + k_4]}; \quad (44)$$

C_j = carga do nó j ;

k_1, k_2, k_3, k_4 = constantes de proporcionalidade obtidas empiricamente;

O_j = capacidade ociosa do nó j .

Estas equações favorecem inicialmente o nó vizinho que possui maior valor-Q e constitui assim, o menor caminho. À medida que os menores caminhos são continuamente utilizados, eles se tornam propensos ao congestionamento. Então, quando o nível de carga do nó candidato atinge o nível de aproximadamente 93% de sua capacidade máxima de processamento, o peso atribuído à capacidade ociosa do nó (representado por ω) passa a crescer mais rapidamente. Desta forma, os nós que estão

congestionados ou próximos desta condição não são selecionados por um determinado período de tempo, durante o qual rotas alternativas são percorridas. Após este período, os menores caminhos se tornam novamente disponíveis para novas chamadas na rede. Assim, quando o nível de carga de um nó decresce e se encontra abaixo de aproximadamente 87% de sua capacidade máxima de processamento, a importância do seu valor- Q na eq.(42) passa a crescer significativamente com relação à sua capacidade ociosa. Esta estratégia procura realizar o balanceamento de carga sobre a rede em uma variedade de situações de tráfego, minimizando continuamente o número de chamadas perdidas pelo sistema.

2 - À medida que o caminho é construído, a capacidade ociosa O_i de cada nó i pertencente a rota selecionada é decrementada por um valor fixo. Se, no momento da construção da rota, um dos nós selecionados não possui capacidade de processamento disponível ou já pertence à rota atual, a chamada é perdida.

3 - Quando a duração de uma chamada chega ao fim, a capacidade de processamento de cada nó pertencente à rota selecionada é incrementada, se tornando novamente disponível para as demais chamadas presentes na rede.

O algoritmo do sistema de roteamento desenvolvido é descrito em pseudo-código a seguir.

Seja:

t = passo de tempo atual, t_{max} = passo de tempo máximo da simulação;

o = nó origem, d = nó destino;

a = nó atual, p = próximo nó;

i = nó qualquer da rede, $i = 1, 2, \dots, n$;

n = número de nós da rede;

C_i = capacidade de processamento máxima do nó i ;

$O_p(t)$ = capacidade ociosa do nó p no passo de tempo t ;

N_a = conjunto de nós vizinhos do nó a .

Função principal:

Repita (para cada t):
 Gere formigas
 Processe formigas
 Gere chamadas
 Processe chamadas
 Até que $t = t_{max}$

Função Gere formigas:

Repita (para cada nó i da rede):
 Selecione aleatoriamente o nó destino d
 Armazene na memória da formiga seus nós origem e destino
 Estabeleça idade da formiga igual a zero
 Lance formiga
 Até que $i = n$.

Função Processe Formigas:

Repita (para cada formiga na rede):
 Se ($a \neq d$)
 Escolha um nó vizinho p de acordo com $Q_a(p, d)$, $p \in N_a$
 Envie formiga através da conexão (a, p)
 $a = p$
 Armazene na memória da formiga o nó p
 Incremente a idade da formiga em um passo
 Se ($o_p(t) > (1/C_p)$)
 Atualize valores-Q da tabela de roteamento de a :
 Incremente $Q_a(p, d)$
 Decremente $Q_a(v, d)$, $\forall v \in N_a, v \neq p$
 Atualize valores-Q da tabela de roteamento de p :
 Incremente $Q_p(a, o)$
 Decremente $Q_p(z, o)$, $\forall z \in N_p, z \neq a$
 Senão
 Atrase a formiga em p
 Incremente a idade da formiga de acordo com o atraso atual
 Fim se
 Senão
 Apague a formiga
 Fim se
 Até que o numero de formigas processadas seja igual ao total de formigas

Função Gere Chamadas:

Gere em média uma chamada
 Selecione os nos origem e destino de acordo com a distribuição de probabilidade
 Determine a duração da(s) chamada(s)

Função Processe Chamadas:

```

Repita (para cada chamada na rede):
  Se ( $\alpha \neq d$ )
    Escolha nó vizinho  $p$  tal que  $p = \max_p A_p(\alpha, d)$ 
    Acrescente o nó  $p$  à rota desta chamada
    Se ( $o_p(t) > (1/C_p)$ )
      Decremente  $o_p(t)$ 
    Senão
      Apague a chamada
      Incremente número de chamadas perdidas na rede
    Fim se
  Senão
    Se ( $\alpha = d$ ) e (duração da chamada  $> 0$ )
      Envie chamada através da rota selecionada
      Decremente duração da chamada em um passo de tempo
    Senão
      Incremente o número de chamadas recebidas na rede
      Apague a chamada
    Fim se
  Fim se
Até que o número de chamadas seja igual ao total de chamadas

```

7.3 - Conclusão

Neste capítulo foi descrito em detalhes o algoritmo de roteamento Agentes-Q desenvolvido neste projeto de pesquisa, evidenciando suas principais características, que constituem aspectos pertinentes a três estratégias de otimização acrescidos de um mecanismo extra de exploração da rede. A justificativa para a seleção destas técnicas foi apresentada, juntamente com o objetivo de sua integração em um único sistema de roteamento.

Dentro deste contexto, o próximo capítulo apresentará os experimentos realizados com este algoritmo, descrevendo os aspectos fundamentais dos componentes da rede à qual ele foi aplicado, juntamente com a definição dos parâmetros utilizados. Assim também, serão apresentados os resultados obtidos pelos agentes-Q diante de diversas condições de operação da rede e comparados com aqueles produzidos pelos algoritmos de roteamento ABC e ABC com agentes espertos, que também foram implementados. As vantagens e limitações de cada algoritmo serão discutidas e analisadas.

8 – Experimentos Realizados

Neste capítulo são descritos os experimentos realizados com o algoritmo de roteamento Agentes-Q e com os algoritmos ABC e ABC com agentes espertos, ambos baseados no comportamento coletivo de formigas. O algoritmo ABC foi selecionado devido ao seu desempenho superior em relação a um algoritmo clássico baseado nos menores caminhos e a um algoritmo baseado em agentes móveis que é utilizado pela empresa de telecomunicações britânica *British Telecom*. O algoritmo ABC com agentes espertos foi escolhido devido ao fato de ter obtido resultados melhores que o algoritmo original ABC.

Inicialmente, são descritas as características das redes de comutação de circuitos às quais os algoritmos foram aplicados, seguidas pelos parâmetros utilizados por eles e pelos objetivos dos testes realizados. Finalmente, os experimentos são discutidos e os principais resultados e limitações dos algoritmos implementados são analisados.

8.1 – Características das Redes Utilizadas

As redes de comutação de circuitos utilizadas em nossos experimentos são: (1) o modelo da rede de telefonia da empresa britânica (Figura 8) apresentado anteriormente na Seção 6.2.1, a qual foi utilizada inicialmente pelos algoritmos ABC e ABC com agentes espertos, e (2) um modelo da rede de telefonia utilizada pela empresa norte-americana AT&T (Figura 9). As características dos componentes das redes foram mantidas idênticas às aquelas utilizadas por Schoonderwoerd et al.(1996; 1997) e Bonabeau et al.(1998).

Figura 8 – Rede de telecomunicações da *British Telecom* com números de identificação.

Figura 9 – Rede de telecomunicações da *AT&T*.

Os modelos das redes utilizadas constituem dois grafos de 30 e 87 nós, onde cada nó i possui as seguintes características:

- ♦ Número de identificação. No primeiro modelo, os números pertencem ao conjunto de números reais $[0, 29]$ e no segundo, ao conjunto $[0, 86]$.
- ♦ Capacidade de processamento de chamadas máxima, C_i , igual a 40 chamadas.

- ♦ Capacidade de processamento de chamadas ociosa, O_i , pertencente ao conjunto de números reais $[0, 100]$, que corresponde a porcentagem da capacidade máxima que encontra-se disponível. A cada chamada processada em um nó, a capacidade ociosa do mesmo é decrementada em 2.5%.
- ♦ Conjunto de vizinhos denominado N_i , formado pelos nós adjacentes a i .
- ♦ Probabilidade de constituir um nó terminal de uma chamada. Estas probabilidades são denominadas probabilidades de chamadas e determinadas através de uma distribuição aleatória em $[0.01, 0.07]$, sendo normalizadas após sua geração. O conjunto de probabilidades de todos os nós da rede constituirão pontos terminais de chamadas caracteriza a distribuição geográfica das mesmas ao longo da rede e representa assim, os padrões de tráfego aplicados a rede.
- ♦ Tabela de roteamento com número de linhas correspondente aos possíveis destinos na rede, no caso do primeiro modelo 30 e do segundo 87, e com número de colunas determinado pelo número de nós em cada conjunto N_i .

Cada laço que conecta dois nós quaisquer da rede possui capacidade de processamento de chamadas ilimitada. Deste modo, a capacidade de processamento dos nós constitui o único fator limitante à operação da rede.

Os valores associados aos parâmetros utilizados pelo sistema de roteamento desenvolvido são descritos a seguir.

8.2 – Parâmetros dos Algoritmos de Roteamento Implementados

Os valores atribuídos aos parâmetros utilizados na atualização dos valores-Q realizada pelos agentes-Q, α_i , w , ϕ , δ , γ_i , α_j , β e γ_j e aqueles usados na seleção das rotas percorridas pelas chamadas, k_1 , k_2 , k_3 e k_4 , foram determinados empiricamente para cada uma das redes utilizadas. Na primeira rede (Figura 8), durante os primeiros 500 passos de tempo, que constituem o período em que não há chamadas telefônicas na rede, mas somente a presença de agentes, foram utilizados os seguintes valores: $\alpha_i = 0.95$, $r_i = 0.0$, $\gamma_i = 1.0$, $\alpha_j = 0.95$, $\rho = 0.0$, $\beta = 6.0$ e $\gamma_j = 1.0$. Na segunda rede (Figura 9), os parâmetros assumiram os seguintes valores: $\alpha_i = 0.1$, $r_i = 0.005$, $\gamma_i = 1.0$, $\alpha_j = 0.1$, $\rho = 0.01$, $\beta = 90.0$ e $\gamma_j = 1.0$. Após 500 passos de tempo, quando então as chamadas são introduzidas na rede, os parâmetros da primeira rede assumem os seguintes valores: $\alpha_i = 0.08$, $w = 0.05$, $\phi = 1.0$, $\delta = 0.1$, $\gamma_i = 1.0$, α_j

$= 0.2$, $\beta = 8.0$ e $\gamma_j = 1.0$. Os parâmetros da segunda rede são os seguintes: $\alpha_i = 0.02$, $w = 0.05$, $\phi = 1.0$, $\delta = 0.1$, $\gamma_i = 0.9$, $\alpha_j = 0.06$, $\beta = 70.0$ e $\gamma_j = 0.8$. Isto se deve ao fato de que inicialmente, quando os agentes não possuem conhecimento sobre o meio ambiente, as correções dos valores-Q devem ser maiores, a fim de que os menores caminhos sejam rapidamente obtidos. A partir do momento em que há tráfego de chamadas na rede, as correções destes valores devem ser menores, de forma que o sistema se adapte de forma suave às mudanças nas características da rede. Os parâmetros utilizados no roteamento de chamadas são: $k_1 = 28.0$, $k_2 = 26.0$, $k_3 = 33.5$ e $k_4 = 30.0$.

Os parâmetros utilizados na política de atualização dos algoritmos ABC e ABC com agentes espertos são os seguintes: $a = 0.08$ e $b = 0.005$. Nos testes envolvendo a segunda rede, estes valores foram modificados para: $a = 0.48$ e $b = 0.015$.

As mudanças destes parâmetros nos experimentos referentes à rede de telecomunicações norte-americana se deve ao fato de que eles foram determinados empiricamente para a primeira rede utilizada, a rede britânica. Deste modo, ao utilizarmos os algoritmos em uma nova rede, que possui topologia diferente e número de nós superior a anterior, tornou-se necessário modificar os parâmetros, a fim de que os algoritmos se aproximassem ou mantivessem o comportamento apresentado nos primeiros experimentos.

Os procedimentos referentes à simulação realizada são descritos a seguir.

8.3 – Detalhes da Simulação

O algoritmo de roteamento proposto foi escrito em C para ambiente Unix em uma *workstation* SunW, Ultra-1. Cada passo de tempo da simulação envolveu as seguintes etapas:

- ♦ Lançamento de uma formiga a partir de cada nó da rede, com destino determinado através de uma distribuição aleatória.
- ♦ Processamento de todas as formigas presentes na rede. Esta etapa compreende a escolha probabilística de um nó vizinho a ser visitado pelo agente e o envio deste indivíduo ao nó selecionado, ou o seu atraso no nó em que ele se encontra, de acordo com o nível de congestionamento do mesmo.
- ♦ Geração de uma chamada em média na rede, através de uma distribuição de Poisson. Os pontos de origem e destino são definidos através da distribuição de chamadas, que é aleatória. A duração da chamada é determinada por uma distribuição exponencial, com média de 170 passos de tempo.

- ♦ Processamento de todas as chamadas presentes na rede. Esta etapa consiste na seleção determinística da rota a ser percorrida pela chamada e na redução da capacidade de processamento ociosa dos nós que a constituem em 2.5%.

Os experimentos consistiram na geração de 10 conjuntos de probabilidades de chamadas, que representam os padrões de tráfego da rede. Cada conjunto destas probabilidades foi utilizado para gerar uma sequência de chamadas com duração de 15.000 passos de tempo. Cada uma destas sequências foi dividida em dois blocos, A e B, correspondentes a um período de adaptação e um período de teste, cada um com duração de 7.500 passos de tempo (Tabela 2).

Tabela 2 - Geração de sequências de chamadas (Schoonderwoerd et al., 1996).

Conjuntos de probabilidades	Período de Adaptação (0 – 7500)	Período de Teste (7500 -15000)
1	1A	1B
2	2A	2B
:	:	:
10	10A	10B

Cada simulação envolveu um período de inicialização correspondente a 500 passos de tempo, durante o qual não houve geração de chamadas na rede. Os valores-Q armazenados nas tabelas de roteamento foram inicializados seguindo uma distribuição uniforme, ou seja, todos os vizinhos de um determinado nó possuíam a mesma probabilidade de serem escolhidos inicialmente. O fator de ruído utilizado na seleção das rotas pelas formigas foi igual a 5%.

A descrição dos testes realizados com os algoritmos implementados é realizada a seguir.

8.4 – Testes Realizados

As redes de telecomunicações apresentam condições de operação dinâmicas e estocásticas, resultantes das constantes variações ocorridas em três parâmetros fundamentais: padrões de tráfego, nível de mensagens presentes na rede e topologia da mesma. Desta forma, um algoritmo de roteamento deve ser capaz de realizar suas funções de forma rápida e eficiente sob situações de mudanças em todos estes aspectos, mantendo um alto nível de desempenho e minimizando o custo de utilização dos recursos da rede.

Dentro deste contexto, os experimentos realizados com os algoritmos de roteamento implementados compreendem variações permanentes e temporárias dos três parâmetros acima descritos. Até o presente momento, os algoritmos de roteamento desenvolvidos foram testados sob variações em apenas uma ou duas destas características, sendo geralmente os padrões de tráfego ou o nível de mensagens presentes na rede.

Foram realizados 18 testes com os três algoritmos implementados em cada uma das redes selecionadas. Em todos os experimentos, o nível de carga na rede durante o período de adaptação foi mantido baixo (geração de 1 chamada em média), realizando-se variações nas condições da rede e do tráfego somente no período de teste. Em cada um dos experimentos, somente um dos parâmetros da rede foi modificado, a fim de verificar o efeito de cada um deles sobre o desempenho dos algoritmos considerados. Os quatro primeiros testes são similares àqueles realizados por Schoonderwoerd et al. (1996, 1997) e compreenderam a variação nos padrões de tráfego da rede, representados pela distribuição de probabilidades de chamadas. O quinto e sexto experimentos envolveram a variação permanente do volume de chamadas presentes na rede para os níveis médio e alto, respectivamente. O sétimo e oitavo testes também realizaram a variação do volume de chamadas presentes na rede para os níveis médio e alto, respectivamente, porém de forma temporária. O nono teste realizado envolveu a desconexão de um nó da rede de forma permanente. O nó selecionado na primeira rede foi o de número 10, por constituir um nó amplamente utilizado pelos agentes, devido ao fato de encontrar-se localizado na região central da rede e apresentar o maior número de vizinhos, igual a 7. Na segunda rede, o nó 60 foi o selecionado, também por ser bastante utilizado pelas chamadas, devido a sua localização e alta probabilidade de ser um nó terminal de uma chamada. O décimo experimento é semelhante ao anterior, porém as desconexões dos nós 10 e 60 foram realizadas de forma temporária. Os testes décimo primeiro a décimo quarto compreenderam a utilização permanente de um fator de ruído na seleção das chamadas. Este fator é igual ao utilizado pelas formigas, ou seja, o nó vizinho a ser percorrido por uma chamada apresenta uma determinada probabilidade de ser selecionado de modo uniformemente aleatório. Nos testes décimo quinto ao décimo oitavo, o ruído foi inserido temporariamente. Estes testes procuraram verificar a robustez dos algoritmos à presença de ruídos nas informações utilizadas na seleção das rotas.

Os quatro primeiros experimentos realizados são definidos da seguinte forma:

1. As chamadas geradas no período de adaptação pertencem ao mesmo conjunto de probabilidades que aquelas geradas no período de teste (Tabela 3), volume baixo

de chamadas na rede (geração de 1 chamada em média a cada passo de tempo) e lançamento contínuo de agentes.

Tabela 3 – Utilização do mesmo conjunto de probabilidades de chamadas nos períodos de adaptação e teste (Schoonderwoerd et al., 1996).

Período de Adaptação (0 – 7500)	Período de Teste (7500 -15000)
1A	1B
2A	2B
:	:
10A	10B

2. Como em (1), porém o lançamento de agentes foi interrompido durante o período de teste.
3. As chamadas geradas no período de adaptação não pertencem ao mesmo conjunto de probabilidades que aquelas geradas no período de teste (Tabela 4), volume baixo de chamadas na rede e lançamento contínuo de agentes.

Tabela 4 - Utilização de conjuntos de probabilidades de chamadas diferentes nos períodos de adaptação e teste.

Período de Adaptação (0 – 7500)	Período de Teste (7500 -15000)
1A	2B
2A	3B
:	:
10A	1B

4. Como em (3), porém o lançamento de agentes foi interrompido durante o período de teste.

A partir do quinto experimento, foi utilizada a mesma distribuição espacial de probabilidades de chamadas nos dois períodos e houve lançamento de formigas durante todo o período de teste. As variações ocorreram com relação ao nível de carga, topologia da rede ou introdução de ruído e são descritas a seguir:

5. A carga da rede foi alterada para o nível médio (geração de 1.5 chamada em média a cada passo de tempo) durante todo o período de teste.
6. A carga da rede foi alterada para o nível alto (geração de 2 chamadas em média a cada passo de tempo) durante todo o período de teste.

7. O nível de carga da rede foi alterado temporariamente para o nível médio (do passo de tempo 8.000 ao 9.000) e posteriormente restabelecido ao nível baixo.
8. O nível de carga da rede foi alterado temporariamente para o nível alto (do passo de tempo 8.000 ao 9.000) e posteriormente restabelecido ao nível baixo.
9. Um nó foi desconectado da rede durante todo o período de teste. No caso da rede britânica, o nó 10 foi selecionado e para a rede norte-americana, o nó 60 foi escolhido.
10. Um nó foi temporariamente desconectado da rede durante o período de teste (do passo de tempo 8.000 ao 9.000) e posteriormente restabelecido. Foram utilizados os mesmos nós do experimento anterior.
11. Foi inserido um fator de ruído na seleção das chamadas igual a 1% durante todo o período de teste.
12. Como em (11), porém o fator de ruído foi igual a 2%.
13. Como em (11), porém o fator de ruído foi igual a 5%.
14. Como em (11), porém o fator de ruído foi igual a 10%.
15. Foi inserido um fator de ruído na seleção das chamadas igual a 1% temporariamente durante o período de teste (do passo 8.000 a 9.000).
16. Como em (12), porém o fator de ruído foi igual a 2%.
17. Como em (12), porém o fator de ruído foi igual a 5%.
18. Como em (12), porém o fator de ruído foi igual a 10%.

A medida de desempenho utilizada e os resultados obtidos em todos os testes realizados são apresentados a seguir.

8.5 – Resultados Obtidos

O desempenho dos algoritmos foi monitorado durante os períodos de adaptação e teste, através da porcentagem média de chamadas perdidas a cada 500 passos de tempo, sobre os 10 conjuntos de probabilidades de chamadas gerados. Os resultados obtidos foram divididos em cinco conjuntos para cada rede utilizada, de acordo com as variações realizadas nos parâmetros das redes considerados. Em cada conjunto, são apresentadas inicialmente tabelas contendo os valores médios percentuais de chamadas perdidas e seus respectivos desvios-padrão. Posteriormente, são apresentados os gráficos com as médias de chamadas perdidas em cada conjunto de experimentos.

Os valores médios percentuais obtidos foram submetidos a um teste estatístico, denominado teste de hipóteses, para verificar se existe diferença significativa entre os resultados apresentados pelos três algoritmos implementados (Montgomery, 1997). A distribuição estatística utilizada foi a distribuição *t* de *Student*, com um nível de significância de 0.05.

8.5.1 – Rede de Telecomunicações Britânica

8.5.1.1 - Variações dos Padrões de Tráfego da Rede

Tabela 5 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,42%	4,59%	0,66%	0,19%
Agentes ABC espertos	4,48%	3,97%	0,52%	0,13%
Agentes-Q	1,65%	1,74%	0,26%	0,03%

Tabela 6 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,32%	4,64%	0,59%	0,12%
Agentes ABC espertos	4,62%	4,19%	0,52%	0,08%
Agentes-Q	1,66%	1,99%	0,27%	0,10%

Tabela 7 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,21%	5,05%	0,58%	0,04%
Agentes ABC espertos	4,40%	4,37%	0,43%	0,05%
Agentes-Q	1,55%	1,78%	0,35%	0,03%

Tabela 8 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,53%	5,88%	0,67%	0,36%
Agentes ABC espertos	4,75%	5,41%	0,45%	0,44%
Agentes-Q	1,61%	2,78%	0,29%	0,46%

Os testes estatísticos realizados demonstraram uma diferença significativa entre todos os resultados obtidos pelos três algoritmos implementados sob variações nos padrões de tráfego da rede. Os agentes-Q produziram uma porcentagem média de chamadas perdidas menor que a dos agentes ABC espertos, que por sua vez foi menor que aquela obtida pelos agentes ABC, em ambos os períodos de adaptação e teste. No caso de probabilidades de chamadas inalteradas e lançamento contínuo de agentes (Tabela 5), os desvios-padrão apresentados pelos algoritmos ABC e ABC com agentes espertos nos períodos de adaptação e teste foram muito maiores que aqueles obtidos pelo algoritmo proposto. Os valores de

desvio produzidos pelos três algoritmos sofreram um decréscimo no período de teste, quando todos os sistemas já se encontravam adaptados às condições da rede. De acordo com os resultados obtidos, observa-se que a interrupção do lançamento de agentes no período de teste (Tabela 6) sob probabilidades de chamadas inalteradas provocou um pequeno aumento na porcentagem de chamadas perdidas em todos os algoritmos implementados. Este comportamento demonstra a capacidade de todos os agentes de se adaptarem continuamente às situações enfrentadas pela rede. Os desvios-padrão obtidos pelos algoritmos ABC e ABC com agentes espertos no período de adaptação foram semelhantes àqueles obtidos no primeiro experimento, enquanto os do período de teste foram menores que os anteriores. Os agentes-Q, por sua vez, apresentaram um aumento nos valores de desvio obtidos no período de teste, indicando o maior efeito destes agentes no desempenho do sistema proposto que o dos demais agentes analisados sob seus respectivos algoritmos. A variação da probabilidade de chamadas (Tabela 7) provocou um acréscimo nos valores médios de chamadas perdidas obtidas pelos algoritmos ABC e ABC com agentes espertos superior àquele apresentado pelos agentes-Q. Estes resultados demonstram que o algoritmo proposto se adapta melhor à variação nos padrões de tráfego da rede considerada que seus competidores. Os desvios-padrão obtidos pelos algoritmos ABC e ABC com agentes espertos no período de adaptação foram menores que aqueles obtidos no primeiro teste, enquanto o algoritmo Agentes-Q apresentou um acréscimo nestes valores. Estas variações de desvio são comuns, uma vez que as decisões dos agentes são probabilísticas. No período de adaptação, o algoritmo proposto produziu o mesmo valor de desvio obtido no primeiro experimento, enquanto os demais obtiveram valores inferiores aos apresentados anteriormente. A interrupção do lançamento de formigas durante o período de teste aliada à variação dos padrões de tráfego da rede (Tabela 8) provocou o maior crescimento das perdas dos algoritmos neste conjunto de experimentos. Mesmo assim, o nosso algoritmo obteve valores médios percentuais de chamadas perdidas menores que os de seus competidores. Este comportamento confirma a capacidade superior dos agentes-Q de reagirem às mudanças nas condições do tráfego. Os desvios-padrão apresentados pelos três algoritmos sofreram um grande acréscimo em relação àqueles obtidos anteriormente, confirmando a influência dos agentes no roteamento adequado das chamadas.

Figura 10 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede britânica.

Figura 11 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.

Figura 12 – Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede britânica.

Figura 13 - Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede britânica.

No primeiro experimento (Figura 10), os algoritmos ABC e ABC com agentes espertos obtiveram uma porcentagem média de chamadas perdidas decrescente no período de teste, relacionada com o aprendizado dos agentes por eles utilizados. O algoritmo Agentes-Q, por sua vez, se adaptou rapidamente às condições da rede e manteve praticamente o mesmo nível de perdas em ambos os períodos de adaptação e teste. A interrupção do lançamento de formigas no segundo teste (Figura 11) reduziu o decréscimo

de chamadas perdidas obtido pelos algoritmos ABC e ABC com agentes espertos no período de teste, enquanto provocou nos agentes-Q um pequeno acréscimo na perda de chamadas. A mudança dos padrões de tráfego (Figura 12) causou inicialmente um aumento na porcentagem média de chamadas perdidas. Após o período de aprendizado, todos os algoritmos se adaptaram à nova situação e mantiveram as perdas em níveis praticamente constantes. O algoritmo proposto apresentou mais uma vez menores oscilações nos resultados obtidos. Sob variação dos padrões de tráfego e interrupção do lançamento de agentes (Figura 13), todos os algoritmos obtiveram um aumento significativo nos resultados. Este comportamento evidencia a importância da presença dos agentes para o roteamento eficiente das chamadas.

8.5.1.2 – Variações do Nível de Carga da Rede

Tabela 9 - Resultados obtidos sob alteração da carga da rede para o nível médio durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt.	Teste	Adapt.	Teste
Agentes ABC	5,47%	7,67%	0,49%	0,47%
Agentes ABC espertos	4,30%	7,18%	0,42%	0,53%
Agentes-Q	1,62%	5,14%	0,38%	0,59%

Tabela 10 - Resultados obtidos sob alteração da carga da rede para o nível alto durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt.	Teste	Adapt.	Teste
Agentes ABC	5,03%	11,52%	0,54%	1,27%
Agentes ABC espertos	4,14%	11,14%	0,40%	1,33%
Agentes-Q	1,35%	9,10%	0,32%	1,38%

Tabela 11 - Resultados obtidos sob alteração temporária da carga da rede para o nível médio durante o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt.	Teste	Adapt.	Teste
Agentes ABC	6,46%	7,50%	0,71%	0,57%
Agentes ABC espertos	5,02%	6,81%	0,41%	0,56%
Agentes-Q	1,80%	4,75%	0,29%	0,63%

Tabela 12 - Resultados obtidos sob alteração temporária da carga da rede para o nível alto durante o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt.	Teste	dapt.	Teste
Agentes ABC	5,28%	10,67%	0,79%	1,29%
Agentes ABC espertos	4,40%	10,28%	0,55%	1,27%
Agentes-Q	1,56%	8,88%	0,77%	1,47%

Os testes estatísticos realizados demonstraram que não houve diferença significativa entre os valores médios de chamadas perdidas apresentados pelos sistemas ABC e ABC com agentes espertos sob todos os experimentos envolvendo variação do nível de carga da rede. Os testes também demonstraram que o sistema Agentes-Q obteve valores médios de chamadas perdidas menores que os demais algoritmos em todas as mudanças realizadas no

nível de carga da rede. As alterações da carga da rede produziram grandes acréscimos na perda de chamadas em todos os algoritmos implementados, que foram proporcionais ao nível de carga considerado. Deste modo, a porcentagem média de chamadas perdidas pelos algoritmos sofreu um acréscimo sob condição permanente de nível alto de carga (Tabela 10) superior àquele obtido sob o nível médio (Tabela 9). Os testes estatísticos demonstraram que não houve diferença significativa entre os valores médios de chamadas perdidas produzidos pelos agentes ABC e ABC espertos. Os agentes-Q obtiveram, mais uma vez, as menores perdas. Os desvios-padrão obtidos pelos algoritmos ABC e ABC com agentes espertos sob variação permanente da carga para o nível médio (Tabela 9) foram similares àqueles apresentados no período de adaptação e muito superiores aos obtidos no primeiro experimento, sob condições constantes da rede (Tabela 5). O algoritmo Agentes-Q, por sua vez, obteve um valor de desvio elevado em relação ao período anterior, indicando a grande influência da mudança no nível de carga sobre o seu desempenho. Os desvios-padrão produzidos pelos três algoritmos sob alteração permanente da carga da rede para o nível alto (Tabela 10) sofreram um grande acréscimo, demonstrando a sensibilidade dos agentes a este tipo de mudança, em particular dos agentes-Q. A alteração temporária da carga da rede para o nível médio (Tabela 11) causou um aumento na porcentagem média de chamadas perdidas inferior àquele obtido sob a sua variação permanente em todos os algoritmos. Os valores de desvio-padrão dos três algoritmos foram um pouco maiores que os produzidos sob a variação permanente da carga para este nível. Este aumento será compreendido através dos gráficos apresentados a seguir. A alteração temporária da carga da rede para o nível alto (Tabela 12) também provocou um acréscimo nos valores de chamadas perdidas inferior àquele obtido sob sua variação permanente. Com relação aos desvios-padrão, isto ocorreu somente no caso dos agentes ABC espertos, demonstrando a sensibilidade dos demais algoritmos a esta situação da rede. Mais uma vez, os agentes-Q apresentaram os maiores valores de desvio-padrão, indicando que o efeito deste tipo de variação sobre o seu desempenho é superior que sobre o de seus competidores.

Figura 14 – Desempenho dos três algoritmos sob o nível médio de carga da rede durante todo o período de teste para a rede britânica.

Figura 15 – Desempenho dos três algoritmos sob o nível alto de carga da rede durante todo o período de teste para a rede britânica.

Figura 16 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível médio (do passo 8.000 ao 9.000) durante o período de teste para a rede britânica.

Figura 17 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível alto (do passo 8.000 ao 9.000) durante o período de teste para a rede britânica.

Os gráficos demonstram que sob variações permanentes da carga da rede, todos os algoritmos produziram inicialmente um grande acréscimo na porcentagem média de chamadas perdidas, proporcional ao nível de carga considerado. Após um período de tempo, os agentes aprenderam e se adaptaram à nova condição da rede. No caso do nível médio de carga (Figura 14), os agentes-Q inicialmente reagiram como seus competidores, produzindo um aumento nos valores de chamadas perdidas, porém inferior ao obtido pelos demais algoritmos. Entretanto, após um determinado período de tempo, também inferior ao

requerido pelos demais algoritmos, os agentes-Q se adaptaram à nova situação e obtiveram um decréscimo nas perdas, mais acentuado que aquele obtido pelos agentes ABC e agentes ABC espertos. Quando a carga da rede foi alterada de forma permanente para o nível alto (Figura 15), o algoritmo proposto obteve um grande acréscimo na porcentagem média de chamadas perdidas, novamente inferior ao produzido pelos demais algoritmos. Após um intervalo de tempo, os agentes-Q obtiveram um grande decréscimo nas perdas, mais uma vez superior ao apresentado pelos agentes ABC e ABC espertos. Sob a variação temporária da carga da rede para o nível médio (Figura 16), os agentes-Q se comportaram de forma semelhante aos demais, apresentando um aumento nos valores de chamadas perdidas. Quando a carga retornou ao nível baixo, todos os algoritmos se adaptaram à nova situação e obtiveram perdas decrescentes. Os agentes ABC e ABC espertos produziram um decréscimo nas perdas superior ao obtido sob a variação permanente da carga para o nível médio. A grande variação das perdas explica o aumento dos desvios-padrão obtidos pelos algoritmos sob esta situação superior a situação de alteração permanente da carga. A mudança temporária da carga da rede para o nível alto (Figura 17) provocou um comportamento similar entre todos os algoritmos, que obtiveram valores crescentes de perdas. Quando a carga retornou ao nível baixo, os agentes produziram um decréscimo nos valores de chamadas perdidas, levemente mais acentuado no caso dos agentes-Q. Como antes, as alterações nas perdas dos algoritmos justificam os elevados valores de desvio apresentados pelos agentes ABC e agentes-Q em relação aos obtidos sob a variação permanente da carga.

8.5.1.3 – Variações da Topologia da Rede

Tabela 13 – Resultados obtidos sob desconexão do nó 10 durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,17%	7,82%	0,58%	1,62%
Agentes ABC espertos	3,58%	6,92%	0,27%	1,67%
Agentes-Q	1,00%	5,07%	0,26%	1,99%

Tabela 14 - Resultados obtidos sob desconexão temporária do nó 10 durante o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	4,75%	6,14%	0,41%	0,51%
Agentes ABC espertos	4,10%	5,22%	0,35%	0,33%
Agentes-Q	1,33%	2,63%	0,25%	0,24%

Os testes estatísticos demonstraram que, em todos os casos envolvendo variação da topologia da rede considerados, os agentes-Q obtiveram valores médios percentuais de chamadas perdidas inferiores aos de seus competidores. A desconexão permanente do nó 10 (Tabela 13) produziu um grande aumento nas perdas obtidas por todos os algoritmos. Os testes estatísticos demonstraram que os valores médios de chamadas perdidas obtidos pelos

agentes ABC e ABC espertos não apresentaram diferença significativa. Os agentes-Q permaneceram com perdas menores que os demais. Os desvios-padrão produzidos também sofreram um acréscimo, mais acentuado no caso dos agentes-Q. Estes resultados indicam que o nó 10 participa ativamente do roteamento das chamadas e que sua retirada produziu maior efeito no desempenho do algoritmo proposto que no de seus competidores. No caso da desconexão temporária do nó 10 (Tabela 14), os três algoritmos apresentaram um aumento nas perdas inferior a desconexão permanente deste nó. Isto demonstra a importância do nó 10 para o roteamento eficiente da rede, como também a capacidade dos agentes de se adaptarem à situação em que ele é restabelecido a rede. Como consequência, os valores de desvio-padrão são significativamente menores que os obtidos sob a desconexão permanente do nó, principalmente no caso dos agentes-Q, mantendo-se praticamente constantes nos períodos de adaptação e teste.

Figura 18 – Desempenho dos algoritmos sob desconexão permanente do nó 10 no período de teste para a rede britânica.

Figura 19 – Desempenho dos algoritmos sob desconexão temporária do nó 10 no período de teste (do passo 8.000 ao 9.000) para a rede britânica.

Os gráficos demonstram que a desconexão do nó 10 (Figura 18) produziu um aumento significativo na porcentagem média de chamadas perdidas pelos três algoritmos implementados. A localização do nó 10 e o grande número de vizinhos que ele possui sugerem que a perda de chamadas sob sua desconexão seria naturalmente elevada e não apresentaria tendência de decréscimo. Quando este nó foi restabelecido a rede (Figura 19), todos os algoritmos se comportaram de modo similar, adaptando-se a nova situação e mantendo os valores de chamadas perdidas em níveis praticamente constantes.

8.5.1.4 – Utilização Permanente de um Fator de Ruído na Seleção das Rotas para as Chamadas

Tabela 15 – Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 1% durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,36%	4,74%	0,68%	0,10%
Agentes ABC espertos	4,54%	4,39%	0,51%	0,06%
Agentes-Q	1,64%	2,17%	0,26%	0,09%

Tabela 16 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 2% durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,38%	4,83%	0,57%	0,05%
Agentes ABC espertos	4,56%	4,57%	0,54%	0,05%
Agentes-Q	1,66%	3,02%	0,27%	0,05%

Tabela 17 – Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 5% durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,29%	5,36%	0,60%	0,16%
Agentes ABC espertos	4,40%	4,93%	0,43%	0,20%
Agentes-Q	1,65%	2,74%	0,27%	0,31%

Tabela 18 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 10% durante todo o período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,29%	6,17%	0,56%	0,52%
Agentes ABC espertos	4,48%	5,84%	0,40%	0,57%
Agentes-Q	1,56%	3,63%	0,36%	0,66%

Os teste estatísticos demonstraram que houve diferença significativa entre os valores produzidos pelos três algoritmos nos experimentos envolvendo a adição de ruído na seleção das chamadas, o que provocou acréscimos nas perdas dos algoritmos proporcionais ao valor considerado. Em todas as situações consideradas, os agentes-Q obtiveram os menores valores médios de chamadas perdidas, sendo seguidos pelos agentes ABC espertos e estes, por sua vez, sucedidos pelos agentes ABC. Assim, a utilização de um ruído igual a 1% (Tabela 15) produziu um pequeno aumento nos valores de chamadas perdidas por todos os algoritmos implementados em relação aos valores obtidos no primeiro experimento (Tabela 5). Mesmo assim, os agentes ABC e ABC espertos obtiveram perdas no período de teste inferiores as produzidos no período de adaptação. Os valores de desvio-padrão apresentados pelos agentes ABC e ABC espertos foram inferiores àqueles obtidos no primeiro experimento (Tabela 5), enquanto os desvios apresentados pelos agentes-Q foram superiores aos obtidos anteriormente. Isto indica que a sensibilidade do algoritmo proposto a ruídos é superior a de seus competidores. Para um fator de ruído igual a 2% (Tabela 16), todos os algoritmos apresentaram novamente um acréscimo nos valores obtidos, neste caso

superior ao obtido sob um ruído de 1%. Os agentes ABC ainda apresentaram um valor médio de chamadas perdidas inferior ao obtido no período de adaptação, enquanto os agentes ABC espertos apresentaram praticamente o mesmo valor obtido naquele período. Os desvios-padrão produzidos foram inferiores aos obtidos no experimento anterior para todos os agentes. Os resultados produzidos pelos agentes sob um ruído de 5% (Tabela 17) foram ainda superiores aos obtidos para os demais valores de ruído. Neste caso, os agentes ABC e ABC espertos apresentaram aumento nas perdas em relação ao período de adaptação. O aumento produzido nos resultados apresentados pelos agentes-Q foi levemente inferior ao obtido sob um ruído de 2%. Os desvios-padrão apresentados pelos três algoritmos sofreram um acréscimo em relação aos experimentos anteriores com ruídos. As maiores perdas ocorreram para a utilização de um fator ruído igual a 10% (Tabela 18), juntamente com os maiores valores de desvio. Apesar dos acréscimos obtidos sob todas as situações consideradas, foi possível verificar que os agentes foram capazes de se adaptar às novas situações enfrentadas pela rede.

Figura 20 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 1% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica.

Figura 21 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 2% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica.

Figura 22 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 5% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica.

Figura 23 - Desempenho dos algoritmos sob utilização de um fator de ruído igual a 10% na seleção das rotas das chamadas durante todo o período de teste para a rede britânica.

Os gráficos ilustram que em todas as situações envolvendo a utilização de um fator de ruído, os algoritmos produziram um aumento nas perdas. Assim, para um fator de ruído de 1% (Figura 20), as perdas sofreram um pequeno acréscimo e após um determinado período de tempo, os agentes aprenderam e mantiveram suas perdas em valores estáveis. A utilização de um ruído igual a 2% (Figura 21) causou em todos os agentes perdas um pouco superiores as obtidas sob o valor de 1% e também após um período, os algoritmos se adaptaram à nova situação. Porém, para os valores de ruídos iguais a 5% (Figura 22) e 10% (Figura 23), as perdas sofreram acréscimos superiores aos apresentados anteriormente e se mantiveram em crescimento. Este comportamento demonstra a influência do ruído nas informações utilizadas na seleção das rotas sobre o comportamento dos algoritmos. Em todas as situações, os agentes-Q apresentaram valores médios de chamadas perdidas inferiores aos obtidos pelos demais agentes. Entretanto, foi possível verificar neste conjunto de experimentos que o algoritmo proposto é mais sensível a utilização de ruído no roteamento de chamadas que seus competidores, apresentando um acréscimo nas perdas obtidas no período de teste em relação as produzidas no período de adaptação superior ao apresentado pelos demais algoritmos.

8.5.1.5 – Utilização Temporária de um Fator de Ruído na Seleção das Rotas para as Chamadas

Tabela 19 – Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 1% no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,19%	4,57%	0,61%	0,17%
Agentes ABC espertos	4,88%	4,23%	0,53%	0,16%
Agentes-Q	1,58%	1,74%	0,29%	0,04%

Tabela 20 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 2% no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,43%	4,74%	0,60%	0,20%
Agentes ABC espertos	4,69%	4,26%	0,42%	0,14%
Agentes-Q	1,52%	1,80%	0,24%	0,04%

Tabela 21 – Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 5% no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,60%	4,75%	0,68%	0,21%
Agentes ABC espertos	4,69%	4,37%	0,39%	0,16%
Agentes-Q	1,53%	1,94%	0,30%	0,05%

Tabela 22 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 10% no período de teste para a rede britânica.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,10%	4,76%	0,58%	0,23%
Agentes ABC espertos	4,82%	4,62%	0,45%	0,20%
Agentes-Q	1,65%	2,21%	0,26%	0,11%

Os teste estatísticos demonstraram que houve diferença significativa entre os valores produzidos pelos três algoritmos em todos os experimentos onde foi utilizado um fator de ruído temporariamente no período de teste. Os agentes-Q obtiveram as menores perdas em todas as situações, sendo seguidos pelos agentes ABC espertos e estes, por sua vez, pelos agentes ABC. Os valores produzidos pelos agentes sob a utilização temporária de ruído na seleção das rotas percorridas pelas chamadas foram menores que os obtidos sob a adição permanente deste fator. Isto demonstra a maior capacidade de todos os agentes de se adaptarem à este tipo de situação em relação a anterior. Sob um ruído temporário de 1% (Tabela 19), os agentes ABC e ABC espertos apresentaram valores médios de chamadas perdidas no período de teste menores que no período de adaptação. Em relação ao primeiro experimento (Tabela 5), onde não havia ruído nas informações armazenadas nas tabelas, os agentes ABC e agentes-Q apresentaram resultados praticamente iguais, enquanto os agentes ABC espertos obtiveram perdas um pouco maiores que anteriormente. Os valores de desvio-padrão apresentados pelos três algoritmos no período de teste foram próximos daqueles obtidos no primeiro experimento. A utilização temporária de um ruído de 2% (Tabela 20)

provocou um pequeno acréscimo nas perdas em relação ao valor de 1%. Os desvios-padrão apresentados pelos algoritmos foram praticamente iguais aos do experimento anterior, confirmando a alta velocidade de resposta dos agentes a esta condição de operação da rede. Sob um ruído temporário de 5% (Tabela 21), as perdas de todos os algoritmos também foram um pouco maiores que sob os demais valores. Os valores de desvio apresentados se mantiveram, mais uma vez, próximos daqueles obtidos sob os demais valores de ruído. As maiores perdas produzidas pelos três algoritmos ocorreram sob um ruído temporário de 10% (Tabela 22), juntamente com os maiores valores de desvio-padrão em relação ao primeiro experimento realizado (Tabela 5). Mesmo assim, o aumento nas perdas não foi acentuado em nenhum dos algoritmos, reafirmando sua capacidade de adaptação às novas situações da rede, tanto no período de utilização do ruído como no posterior retorno às condições normais de operação.

Figura 24 - Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 1% na seleção das rotas das chamadas no período de teste para a rede britânica.

Figura 25 - Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 2% na seleção das rotas das chamadas no período de teste para a rede britânica.

Figura 26 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 5% na seleção das rotas das chamadas no período de teste para a rede britânica.

Figura 27 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 a 9.000) de um fator de ruído igual a 10% na seleção das rotas das chamadas no período de teste para a rede britânica.

Os gráficos demonstram que a utilização temporária de um fator de ruído na seleção das rotas para as chamadas produziu pequenos acréscimos nos valores médios de chamadas perdidas pelos três algoritmos implementados. Sob um ruído de 1% (Figura 24), os algoritmos apresentaram um pequeno acréscimo nas perdas no período em que este fator é introduzido, sendo menos acentuado no caso dos agentes-Q. Quando a seleção das rotas retornou ao processo normal, todos os agentes apresentaram comportamento semelhante ao do primeiro experimento, se adaptando à nova situação (Figura 10). Assim, os agentes ABC e ABC espertos obtiveram queda nos valores médios de chamadas perdidas, enquanto os agentes-Q mantiveram suas perdas em um determinado nível. A utilização de um ruído de 2% (Figura 25) provocou perdas maiores nos três algoritmos em relação ao valor anterior. Assim também, após o restabelecimento das condições normais da rede, todos os algoritmos mantiveram o comportamento apresentado anteriormente. Sob um ruído de 5% (Figura 26), os agentes produziram um aumento nas perdas obtidas superior aos ocorridos sob os demais valores e sob a ausência de ruído, os agentes ABC e ABC espertos obtiveram decréscimo nas perdas, enquanto os agentes-Q produziram perdas em um determinado patamar. O maior acréscimo nas perdas dos três algoritmos foi obtido sob um fator de ruído igual a 10% (Figura 27). Mesmo assim, após a retirada deste fator todos os agentes se adaptaram à nova situação, apresentando queda nos valores médios de chamadas perdidas. Verificamos, então, que os algoritmos respondem corretamente ao ruído temporário no roteamento das

chamadas, em especial os agentes-Q, que produzem os menores resultados sob todas as situações consideradas.

8.5.2 – Rede de Telecomunicações Norte-Americana

8.5.2.1 – Variações dos Padrões de Tráfego da Rede

Tabela 23 – Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,61%	4,14%	0,84%	0,24%
Agentes ABC espertos	3,66%	3,00%	0,32%	0,12%
Agentes-Q	1,91%	1,57%	0,36%	0,07%

Tabela 24 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,41%	4,19%	0,78%	0,13%
Agentes ABC espertos	3,77%	3,32%	0,40%	0,06%
Agentes-Q	1,84%	1,77%	0,30%	0,09%

Tabela 25 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,68%	4,24%	0,84%	0,23%
Agentes ABC espertos	3,46%	3,00%	0,28%	0,06%
Agentes-Q	1,88%	1,57%	0,31%	0,06%

Tabela 26 - Resultados obtidos sob nível baixo de carga, probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,73%	4,63%	0,88%	0,05%
Agentes ABC espertos	3,61%	3,58%	0,37%	0,18%
Agentes-Q	2,04%	2,05%	0,46%	0,15%

Os testes estatísticos demonstraram que houve diferença significativa entre os resultados produzidos pelos três algoritmos em todos os experimentos envolvendo variações dos padrões de tráfego. As perdas obtidas no período de adaptação pelos agentes ABC e agentes-Q foram levemente superiores as produzidos no mesmo período para os testes com a rede britânica, enquanto as obtidas no período de teste por todos os agentes foram levemente menores que as produzidas anteriormente. Isto demonstra a capacidade de generalização dos três algoritmos, que se adaptaram à nova topologia considerada. Os agentes ABC obtiveram a maior porcentagem de chamadas perdidas sob todos os testes envolvendo variação dos padrões de tráfego, sendo seguidos pelos agentes ABC espertos e estes, pelos agentes-Q. Sob a situação de probabilidades de chamadas inalteradas e lançamento contínuo de formigas (Tabela 23), no período de adaptação o desvio-padrão apresentado pelos agentes ABC foi

superior aos obtidos pelos demais agentes. No período de teste, os desvios sofreram um decréscimo, porém o valor apresentado pelos agentes ABC se manteve superior aos demais. A interrupção do lançamento de formigas no período de teste sob probabilidades de chamadas inalteradas (Tabela 24) provocou um pequeno acréscimo nos valores médios percentuais de chamadas perdidas pelos três algoritmos. Este comportamento evidencia a influência dos agentes sobre o bom desempenho da rede. O valor de desvio-padrão obtido pelo algoritmo proposto no período de teste foi maior que o apresentado no primeiro experimento, enquanto o dos demais foi menor, confirmando a maior necessidade da presença dos agentes-Q para um roteamento eficiente das chamadas em relação aos demais agentes. A alteração das probabilidades de chamadas utilizadas nos períodos de adaptação e teste (Tabela 25) causou perdas um pouco maiores que no primeiro teste (Tabela 23) somente no caso dos agentes ABC. Os agentes ABC espertos e agentes-Q não obtiveram acréscimo nos resultados, demonstrando que eles se adaptam melhor à este tipo de mudança que seu competidor. Os desvios-padrão apresentados no período de teste são semelhantes aos obtidos no primeiro experimento pelos algoritmos ABC e Agentes-Q. O valor de desvio obtido pelos agentes ABC espertos sofreu um pequeno decréscimo em relação ao primeiro teste, reafirmando a capacidade destes algoritmos de acompanhar mudanças nos padrões de tráfego da rede. A variação das probabilidades de chamadas combinada com a interrupção do lançamento de agentes (Tabela 26) produziu um acréscimo nos valores percentuais médios de chamadas perdidas obtidos pelos três algoritmos maior que nos demais experimentos. Ainda assim, as perdas apresentadas pelo algoritmo proposto foram menores que aquelas produzidas pelos seus competidores. Este comportamento demonstra simultaneamente a maior capacidade de adaptação do algoritmo proposto à variação dos padrões de tráfego considerada e a importância dos agentes-Q para o constante balanceamento de carga sobre a rede. As perdas obtidas pelos algoritmos ABC e ABC com agentes espertos evidenciam também a influência dos agentes sobre o desempenho da rede e sua capacidade levemente inferior de acompanhar a variação dos padrões de tráfego realizada.

Figura 28 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.

Figura 29 – Desempenho dos três algoritmos sob probabilidades de chamadas inalteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.

Figura 30 – Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e lançamento contínuo de agentes no período de teste para a rede norte-americana.

Figura 31 – Desempenho dos três algoritmos sob probabilidades de chamadas alteradas e interrupção do lançamento de agentes no período de teste para a rede norte-americana.

Os gráficos esboçam que o comportamento dos três algoritmos implementados sob variações dos padrões de tráfego da rede norte-americana se manteve similar àquele apresentado nos experimentos envolvendo a rede britânica. Entretanto, alguns aspectos dos algoritmos se tornaram mais acentuados nos casos relacionados com a segunda rede. No primeiro experimento (Figura 28), os agentes ABC obtiveram inicialmente perdas superiores ao teste envolvendo a rede britânica (Figura 10), que depois sofreram um acréscimo. Após

um determinado período de tempo, as perdas dos agentes ABC decresceram de forma mais acentuada que na primeira rede utilizada. Os agentes ABC espertos também produziram inicialmente valores médios de chamadas perdidas maiores que sob a rede britânica, as quais sofreram posteriormente um acréscimo, que foi inferior ao apresentado no teste envolvendo a rede anterior. Como os agentes ABC, os agentes espertos apresentaram posteriormente um declínio maior nos valores médios de chamadas perdidas em relação ao experimento com a primeira rede considerada. As perdas produzidas pelos agentes-Q também partiram de um valor superior ao obtido na primeira rede considerada, e apresentaram uma queda acentuada a seguir, aprendendo rapidamente uma política de roteamento mais eficiente que os demais. Sob a interrupção do lançamento de formigas no período de teste com probabilidades de chamadas inalteradas (Figura 29), os agentes-Q e os agentes ABC espertos apresentaram um acréscimo nos valores médios percentuais de chamadas perdidas, enquanto os agentes ABC não obtiveram o decréscimo evidenciado no primeiro experimento. A variação das probabilidades de chamadas no período de teste (Figura 30) produziu praticamente os mesmos resultados apresentados no primeiro experimento pelos três algoritmos considerados, confirmando sua capacidade de adaptação à mudança deste parâmetro. Entretanto, a variação das probabilidades de chamadas aliada à interrupção do lançamento de agentes (Figura 31) causaram um pequeno acréscimo nos valores de chamadas perdidas apresentados pelos agentes ABC após um determinado período de tempo. Os agentes-Q e agentes ABC espertos obtiveram maior acréscimo nos seus resultados, evidenciando a influência dos agentes sobre o seu desempenho.

8.5.2.2 – Variações do Nível de Carga da Rede

Tabela 27 – Resultados obtidos sob alteração da carga da rede para o nível médio durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,82%	8,76%	0,70%	1,28%
Agentes ABC espertos	3,63%	7,73%	0,27%	1,52%
Agentes-Q	2,16%	6,76%	0,47%	1,92%

Tabela 28 – Resultados obtidos sob alteração da carga da rede para o nível alto durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	6,67%	14,56%	1,12%	2,71%
Agentes ABC espertos	4,37%	13,72%	0,65%	3,07%
Agentes-Q	2,42%	13,98%	0,68%	4,17%

Tabela 29 – Resultados obtidos sob alteração temporária da carga da rede para o nível médio no período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		esvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	5,77%	5,54%	0,82%	0,50%
Agentes ABC espertos	3,52%	4,31%	0,37%	0,38%
Agentes-Q	1,84%	2,91%	0,38%	0,33%

Tabela 30 – Resultados obtidos sob alteração temporária da carga da rede para o nível alto no período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		esvio-Padrão	
	Adapt	Teste	Adapt.	Teste
gentes ABC	4,89%	7,90%	0,48%	0,83%
Agentes ABC espertos	3,08%	6,89%	0,16%	0,77%
Agentes-Q	1,66%	5,92%	0,22%	0,83%

Os testes estatísticos comprovaram que, em geral, houve diferença significativa entre os resultados obtidos pelos três algoritmos implementados sob as variações do nível de carga da rede consideradas. Estas alterações provocaram um aumento significativo nas perdas para todos os algoritmos analisados, demonstrando a influência deste parâmetro sobre o desempenho da rede. A mudança da carga para o nível médio durante todo o período de teste (Tabela 27) produziu um aumento elevado de chamadas perdidas pelos três algoritmos em relação ao período de adaptação, principalmente para os agentes-Q. Os desvios-padrão obtidos pelos agentes também sofreram um acréscimo, mais acentuado no caso dos agentes-Q, evidenciando a maior sensibilidade do algoritmo proposto à variação de carga da rede. Os testes estatísticos demonstraram que os valores obtidos pelos agentes ABC espertos e agentes-Q não foram significativamente diferentes. A alteração da carga da rede para o nível alto (Tabela 28) provocou aumentos nas perdas produzidas pelos três algoritmos superiores ao experimento anterior, fenômeno que também se evidencia nos valores de desvio-padrão obtidos. Mais uma vez, os agentes-Q apresentaram o maior desvio obtido, confirmando o grande efeito deste tipo de mudança para o roteamento da rede. Neste caso, os testes estatísticos demonstraram que os resultados produzidos pelos três algoritmos no período de teste não apresentaram diferença significativa. Sob a situação de mudança temporária da carga da rede para o nível médio (Tabela 29), todos os algoritmos apresentaram novamente aumento nos valores de chamadas perdidas, porém estes foram inferiores àqueles obtidos sob a variação permanente da carga. Os desvios-padrão obtidos também foram muito inferiores aos produzidos no caso de mudança permanente da carga, porém ainda se distanciaram um pouco dos valores obtidos sob variações dos padrões de tráfego da rede. Este comportamento também pode ser verificado na situação de alteração temporária da carga para o nível alto (Tabela 30), quando os resultados obtidos pelos agentes, tanto com relação aos valores médios de chamadas perdidas como aos desvios-padrão, foram inferiores àqueles observados sob variação permanente da carga para este nível.

Figura 32 – Desempenho dos três algoritmos sob o nível médio de carga da rede durante todo o período de teste para a rede norte-americana.

Figura 33 – Desempenho dos três algoritmos sob o nível alto de carga da rede durante todo o período de teste para a rede norte-americana.

Figura 34 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível médio (do passo 8.000 ao 9.000) durante o período de teste para a rede norte-americana.

Figura 35 – Desempenho dos três algoritmos sob variação temporária da carga da rede para o nível alto (do passo 8.000 ao 9.000) durante o período de teste para a rede norte-americana.

Os gráficos ilustram que a mudança permanente da carga para o nível médio (Figura 32) provocou um acréscimo contínuo nos valores médios percentuais de chamadas perdidas produzidos por todos os algoritmos considerados. Para este modelo de rede e para os blocos de chamadas gerados, os algoritmos não apresentaram níveis decrescentes de valores após um determinado período de tempo, como aconteceu na rede britânica. Este comportamento pode estar relacionado com o tamanho desta rede, superior à anterior, que aumentam a complexidade do problema de roteamento. Além disso, os nós desta rede encontram-se

conectados a um menor número de vizinhos que a rede anterior, o que limita o número de rotas a serem usadas por eles. Sob alteração permanente da carga para o nível alto (Figura 33), o mesmo comportamento foi observado, porém as perdas de chamadas foram superiores ao experimento anterior, sendo proporcional ao nível de carga considerado. Sob a situação de mudança temporária da carga para o nível médio (Figura 34), todos os algoritmos sofreram inicialmente um acréscimo nas perdas obtidas. Após o restabelecimento do nível de carga da rede para o nível baixo, os agentes aprenderam e apresentaram decréscimo nos valores de chamadas perdidas, evidenciando sua capacidade de se adaptar à nova situação da rede. A alteração temporária da carga para o nível alto (Figura 35) produziu um grande aumento na porcentagem média de chamadas perdidas. Mais uma vez, o posterior retorno da carga da rede para o nível baixo produziu valores decrescentes de chamadas perdidas, confirmando a capacidade de adaptação dos agentes. Nestes dois últimos testes, os agentes-Q obtiveram menores perdas que os demais algoritmos, indicando sua capacidade superior de adaptação às novas condições da rede.

As perdas produzidas pelos três algoritmos sob variações permanentes do nível de carga da rede norte-americana foram superiores às obtidas para a rede britânica. Mais uma vez, isto pode estar vinculado a maior complexidade do problema devido ao maior número de nós da segunda rede, como também ao menor número de vizinhos dos mesmos. Sob variação temporária do nível de carga da rede, porém, todos os algoritmos obtiveram perdas menores que as produzidas na rede britânica. Isto demonstra a maior velocidade de resposta dos agentes à mudanças momentâneas nas condições da rede, independente de seu tamanho, que constitui uma característica de extrema importância do sistema de roteamento, uma vez que variações temporárias dos parâmetros da rede são mais comuns nas redes de telecomunicações reais que as permanentes.

8.5.2.3 – Variações na Topologia da Rede

Tabela 31 – Resultados obtidos sob desconexão do nó 60 durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,36%	7,18%	0,75%	1,55%
Agentes ABC espertos	3,40%	5,92%	0,38%	1,59%
Agentes-Q	1,70%	2,93%	0,34%	0,56%

Tabela 32 - Resultados obtidos sob desconexão temporária do nó 60 durante o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	5,98%	5,31%	0,87%	0,12%
Agentes ABC espertos	3,51%	3,86%	0,36%	0,10%
Agentes-Q	2,02%	2,54%	0,41%	0,17%

Os testes estatísticos atestaram que a desconexão permanente do nó 60 (Tabela 31) influenciou de forma mais intensa o comportamento dos agentes ABC e ABC espertos em relação aos agentes-Q, causando maiores perdas para aqueles algoritmos. Todos os algoritmos apresentaram acréscimo nos valores de desvio-padrão obtidos no período de teste, mais acentuado no caso dos agentes ABC espertos, demonstrando sua sensibilidade a este tipo de mudança na rede. A desconexão temporária do nó 60 (Tabela 32) produziu menores perdas em todos os algoritmos que sua desconexão permanente. Neste caso, os agentes ABC espertos e agentes-Q obtiveram um pequeno acréscimo nos valores médios de chamadas perdidas em relação aos resultados produzidos no período de adaptação, enquanto os agentes ABC sofreram um decréscimo nestes valores. Os desvios-padrão apresentados pelos três algoritmos no período de teste foram menores que os produzidos no teste anterior, demonstrando a capacidade de adaptação dos agentes à esta situação de operação da rede. Os agentes-Q obtiveram o maior valor de desvio, indicando sua sensibilidade a retirada do nó selecionado.

Figura 36 – Desempenho dos três algoritmos sob desconexão permanente do nó 60 no período de teste para a rede norte-americana.

Figura 37 – Desempenho dos três algoritmos sob desconexão temporária do nó 60 (do passo 8.000 ao 9.000) no período de teste para a rede norte-americana.

Os gráficos esboçam que a desconexão permanente do nó 60 (Figura 36) produziu um grande acréscimo nos valores médios de chamadas perdidas obtidos pelos agentes ABC e ABC espertos. Os agentes-Q, por sua vez, apresentaram um menor aumento nas perdas que seus competidores, demonstrando sua capacidade de adaptação superior a este tipo de mudança na topologia em relação aos demais algoritmos. Como no experimento envolvendo a rede britânica, este comportamento dos algoritmos pode estar vinculado com um dos

seguintes fatores ou com uma combinação de ambos: a capacidade de adaptação dos algoritmos e a influência do nó selecionado para o roteamento da rede. Quando o nó 60 foi restabelecido à rede (Figura 37), os agentes ABC e ABC espertos praticamente mantiveram seus valores médios de chamadas perdidas em um determinado nível, com algumas oscilações. Os agentes-Q, por sua vez, apresentaram queda nas perdas, com oscilações inferiores às obtidas pelos demais agentes. Este comportamento evidencia a adaptação mais eficiente do algoritmo proposto à situação considerada em relação aos seus competidores.

Os valores médios de chamadas perdidas produzidos pelos três algoritmos sob a desconexão permanente e temporária de um nó da rede norte-americana foram menores que aqueles obtidos para a rede britânica. Este comportamento evidencia a menor influência deste parâmetro sobre o desempenho do sistema de roteamento em relação ao nível de carga da rede, onde os resultados produzidos pelos agentes foram maiores na rede norte-americana que na rede britânica. Assim também, verificamos a capacidade de adaptação dos algoritmos à esta nova situação da rede, independente do número de nós que a compõem.

8.5.2.4 – Utilização Permanente de Um Fator de Ruído na Seleção das Rotas para as Chamadas

Tabela 33 – Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 1% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,39%	6,23%	0,40%	0,10%
Agentes ABC espertos	4,33%	4,49%	0,25%	0,22%
Agentes-Q	2,11%	2,57%	0,45%	0,34%

Tabela 34 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 2% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,04%	6,52%	0,46%	0,33%
Agentes ABC espertos	3,58%	4,22%	0,31%	0,43%
Agentes-Q	2,02%	3,00%	0,46%	0,54%

Tabela 35 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 5% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	7,17%	8,28%	0,50%	0,66%
Agentes ABC espertos	3,65%	5,51%	0,33%	0,97%
Agentes-Q	1,89%	4,41%	0,37%	1,09%

Tabela 36 - Resultados obtidos sob utilização de um fator de ruído na seleção das rotas das chamadas igual a 10% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,27%	9,64%	0,48%	1,57%
Agentes ABC espertos	5,01%	8,65%	0,42%	1,71%
Agentes-Q	2,12%	6,66%	0,42%	2,13%

Os testes estatísticos demonstraram que os valores médios de chamadas perdidas pelos três algoritmos apresentaram diferença significativa em todos os experimentos envolvendo a utilização permanente de um fator de ruído. As perdas produzidas pelos agentes-Q foram menores que as dos demais algoritmos em todas as situações consideradas, enquanto os agentes ABC espertos obtiveram resultados menores que os agentes ABC. No caso de um ruído constante igual a 1% (Tabela 33), os algoritmos apresentaram um acréscimo nos valores obtidos em relação ao primeiro experimento (Tabela 23), onde não foi utilizado este fator. Como no caso da primeira rede considerada, os agentes ABC não obtiveram aumento nas perdas no período de teste em relação ao período de adaptação. Os agentes-Q e ABC espertos, por sua vez, produziram um acréscimo nos valores médios de chamadas perdidas obtidos no período de adaptação. Este comportamento salienta a maior sensibilidade destes algoritmos a este tipo de mudança nas condições da rede em relação ao algoritmo ABC. Da mesma forma, o valor de desvio-padrão produzido pelo algoritmo ABC foi menor que o obtido no primeiro experimento (Tabela 23), enquanto os valores apresentados pelos agentes ABC espertos e agentes-Q foram superiores aos obtidos anteriormente. Sob a utilização de um ruído igual a 2% (Tabela 34), todos os algoritmos apresentaram acréscimo nas perdas em relação ao período de adaptação, sendo maiores que os valores obtidos no teste anterior para os agentes ABC e agentes-Q. Assim também, os desvios-padrão sofreram um aumento em relação ao valor de 1% para os três tipos de agentes. O comportamento de perdas crescentes se evidencia para os demais fatores de ruído utilizados. Deste modo, para um fator de 5% (Tabela 35), os valores médios de chamadas perdidas são superiores aos obtidos anteriormente, assim como os desvios-padrão. As maiores perdas foram produzidas sob um ruído de 10% (Tabela 36) e neste caso os resultados apresentados pelos agentes ABC não foram significativamente diferentes daqueles obtidos pelos agentes ABC espertos, de acordo com os testes estatísticos realizados. Os agentes-Q, mais uma vez, produziram as menores perdas. Os valores de desvio-padrão obtidos pelos três algoritmos acompanharam a tendência das chamadas perdidas em média, apresentando grande acréscimo em relação aos obtidos anteriormente.

Figura 38 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 1% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.

Figura 39 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 2% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.

Figura 40 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 5% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.

Figura 41 – Desempenho dos algoritmos sob utilização de um fator de ruído igual a 10% na seleção das rotas das chamadas durante todo o período de teste para a rede norte-americana.

Os gráficos demonstram que a utilização de um fator de ruído na seleção das chamadas provocou aumentos nas perdas dos três algoritmos, que acompanham o crescimento do ruído. Assim, sob um fator de ruído de 1% (Figura 38), as perdas apresentadas pelo algoritmo ABC sofreram um pequeno aumento, porém se mantiveram estáveis em um determinado patamar. Os agentes ABC espertos e agentes-Q, por sua vez,

produziram valores médios de chamadas perdidas crescentes, sendo mais acentuado no caso do algoritmo proposto, demonstrando a maior influência deste tipo de variação sobre o desempenho dos agentes-Q em relação aos demais. Para um ruído igual a 2% (Figura 39), todos os agentes obtiveram perdas crescentes. O mesmo comportamento pode ser verificado para os demais valores de ruído considerados. Assim, os resultados produzidos sob um ruído de 5% (Figura 40) sofreram acréscimo contínuo e foram superiores aos obtidos sob os valores utilizados anteriormente. As maiores perdas foram apresentadas para o valor de ruído igual a 10% (Figura 41), onde os resultados mantiveram a tendência crescente, neste caso de forma mais acentuada. O comportamento dos algoritmos demonstra a influência da variação das informações utilizadas na seleção das rotas, confirmando a necessidade da presença de dados coerentes com a situação corrente da rede nas tabelas dos nós, para que haja o roteamento adequado das chamadas. Assim, verificamos que os algoritmos não apresentam robustez a ruídos a partir de um determinado valor considerado.

8.5.2.5 – Utilização Temporária de Um Fator de Ruído na Seleção das Rotas para as Chamadas

Tabela 37 – Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 1% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,18%	5,63%	0,52%	0,24%
Agentes ABC espertos	3,54%	3,14%	0,30%	0,13%
Agentes-Q	2,09%	1,87%	0,44%	0,10%

Tabela 38 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 2% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,22%	5,56%	0,46%	0,26%
Agentes ABC espertos	3,76%	3,37%	0,45%	0,16%
Agentes-Q	1,94%	1,87%	0,40%	0,10%

Tabela 39 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 5% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,25%	6,07%	0,44%	0,30%
Agentes ABC espertos	5,15%	5,06%	0,51%	0,20%
Agentes-Q	1,99%	2,31%	0,45%	0,18%

Tabela 40 - Resultados obtidos sob utilização temporária de um fator de ruído na seleção das rotas das chamadas igual a 10% durante todo o período de teste para a rede norte-americana.

Algoritmos	Média de Chamadas Perdidas		Desvio-Padrão	
	Adapt	Teste	Adapt.	Teste
Agentes ABC	6,15%	6,50%	0,42%	0,39%
Agentes ABC espertos	5,00%	5,43%	0,40%	0,29%
Agentes-Q	1,98%	2,88%	0,36%	0,29%

Os testes estatísticos demonstraram que os valores médios de chamadas perdidas pelos três algoritmos apresentaram diferença significativa em todos os experimentos envolvendo a utilização de um fator de ruído. Os agentes-Q obtiveram as menores perdas em todas as situações, sendo seguidos pelos agentes ABC espertos e estes, pelos agentes ABC. Os resultados produzidos pelos três algoritmos sob a adição temporária de ruído na seleção das rotas para as chamadas foram menores que os obtidos nos casos de utilização permanente deste fator. Para um ruído temporário de 1% (Tabela 37), todos os algoritmos apresentaram um pequeno acréscimo nos valores médios de chamadas perdidas em relação ao primeiro experimento (Tabela 23). Mesmo assim, os valores produzidos no período de teste foram menores que aqueles obtidos no período de adaptação, demonstrando a capacidade dos agentes de aprenderem a lidar com a nova situação da rede. Os desvios-padrão apresentados pelos três algoritmos foram semelhantes aos obtidos no primeiro experimento (Tabela 23), confirmando sua capacidade de resposta à variação da rede. Sob um ruído temporário de 2% (Tabela 38), as perdas dos agentes ABC espertos foram um pouco superiores às obtidas no caso anterior, juntamente com os valores de desvio-padrão. Os agentes-Q, por sua vez, produziram os mesmos resultados que antes e os agentes ABC apresentaram queda no valor médio de chamadas perdidas. Os desvios obtidos pelos três algoritmos foram próximos dos produzidos no experimento anterior. A adição de um ruído temporário de 5% (Tabela 39) provocou maiores perdas em todos os algoritmos, juntamente com maiores desvios-padrão. Os agentes ABC e ABC espertos mantiveram obtendo perdas no período de teste menores que no período de adaptação, enquanto os agentes-Q apresentaram um aumento nos valores médios de chamadas perdidas produzidos. Sob a adição temporária de um ruído de 10% (Tabela 40), todos os algoritmos produziram valores médios de chamadas perdidas no período de teste maiores que os obtidos no período de adaptação. As perdas foram superiores as produzidas sob os demais valores de ruído. Os desvios produzidos pelos três algoritmos também sofreram acréscimo, mais acentuado no caso dos agentes ABC. Apesar dos aumentos nas perdas, todos os agentes foram capazes de se adaptar às novas condições de operação da rede.

Figura 42 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 1% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.

Figura 43 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 2% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.

Figura 44 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 5% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.

Figura 45 – Desempenho dos algoritmos sob utilização temporária (do passo 8.000 ao 9.000) de um fator de ruído igual a 10% na seleção das rotas das chamadas durante o período de teste para a rede norte-americana.

Os gráficos demonstram que todos os algoritmos obtiveram acréscimos nos valores médios de chamadas perdidas sob a adição temporária de um fator de ruído na seleção das rotas a serem percorridas pelas chamadas. Entretanto, estes aumentos foram inferiores aos produzidos sob a utilização permanente deste fator. Para um ruído temporário de 1% (Figura

42), os três agentes obtiveram um pequeno aumento nas perdas. Quando não havia mais a presença deste fator, todos os algoritmos mantiveram o desempenho apresentado no primeiro experimento (Figura 28), obtendo uma leve queda nas perdas. A adição temporária de um ruído de 2% (Figura 43) provocou um aumento nas perdas dos três algoritmos superior ao obtido sob um ruído de 1%. Após a retirada do fator de ruído, todos os agentes se adaptaram à nova situação e apresentaram decréscimo nos valores médios de chamadas perdidas, com menos oscilações no caso dos agentes-Q. Sob um ruído de 5% (Figura 44), o acréscimo produzido nas perdas dos agentes foi superior aos apresentados sob os demais valores deste fator. A queda nas perdas, ocorrida após o restabelecimento das condições normais de operação da rede, foi mais acentuada neste caso que anteriormente. Os maiores aumentos nos resultados produzidos pelos três algoritmos ocorreram sob a utilização temporária de um fator de ruído de 10% (Figura 45). Mais uma vez, após a retirada deste fator, os agentes aprenderam e obtiveram decréscimo acentuado nos valores médios de chamadas perdidas. Os agentes-Q produziram os menores resultados em todas as situações, demonstrando capacidade superior de adaptação à nova situação da rede em relação aos seus competidores.

8.6 – Discussão dos Resultados

O algoritmo Agentes-Q confirmou a capacidade dos algoritmos de aprendizagem por reforço para solucionar o problema de roteamento em redes de telecomunicações. Este é um problema real de otimização combinatorial, que possui características dinâmicas, além de múltiplos objetivos e restrições. A união de características pertinentes a dois algoritmos de roteamento baseados em aprendizagem-Q aplicados a redes de comutação de pacotes e a um algoritmo baseado no comportamento coletivo de formigas aplicado a redes de comutação de circuitos, aliada à adição de alguns mecanismos de otimização, produziu um sistema distribuído, adaptativo e robusto face a diversas mudanças nas condições de operação da rede.

O sistema proposto aprende de forma contínua e incremental, através da análise *online* das rotas selecionadas pelos agentes. As decisões de roteamento são distribuídas e baseadas somente em informação local sobre as condições da rede. A interação entre os agentes através de um mecanismo de comunicação indireta torna possível a descoberta dos melhores caminhos, minimizando o número de nós visitados e a possibilidade de congestionamento dos mesmos. A utilização de uma política de decisão probabilística permite que rotas alternativas também sejam encontradas, otimizando o desempenho da rede sob condições de bloqueio dos melhores caminhos. Esta estratégia foi baseada no

comportamento coletivo de formigas e constitui um avanço em relação aos algoritmos de roteamento baseados somente na aprendizagem-Q, os quais utilizam uma política de decisão determinística.

O algoritmo Agentes-Q constitui o primeiro sistema a realizar experimentos sob variações nos três principais parâmetros de operação da rede: padrões de tráfego, níveis de carga e topologia. Além disto, foram realizados testes sob a utilização de ruído nas informações usadas na seleção das rotas percorridas pelas chamadas. Estes últimos testes simulam a falha de um roteador relacionada com a corrupção de seu estado interno, que o leva a tomar decisões incorretas em alguns instantes de tempo e provavelmente provoca o congestionamento de alguns nós da rede. Cada um destes parâmetros da rede foi individualmente modificado, a fim de estudar a sua influência sobre o desempenho dos algoritmos considerados. Todas os experimentos realizados procuraram refletir situações reais enfrentadas pelas redes. Verificou-se que as variações do nível de carga da rede produziram os maiores efeitos sobre o volume de chamadas perdidas pelos algoritmos.

Os resultados apresentados e os testes estatísticos realizados demonstraram o desempenho superior do algoritmo proposto em relação aos seus competidores em praticamente todas as situações consideradas. Os agentes-Q foram capazes de aprender a lidar com as diversas condições da rede, demonstrando robustez e flexibilidade às características dinâmicas do ambiente. Este comportamento está relacionado com a combinação dos melhores aspectos dos algoritmos nos quais nosso sistema foi baseado, descritos acima, aliada aos seguintes mecanismos de exploração do ambiente que foram adicionados ao nosso algoritmo:

- ♦ A seleção das rotas a serem percorridas pelas chamadas através da combinação da informação armazenada nas tabelas sobre a qualidade das rotas com a informação sobre a situação atual dos nós visitados também permite que os agentes detectem rapidamente as mudanças ocorridas na rede e reajam corretamente à elas.
- ♦ A utilização de um sinal de reforço negativo na exploração realizada na direção do movimento do agente, quando um nó se encontra processando um número de chamadas superior a 60% de sua capacidade máxima de processamento, reduz a probabilidade de escolha deste nó e favorece nós vizinhos que se encontram menos carregados. Este mecanismo procura atrasar ou até mesmo evitar o congestionamento dos melhores caminhos, balanceando a carga através da rede.

É importante ressaltar que nos casos em que houve variação temporária dos parâmetros da rede, quando as condições do ambiente foram retornadas ao seu estado normal, nenhum dos algoritmos analisados produziu os melhores caminhos obtidos antes da variação. Esta constitui uma limitação destes sistemas, evidenciada anteriormente pelo algoritmo Roteamento-Q, baseado na aprendizagem-Q.

Outra questão de extrema relevância reside no fato de que o único experimento em que o algoritmo proposto produz resultados semelhantes aos seus competidores, que envolve a variação permanente da carga da rede norte-americana, constitui uma situação extrema enfrentada pelas redes de telecomunicações. Em geral, as variações deste parâmetro são temporárias, como também foi simulado. Assim também, os valores estabelecidos para os níveis de carga baixo, médio e alto foram selecionados de forma intuitiva, uma vez que dados reais de operação das redes constituem informação confidencial das empresas e não são divulgados. Da mesma forma, os valores de ruído utilizados foram determinados

Finalmente, devemos enfatizar também que os resultados obtidos pelos três algoritmos implementados e apresentados neste trabalho constituem valores típicos e não os melhores valores por eles produzidos.

8.7 – Conclusão

Neste capítulo foram descritas as características das duas redes de telecomunicações utilizadas nos experimentos realizados, juntamente com as condições de operação da rede consideradas. A realização destes experimentos teve como objetivo verificar o comportamento do algoritmo proposto sob condições de variação dos três parâmetros principais de operação da rede: padrões de tráfego, nível de carga e topologia. De forma simultânea, procurou-se comparar o comportamento do algoritmo desenvolvido com o de dois algoritmos baseados em formigas, os quais têm demonstrado desempenho superior a métodos clássicos e recentes de roteamento aplicados às redes de comutação de circuitos.

Os experimentos demonstraram que o algoritmo de roteamento Agentes-Q obteve melhores resultados que os demais sob todas as situações simuladas. Isto se deve ao fato de que, além de apresentar características semelhantes aos dois algoritmos implementados, o algoritmo proposto utiliza aspectos pertinentes a aprendizagem-Q e a aprendizagem por reforço dual, que o permitem aprender uma política de roteamento em um período de tempo inferior aos demais. Além disso, a presença de alguns mecanismos de exploração extras possibilitou que o algoritmo Agentes-Q se adaptasse de forma mais eficiente e eficaz às diversas variações dos parâmetros da rede realizadas.

É importante salientar que os demais algoritmos considerados foram submetidos somente aos quatro primeiros testes realizados neste trabalho. Deste modo, não havia conhecimento sobre seu comportamento sob condições de mudança do nível de carga e topologia da rede, que foi possível através deste projeto de pesquisa.

Logo, verifica-se que o algoritmo de roteamento desenvolvido apresenta o potencial para se adaptar às principais situações enfrentadas atualmente pelas redes de comutação de circuitos.

No próximo capítulo, serão apresentadas as conclusões obtidas ao longo das atividades realizadas neste projeto de pesquisa.

9 – Conclusões

As redes de telecomunicações devem transmitir informação entre seus usuários de forma rápida, segura e confiável, proporcionando serviços de baixo custo e alta qualidade. Para que isto ocorra, a utilização dos recursos da rede envolvidos na transmissão de dados deve ser minimizada, a fim de que o número de usuários atendidos seja maximizado. Desta forma, a seleção das rotas a serem percorridas pela informação em direção ao destino desejado e sua condução correta a este local constitui um dos principais problemas enfrentados pelas redes de telecomunicações. Esta tarefa é responsabilidade do sistema de roteamento da rede, que deve ser eficiente e eficaz o suficiente para selecionar os melhores caminhos ao longo da rede sob quaisquer condições de operação da mesma. O avanço das tecnologias utilizadas pelas redes proporcionou grande aumento de sua capacidade de transmissão e redução dos custos. Esta mudança vem levando um número cada vez maior de usuários a utilizar serviços de telecomunicação. Deste modo, as redes atuais são concebidas como ambientes altamente dinâmicos e complexos, onde a demanda por serviços diversificados e diferenciados é crescente.

Diante da necessidade de novos sistemas de roteamento, que sejam capazes de lidar corretamente com o novo perfil das redes de telecomunicações, este projeto de pesquisa realizou o desenvolvimento de um algoritmo de roteamento dinâmico e distribuído, resultante da integração de propriedades de três técnicas de aprendizagem e da adição de alguns mecanismos de otimização. As estratégias utilizadas foram a aprendizagem-Q, aprendizagem por reforço dual e aprendizagem baseada no comportamento coletivo de formigas. Deste modo, o algoritmo proposto é constituído por um conjunto de agentes autônomos, que aprendem a lidar com as condições da rede de forma contínua e *online*, baseados somente na experiência sobre as ações realizadas. Os agentes realizam suas ações de forma distribuída e assíncrona, cooperando entre si na busca das melhores rotas sobre a rede através de um processo de comunicação indireta, que permite adquirir informação sobre a qualidade dos caminhos percorridos e o efeito destes sobre as características de operação da rede. Os mecanismos extras de otimização inseridos ao sistema estão relacionados com a seleção das rotas a serem percorridas pelas chamadas e com a atualização das informações utilizadas nesta seleção. Estes mecanismos procuram combinar a informação armazenada nas

tabelas de roteamento dos nós com a situação corrente dos candidatos a serem percorridos pelas chamadas, como também diminuir a chance de selecionar um nó que constitui o menor caminho em direção ao destino desejado, mas que em um determinado momento encontra-se altamente carregado. Esta estratégia procura minimizar a diferença entre os dados armazenados nas tabelas sobre a qualidade das rotas e as condições atuais da rede, que variam continuamente ao longo do tempo, muitas vezes de forma inesperada. Assim, procuramos adequar a velocidade e capacidade de adaptação do algoritmo às diversas situações enfrentadas pelas redes de telecomunicações atuais.

O algoritmo proposto proporcionou algumas melhorias nas abordagens desenvolvidas anteriormente devido em parte à integração das melhores características de três abordagens de aprendizagem e principalmente, à utilização dos dois mecanismos extras de otimização descritos acima. Ao levar em consideração na seleção das rotas a informação utilizada pelos algoritmos competidores combinada com dados atuais sobre as características da rede, os agentes-Q se ajustam melhor às condições dinâmicas e estocásticas da rede em relação aos demais. Assim também, o envio de um sinal de reforço negativo para os nós que se encontram altamente carregados proporciona melhor balanceamento da carga sobre a rede que aquele realizado pelos demais algoritmos considerados, que utilizam sempre um sinal de reforço positivo para os nós selecionados, aumentando suas chances de serem escolhidos independente do nível de carga que eles se encontram processando. Desta forma, os agentes-Q são capazes de distribuir de forma mais eficiente a carga sobre os diversos nós, retardando ou até mesmo reduzindo o período de congestionamento dos melhores caminhos da rede.

Os experimentos realizados compararam o desempenho do algoritmo proposto com o de dois algoritmos recentemente desenvolvidos e aplicados com sucesso ao problema de roteamento, obtendo melhores resultados que algoritmos utilizados por empresas operadoras das redes de telecomunicações. Os experimentos envolveram variações dos principais parâmetros da rede: padrões de tráfego, nível de carga e topologia. Além disto, foi utilizado um fator de ruído na seleção das chamadas, simulando a corrupção do estado interno do roteador. Os testes realizados foram selecionados de forma a considerar as principais situações enfrentadas pelas redes de telecomunicações atuais. Em todas as condições consideradas, os agentes-Q se comportaram de forma mais eficiente e eficaz que os demais, realizando balanço mais adequado entre dinamismo e estabilidade. O algoritmo desenvolvido foi aplicado a duas redes de comutação de circuitos, uma rede britânica e uma norte-americana, para verificar sua capacidade de adaptação à diferentes topologias e tamanhos de redes. O comportamento dos agentes-Q se manteve em ambas as redes e eles demonstraram ser capazes de acompanhar rápida e corretamente as diversas variações realizadas nos

parâmetros de operação das redes utilizadas. Entretanto, as perdas produzidas pelo algoritmo proposto sofreram um pequeno acréscimo na rede norte-americana em relação a rede britânica, possivelmente devido ao maior número de nós e diferente disposição dos mesmos sobre a rede, fatores que aumentam a complexidade do problema de roteamento.

Somente em um dos testes realizados, aquele que envolve a variação permanente da carga da rede norte-americana para o nível alto, o algoritmo proposto apresentou resultados semelhantes aos demais. Nos demais casos, os agentes-Q obtiveram perdas inferiores aos seus competidores. Foi possível verificar que o algoritmo proposto necessita de alguns ajustes para aprimorar sua robustez a ruídos, uma vez que os resultados produzidos demonstraram a maior influência deste aspecto sobre o seu desempenho em relação aos demais algoritmos considerados. Deste modo, uma questão importante de pesquisa futura será investigar novos mecanismos de otimização que devem ser adicionados aos agentes-Q, para aumentar sua robustez a ruídos e exploração do ambiente em casos de variação permanente do nível de carga. O comportamento do sistema sob estas situações indica que o compromisso entre a exploração do ambiente em busca de possíveis soluções (*exploration*) e a exploração para melhoria de uma solução promissora (*explotation*) necessita ser aprimorado.

É importante mencionar que os experimentos realizados envolveram algumas simplificações das características reais das redes de telecomunicações, como a capacidade de transmissão ilimitada das conexões e a mesma capacidade de processamento de chamadas de todos os nós da rede. Isto está relacionado em parte com a dificuldade de simulação de todos os aspectos dinâmicos das redes, devido às limitações de processamento e memória do computador utilizado, assim como às limitações de tempo. Assim também, a impossibilidade de obter dados reais sobre as condições de operação das redes e principalmente sobre o seu desempenho dificulta a simulação de todos os aspectos reais das redes de telecomunicações. Assim, optamos por manter as características das redes consideradas iguais às utilizadas pelos autores dos demais algoritmos implementados.

Os resultados obtidos pelo algoritmo proposto em uma rede de comutação de circuitos indicam que sua estrutura poderá ser aplicada com sucesso a redes de comutação de pacotes e redes de alta velocidade, que combinam características pertencentes as duas tecnologias anteriores. Nestes casos, pretendemos levar em consideração as diversas limitações de operação das redes, como as diferentes capacidades de processamento dos roteadores e de transmissão das conexões. Além disso, procuraremos analisar o comportamento do algoritmo sob diversas topologias de rede, para verificar sua capacidade de generalização. Para melhorar a capacidade de adaptação do algoritmo proposto, aumentando sua velocidade de

resposta e qualidade das ações realizadas, estudaremos novos mecanismos de otimização que devam ser anexados ao sistema, como também a modificação dos já existentes.

A realização deste projeto de pesquisa permitiu o conhecimento das principais características das redes de telecomunicações, e em particular do roteamento, que constitui um dos mecanismos responsáveis pela segurança, qualidade e confiabilidade dos serviços prestados pelas mesmas. Desta forma, foi possível compreender os aspectos fundamentais do planejamento e gerenciamento de uma rede de telecomunicação.

Dentro deste contexto, o desenvolvimento de um sistema de roteamento em uma rede de telecomunicação constitui um trabalho avançado e de extrema relevância, tanto tecnológica como econômica. A realização de pesquisas nesta área permite o posicionamento do Brasil entre os detentores de conhecimento em um ambiente restrito e intimamente ligado ao desenvolvimento de um país.

10 – Referências Bibliográficas

Akinpelu, J. M. (1984). The overload performance of engineered networks with nonhierarchical and hierarchical routing. *AT&T Bell Laboratories Technical Journal*, vol. 63, n° 7, September.

Appleby, S. & Steward, S. (1994). Mobile software agents for control in telecommunications networks. *British Telecom Technological Journal*, vol. 12, pp. 104-113.

Ash, G. R. (1997). *Dynamic routing in telecommunications networks*. McGraw-Hill.

Ash, G. R. (1995). Dynamic network evolution, with examples from AT&T's evolving dynamic network. *IEEE Communications Magazine*, July.

Barto, A. G. & Singh, S. P. (1990). On the computational economics of reinforcement learning. In D. S. Touretzky, J. Elman, T. J. Sejnowski & G. E. Hinton, (Eds.), *Proceedings of the 1990 Connectionist Models Summer School*. San Mateo, CA: Morgan Kaufmann.

Beckers, R., Deneubourg, J. L. & Goss, S. (1992). Trails and u-turns in the selection of a path by the ant *lasius niger*. In *Journal of Theoretical Biology*, vol. 159, pp. 397-415.

Bellman, R. (1957). *Dynamic programming*. Princeton University Press.

Bellman, R. (1958). On a routing problem. *Quarterly of Applied Mathematics*, vol. 16, n° 1, pp. 87-90.

Berry, D. A. & Fristedt, B. (1985). *Bandit problems: sequential allocation of experiments*. London, UK: Chapman and Hall.

Bertsekas, D. (1995). *Dynamic programming and optimal control*. Athena Scientific.

Bertsekas, D. & Gallager, R. (1992). *Data networks*. Prentice-Hall.

Bonabeau, E., Theraulaz, G., Deneubourg, J-L., Aron, S. & Camazine, S. (1997). Self-organization in social insects. Working Paper 97-04-032, Santa Fe Institute.

Bonabeau, E., Henaux, F., Guérin, S., Snyers, D., Kunts, P. & Theraulaz, G. (1998). Routing in telecommunications networks with "smart" ant-like agents telecommunication applications. In *Proceedings of IATA'98, Second Int. Workshop on Intelligent Agents for Telecommunication Applications*. Lectures Notes in Artificial Intelligence, vol. 1437, Springer Verlag.

Boyan, J. A. & Littman, M. L. (1994). Packet routing in dynamically changing networks: a reinforcement learning approach. *Advance in Neural Information Processing Systems 6*, 671-678. San Mateo, CA: Morgan Kaufmann.

Bullnheimer, B., Hartl, R. F. & Strauss, C. (1998). Applying the ant system to the vehicle routing problem. In Osman, I. H., Vob, S., Martello, S. & Roucairol, C. (Eds.), *Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization*, pp. 109-120. Kluwer Academics.

Chapman, D. & Kaelbling, L. P. (1991). Input generalization in delayed reinforcement learning: an algorithm and performance comparisons. In *Proceedings of the 1991 International Joint Conference on Artificial Intelligence*, pp. 726-731.

Choi, S. P. M. & Yeung, D.-Y., (1996). Predictive Q-routing: A memory-based reinforcement learning approach to adaptive traffic control. In Touretzky, D. S., Mozer, M. C., Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems*, vol.8, pp. 945-951. Cambridge, MA: MIT Press.

Clarke, A B. & Disney, R. L. (1979). Probabilidade e processos estocásticos. Livros Técnicos e Científicos. Rio de Janeiro.

Coloni, A, Dorigo, M. & Maniezzo, V. (1991). Distributed optimization by ant colonies. In *Proceedings of the European Conference on Artificial Life (ECAL 91)*, pp. 134-142. Elsevier.

Colomi, A, Dorigo, M., Maniezzo, V. & Trubian, M. (1994). Ant system for job-shop scheduling. *Belgian Journal of Operations Research, Statistics and Computer Science (JORBEL)*, vol. 34, pp.39-53.

DaSilva, L. A (1999). Telecommunications networks – Lecture 2: Telephone networks. <http://www.edu/~ee4614ld/Lecture2.html>

Deneubourg, J.-L. (1977) Application de l'ordre par fluctuations la description de certaines etapes de la construction du nid chez les termites, *Insect. Soc.*, vol. 24, pp.117-130.

Deneubourg, J.-L. & Goss, S. (1989). Collective patterns and decision-making. *Ethology, Ecology and Evolution 1*, pp. 295-311.

Deneubourg, J.-L., Aron, S., Goss, S. & Pasteels, J. M. (1990). The self-organizing exploratory pattern of the argentine ant. *Journal of Insect Behavior*, vol. 3, pp. 159-168.

Di Caro, G. & Dorigo, M. (1997). AntNet: a mobile agents approach to adaptive routing. Technical Report 97-12, IRIDIA, Université Libre de Bruxelles.

Di Caro, G. & Dorigo, M. (1998). AntNet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research (JAIR)*, vol. 9, pp. 317-365, December.

Dorigo, M. (1992). Optimization, learning and natural algorithms (in Italian). *PhD thesis*, Dipartimento di Elettronica e Informazione, Politecnico di Milano, IT.

Dorigo, M. & Gambardella, L. M. (1997). Ant colonies for the traveling salesman problem. *BioSystems*, vol. 43, pp. 73-81.

Dorigo, M. & Di Caro, G. (1998). AntNet: distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research*, vol. 9, pp. 317-365.

Dorigo, M., Maniezzo, V. & Colomi, A. (1991) Positive feedback as a search strategy. Technical Report 91-106, Dipartimento di Elettronica, Politecnico di Milano, IT.

Dorigo, M., Maniezzo, V. & Colomi, A (1996). The ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics B*, vol. 26, pp.1-13.

Dorigo, M., Di Caro, G. & Gambardella, L. M (1999). Ant algorithms for discrete optimization. In *Artificial Life*, MIT Press.

Downes, K., Ford, M., Kim Lew, H., Spanier, S. & Stevenson, T. (1998). *Internetworking technologies handbook*, Second Edition. Cisco Press.

Ford, L. & Fulkerson, D. (1962). *Flows in networks*. Prentice-Hall.

Franks, N. R. (1989). Army ants: A collective intelligence. *American Scientist*, vol. 77, March-April.

Gambardella, L. M. & Dorigo, M. (1997). HAS-SOP: A hybrid ant system for the sequential ordering problem. Technical Report 11-97, IDSIA, Lugano, CH.

Goetz, P., Kumar, S. & Miikkulainen, R. (1996). On-line adaptation of signal predistorter through dual reinforcement learning, *Proc. Machine Learning: Proceedings of the 13th Annual Conference* (Bari, Italy).

Goss, S., Aron, S., Deneubourg, J.-L & Pasteels, J. M. (1989). Self-organized shortcuts in the Argentine ant. *Naturwissenschaften*, vol. 76, pp. 579-581.

Grassé, P. P. (1946). *Les insects dans leur univers*. Ed. du Palais de la découverte, Paris.

Guérin, S. (1997). Optimisation multi-agents en environnement dynamique: application au routage dans les réseaux de télécommunications. DEA Dissertation, University of Rennes I and Ecole Nationale Supérieure des Télécommunications de Bretagne.

Haken, H. (1977). *Synergetic*. Springer-Verlag.

Hölldobler, B. & Wilson, E. O (1994). *Journey to the ants*. Bellknap Press / Harvard University Press.

Kaelbling, L. P., Littman, M. L. & Moore, A W. (1996). Reinforcement learning: A survey. *Journal of AI Research*, vol. 4, pp. 237-285.

Kumar, S. & Miikkulainen, R. (1997). Dual reinforcement Q-routing: an on-line adaptive routing algorithm. In *Proceedings of Artificial Neural Networks in Engineering*.

Lin, L. (1992). Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, vol. 8.

Littman, M. & Boyan, J. A (1993). A distributed reinforcement learning scheme for network routing. *Proceedings of the First International Workshop on Applications of Neural Networks to Telecommunications*, Hillside, New Jersey, pp. 45-51. Erlbaum.

Mahadevan & Connell (1991). Automatic programming of behavior-based robots using reinforcement learning. In *Proceedings of the 1991 National Conference on AI*, pp. 768-773.

Michel, R. & Middendorf, M. (1998). An island model based ant system with lookahead for the shortest supersequence problem. In Eiben, A E., Back, T., Schoenauer, M. & Schwefel, H.-P. (Eds), *Proceedings of PPSN-V, Fifth International Conference on Parallel Problem Solving from Nature*, pp. 692-701. Springer –Verlag.

Montgmorey, D. C. (1997). *Design and analysis of experiments*. New York: John Wiley.

Nicolis, G. & Prigogine, I. (1977). *Self-organization in non-equilibrium systems*. Wiley.

Peng, J. & Williams, R. (1994). Incremental multi-step Q-learning. In W. W. Cohen and H. Hirsh (eds.), *Proceedings of the Eleventh International Conference on Machine Learning*, pp. 226-232. Morgan Kaufmann, San Francisco.

Peng, J. & Williams, R. (1996). Incremental multi-step Q-learning. *Machine Learning*, vol. 22, pp. 283-290.

Russel, S. & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Prentice-Hall, London.

Schoonderwoerd, R., Holland, O., Bruten, J. & Rothkrantz, L. (1996). Ant-based load balancing in telecommunications networks. *Adaptive Behavior*, vol. 5, n° 2, pp.169-207.

Schoonderwoerd, R., Holland, O. & Bruten, J. (1997). Ant-like agents for load balancing in telecommunications networks. In *Proceedings of the First International Conference on Autonomous Agents*, pp. 209-216. ACM Press.

Schmidhuber, J. H. (1996). A general method for multi-agent learning and incremental self-improvement in unrestricted environments. In X. Yao, editor, *Evolutionary Computation: Theory and Applications*. Scientific Publ. Co., Singapore.

Singh, S. P. (1994). Learning to solve markovian decision processes. *PhD thesis*. Department of Computer Science, University of Massachusetts.

Soares, L. F. G., Lemos, G. & Colcher, S. (1995). *Redes de computadores: das LANs, MANs e WANs às redes ATM*. Rio de Janeiro, Campus.

Steenstrup, M. E. (1995). *Routing in telecommunications networks*. Prentice-Hall.

Stickland, T. R. Tofts, C. M. N. & Franks, N. R. (1992). A path choice algorithm for ants. In *Naturwissenschaften*, vol. 79, pp. 567-572.

Stützle, T. & Hoos, H. (1997). The MAX-MIN ant system and local search for the traveling salesman problem. In Baeck, T., Michalewicz, Z. & Yao, X., (Eds), *Proceedings of IEEE-ICEC-EPS'97, IEEE International Conference on Evolutionary Computation and Evolutionary Programming Conference*, pp. 309-314. IEEE Press.

Sutton, R. S. (1988). Learning to predict by the methods of temporal difference. *Machine Learning*, vol. 3, pp. 9-44.

Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the Seventh International Conference on Machine Learning*. San Mateo, CA: Morgan Kaufmann.

Sutton, R. S. & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Theraulaz, G. & Bonabeau, E. (1999) A brief history of stigmergy. *Artificial Life*, vol. 5, n° 2, pp. 97-116.

Theraulaz, G. & Bonabeau, E. (1995) Coordination in distributed building, *Science*, vol. 269, pp. 686-688.

Watkins, C. J. C. H. (1989). Models of delayed reinforcement learning. *PhD thesis*, Psychology Department, Cambridge University.

Watkins, C. J. C. H. & Dayan, P. (1992). Q-learning. *Machine Learning*, vol. 8, n° 3, pp. 279-292.

White, T., Pagurek, B. & Oppacher, F. (1998). Connection management using adaptive mobile agents. In Arabnia, H. R., (Ed), *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA '98)*, pp. 802-809. CSREA Press.

Wilson, E. O (1975). *Sociobiology*. Belknap Press / Harvard University Press.

Apêndice I

I.1 – Algoritmo TD(λ)

O algoritmo geral de diferença temporal TD(λ) realiza a integração das estratégias de programação dinâmica, métodos de Monte Carlo e métodos de diferença temporal utilizando traços de elegibilidade. Estes são um dos mecanismos básicos para resolver o problema da atribuição de crédito temporal enfrentado pelos algoritmos de aprendizagem por reforço (Sutton & Barto, 1998).

Um traço de elegibilidade é um registro temporário sobre a ocorrência de um determinado evento, como a visita a um estado do ambiente ou a seleção de uma ação. O traço atribui os parâmetros de memória associados com os eventos ocorridos, de forma a classificá-los como elegíveis para mudanças no aprendizado em andamento. Quando o processo incorre em erro, somente os estados ou ações elegíveis recebem crédito ou culpa pelo erro. Desta forma, os traços de elegibilidade ajudam a integrar os eventos e a informação de treinamento relacionadas com a ocorrência dos mesmos.

Em sua forma mais simples, é atribuído um traço a cada estado do ambiente. O traço de elegibilidade de um estado s no tempo t é representado por $e_t(s)$. Em cada passo da trajetória, os traços para todos os estados visitados decaem de acordo com uma taxa representada por $\gamma\lambda$ e o traço correspondente ao estado visitado s é incrementado por 1:

$$e_t(s) = \begin{cases} \gamma\lambda e_{t-1}(s) & \text{se } s \neq s_t \\ \gamma\lambda e_{t-1}(s) + 1 & \text{se } s = s_t \end{cases} \quad (\text{I.1})$$

no qual γ é o parâmetro de desconto usualmente utilizado e λ o parâmetro de decaimento do traço de elegibilidade, $0 \leq \lambda \leq 1$.

Este tipo de traço de elegibilidade é denominado traço acumulativo, uma vez que ele é incrementado a cada vez que o estado é visitado e decai gradualmente quando o contrário ocorre. A qualquer instante de tempo, o conjunto de traços registra quão recentemente um estado foi visitado. O traço indica, então, o quanto os estados são importantes para que o aprendizado ocorra.

O algoritmo $TD(\lambda)$ é obtido através da combinação dos traços de elegibilidade descritos acima com o erro de diferença temporal utilizado no algoritmo TD, apresentado na equação (I.3):

$$\Delta V_t(s) = \alpha [r_{t+1} + \gamma V(s_{t+1}) - V(s_t)] e_t(s), \text{ para } \forall s \in S \quad (I.2)$$

no qual $e_t(s)$ corresponde ao traço de elegibilidade de um estado s no tempo t .

O erro de diferença temporal para a predição do valor do estado no tempo t é representado por δ_t :

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \quad (I.3)$$

Observa-se que o erro TD sofre variação a cada passo de tempo, mas não de um estado a outro, uma vez que ele é utilizado na atualização de todos os estados do ambiente a cada passo de tempo. Assim, estados diferentes recebem crédito através deste erro a cada passo de tempo, de acordo com o valor do seu traço de elegibilidade no instante anterior. Assim, as atualizações são proporcionais ao valor do traço de elegibilidade de todos os estados visitados:

$$\Delta V_t(s) = \alpha \delta_t e_t(s) \text{ para } \forall s \in S \quad (I.4)$$

O algoritmo $TD(\lambda)$ é definido como uma família de algoritmos parametrizada por λ , o fator de decaimento do traço. Os dois algoritmos localizados nos extremos desta família são denominados $TD(0)$ e $TD(1)$, respectivamente. Se $\lambda = 0$, então através da eq.(I.1) todos os traços recebem valor zero no tempo t , exceto o traço correspondente ao estado s_t . Se $\lambda = 1$, o traço associado a cada estado s no instante anterior decai de acordo com γ , enquanto o traço correspondente ao estado s_t recebe este decaimento e é incrementado por 1.

Na atualização online, os incrementos de $V(s)$ são realizados a cada passo de tempo. No caso do modelo de horizonte finito, os incrementos podem ser realizados *offline*, ou seja, após o fim da trajetória do agente, ou através da atualização *batch*, em que os incrementos são acumulados sem que haja mudança nas estimativas enquanto um conjunto de tentativas é apresentado. Esta última é raramente utilizada na prática.

O algoritmo $TD(\lambda)$ que realiza os incrementos passo a passo é mostrado a seguir:

Inicialize $V(s)$ arbitrariamente e $e(s) = 0$, para todo $s \in S$;
 Repita (para cada trajetória):
 Perceba o estado atual do ambiente s_t ;
 Repita (para cada passo da trajetória):
 Escolha ação a_t ;
 Realize a ação a_t , observe o retorno r_{t+1} e o próximo estado s_{t+1} ;
 Calcule o erro e o traço de elegibilidade do estado s_t :
 $\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$
 $e_t(s) = \gamma \lambda e_{t-1}(s) + 1$
 Calcule o traço de elegibilidade e atualize o valor $V_t(s)$ para todos os estados $s \in S$:
 $e_t(s) = \gamma \lambda e_{t-1}(s)$
 $\Delta V_t(s) = \alpha \delta_t e_t(s)$
 Faça $s_t = s_{t+1}$;
 Até que s_t seja o estado final do ambiente.
 Até que a trajetória seja a última.

Neste caso, a_t = ação dada pela política π para s_t e $t = 0, 1, 2, 3, \dots$

I.2 – Algoritmo Sarsa(0)

O algoritmo Sarsa estima $Q_\pi(s_t, a_t)$ para a política de comportamento atual π , para todos os estados $s \in S$ e todas as ações $a \in A$. Isto pode ser realizado do mesmo modo descrito anteriormente para estimar $V_\pi(s_t)$. Considera-se, assim, transições de pares estado-ação e aprende-se os valores destes pares. Os teoremas que asseguram a convergência do algoritmo TD(0) também são aplicados ao algoritmo Sarsa. Sua função de avaliação é atualizada como se segue:

$$\Delta Q(s_t, a_t) = \alpha [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (I.5)$$

O nome do algoritmo é decorrente dos elementos $(s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1})$ utilizados na regra de atualização dos valores dos pares estado-ação. Esta atualização é realizada após cada transição, até que o estado final do problema seja alcançado. Se o estado s_{t+1} é o estado final, a função $\Delta Q(s_{t+1}, a_{t+1})$ deve ser levada a zero.

Como nos demais métodos dependentes da política, estima-se continuamente $Q_\pi(s_t, a_t)$ para a política π e ao mesmo tempo realiza-se a mudança desta política em direção a uma política avara² com relação a Q_π .

² Uma política avara é aquela que seleciona ações que apresentam o maior valor estimado (Kaelbling et al., 1996).

A forma geral deste algoritmo é mostrada a seguir:

Inicialize $Q(s, a)$ arbitrariamente;
 Repita (para cada trajetória):
 Perceba o estado atual do ambiente s_t ;
 Repita (para cada passo da trajetória):
 Escolha uma ação a_t ;
 Realize a ação a_t , observe o retorno r_{t+1} e o próximo estado s_{t+1} ;
 Escolha uma ação a_{t+1} ;
 Atualize o valor $Q(s_t, a_t)$:
 $Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$
 Faça $s_t = s_{t+1}$; $a_t = a_{t+1}$;
 Até que s_t seja o estado final.
 Até que a trajetória seja a última.

Neste caso: a_t = ação dada pela política π para o estado s_t ;
 a_{t+1} = ação dada pela política π para o estado s_{t+1} .
 $t = 0, 1, 2, 3, \dots$

Sarsa converge com probabilidade 1 para uma política ótima e função de avaliação da ação se todos os pares estado-ação forem visitados um número infinito de vezes e a política convergir para uma política avara (Sutton & Barto, 1998).

I.3 – Sarsa(λ)

O algoritmo Sarsa(λ) aplica o método TD(λ) a pares estado-ação, ao invés de estados. Assim, é atribuído a cada par estado-ação (s, a) um traço de elegibilidade, representado por $e_t(s, a)$.

A regra de atualização do algoritmo é a seguinte:

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t e_t(s, a) \quad \text{para } \forall s, a \quad (\text{I.6})$$

$$\text{na qual: } \delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \quad (\text{I.7})$$

$$e_t(s, a) = \begin{cases} \gamma \lambda e_{t-1}(s, a) + 1 & \text{se } s = s_t \text{ e } a = a_t \text{ para } \forall s, a \\ \gamma \lambda e_{t-1}(s, a) & \text{caso contrário} \end{cases} \quad (\text{I.8})$$

A forma geral deste algoritmo é mostrada a seguir:

Inicialize $Q(s, a)$ arbitrariamente e $e(s, a) = 0$ para todo s, a ;
 Repita (para cada trajetória):

Inicialize s, a ;
 Perceba o estado atual do ambiente s_t ;
 Repita (para cada passo da trajetória):
 Escolha uma ação a_t ;
 Realize a ação a_t , observe o retorno r_{t+1} e o próximo estado s_{t+1} ;
 Escolha a ação a_{t+1} de s_{t+1} usando política derivada de Q;
 Calcule o erro e o traço de elegibilidade do par estado-ação (s_t, a_t) :
 $\delta_t = r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)$
 $e(s_t, a_t) = e(s_t, a_t) + 1$
 Atualize os valores $Q(s, a)$ e $e(s, a)$ para todo s, a :
 $Q(s, a) = Q(s, a) + \alpha \delta_t e(s, a)$
 $e(s, a) = \gamma \lambda e(s, a)$
 Faça $s_t = s_{t+1}$; $a_t = a_{t+1}$;
 Até que s_t seja o estado final.
 Até que a trajetória seja a última.

Neste caso, a_t = ação dada pela política π para o estado s_t e $t = 0, 1, 2, 3, \dots$
 a_{t+1} = ação dada pela política π para o estado s_{t+1} .
 $t = 0, 1, 2, 3, \dots$

Como o algoritmo Sarsa(0), Sarsa(λ) é um método dependente da política.

I.4 – Aprendizagem-Q(λ) de Watkins

O algoritmo Q(λ) de Watkins aplica o método TD(λ) a pares estado-ação, ao invés de somente aos estados. Os traços de elegibilidade são utilizados da mesma forma que anteriormente, porém é atribuído um traço a cada par estado-ação.

O algoritmo de aprendizagem-Q é um método independente da política, isto é, a política aprendida pode ser diferente daquela seguida para selecionar as ações realizadas. De modo particular, a aprendizagem-Q aprende uma política avara, enquanto segue uma política que envolve ações exploratórias, ou seja, ocasionalmente são selecionadas ações subótimas com relação a função de avaliação $Q(s_t, a_t)$.

Assim, suponha que é atribuído crédito ao par estado-ação (s_t, a_t) no tempo t e que nos próximos dois passos de tempo o agente seleciona uma ação avara; porém, no terceiro instante, o agente seleciona uma ação exploratória. Ao aprender o valor de uma política avara para (s_t, a_t) , a experiência obtida após este instante só poderá ser utilizada enquanto uma política avara estiver sendo seguida. Logo, os retornos obtidos no primeiro e segundo passos de tempo poderão ser utilizados, porém aquele obtido no terceiro passo será excluído, recebendo valor igual a zero. Do mesmo modo, os retornos obtidos após o terceiro instante de tempo não serão considerados.

Então, de forma diferente do algoritmo TD(λ) ou Sarsa(λ), a aprendizagem-Q(λ) de Watkins não se baseia em todo o caminho até o fim da tentativa para realizar a atualização dos valores-Q. Ela se baseia somente nos valores decorridos de ações não exploratórias: considera os retornos obtidos até que uma ação deste tipo ocorra.

A implementação do algoritmo Q(λ) envolve dois procedimentos. Primeiro, os traços para todos os pares estado-ação decaem por um fator $\gamma\lambda$, ou, se uma ação exploratória foi realizada, os traços correspondentes aos estados envolvidos recebem valor zero. Então, o traço correspondente ao estado e ação atuais é incrementado por 1. Assim, são realizadas as atualizações descritas a seguir:

Para o par (s_t, a_t) :

$$e_t(s_t, a_t) = \begin{cases} \gamma\lambda e_{t-1}(s_t, a_t) + 1 & \text{se } a_t = \max_a Q(s_t, a) \\ 1 & \text{caso contrário} \end{cases} \quad (\text{I.9})$$

e para todos os outros pares (s, a) :

$$e_t(s, a) = \begin{cases} \gamma\lambda e_{t-1}(s, a) + 1 & \text{se } a = \max_a Q(s, a) \\ 0 & \text{caso contrário} \end{cases} \quad (\text{I.10})$$

O algoritmo Q(λ) de Watkins é mostrado a seguir:

Inicialize $Q(s, a)$ arbitrariamente e $e(s) = 0$, para todo s, a ;
 Repita (para cada trajetória):
 Perceba o estado atual do ambiente s_t ;
 Repita (para cada passo da trajetória):
 Escolha ação a_t ;
 Realize a ação a_t , observe o retorno r_{t+1} e o próximo estado s_{t+1} ;
 Escolha a ação a_{t+1} ;
 Selecione a melhor ação:

$$a^* = \arg \max_b Q(s_{t+1}, b)$$

 Calcule o traço de eligibilidade e o erro para o par (s_t, a_t) :

$$\delta_t = r_{t+1} + \gamma Q(s_{t+1}, a^*) - Q(s_t, a_t)$$

 Se $a_{t+1} = a^*$, então $e_t(s_t, a_t) = e_t(s_t, a_t) + 1$
 Senão $e_t(s_t, a_t) = 1$
 Atualize o valor de todos os pares (s, a) :

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \delta_t e_t(s, a)$$

 Se $a_{t+1} = a^*$, então $e_t(s, a) = \gamma \lambda e_t(s, a)$
 Senão $e_t(s, a) = 0$
 Faça $s_t = s_{t+1}$ e $a_t = a_{t+1}$
 Até que s_t seja o estado final.
 Até que a trajetória seja igual a última.

Neste caso: a_t = ação dada pela política π para o estado s_t ;
 a_{t+1} = ação dada pela política π para o estado s_{t+1} .
 $t = 0, 1, 2, 3, \dots$

O algoritmo de aprendizagem-Q(λ) de Watkins aprende uma função de avaliação no espaço estado-ação, ao invés de somente no espaço de estados, como TD(λ). Assim, ele é capaz de discriminar os efeitos produzidos pela escolha de ações diferentes em cada estado do ambiente.

I.5 – Aprendizagem-Q(λ) de Peng & Williams

Ao atribuir valor zero aos traços de elegibilidade quando ações exploratórias são realizadas, o algoritmo Q(λ) de Watkins perde grande parte da vantagem da utilização desta estratégia. Se as ações exploratórias são frequentes, o que ocorre geralmente no início do aprendizado, então somente serão realizadas atualizações baseadas em mais de um ou dois passos, o que torna o aprendizado pouco mais rápido que o do algoritmo de aprendizagem-Q de um passo, desenvolvido por Watkins (1989). O algoritmo Q(λ) de Peng & Williams (1994, 1996) procura resolver este problema não fazendo distinção entre ações baseadas em

uma política avara e ações exploratórias. Desta forma, ele pode ser visto como um método híbrido dos algoritmos Sarsa(λ) e Q(λ) de Watkins.

Assumindo que a política de decisão do agente é uma política avara, são calculados dois valores de erros:

$$e_t = r_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t) \quad (\text{I.11})$$

e

$$e'_t = r_{t+1} + \gamma V_t(s_{t+1}) - Q_t(s_t, a_t) \quad (\text{I.12})$$

nas quais $V(s_t) = \max_{a_t} Q(s_t, a_t)$.

Assim, a eq. (I.12) é utilizada para um passo e a eq. (I.11) deste valor em diante. A eq.(I.11) representa o erro considerando a estimativa da ação que maximiza a função de avaliação do par estado-ação, enquanto a eq.(I.12) representa o erro considerando a ação efetivamente realizada.

A cada passo de tempo, o algoritmo Q(λ) atualiza um conjunto de pares estado-ação que cresce linearmente com o tempo. No pior caso, este conjunto pode ser composto por todo o espaço estado-ação. Entretanto, o número requerido de pares pode ser melhor gerenciado mantendo somente os pares para os quais o traço de elegibilidade $(\gamma\lambda)^n$ é significativo, uma vez que este valor decai exponencialmente quando $\gamma\lambda < 1$.

O algoritmo é descrito a seguir:

Inicialize $Q(s, a) = 0$ e $e(s) = 0$, para todo $s \in S$ e $a \in A$;

Repita (para cada trajetória):

Perceba o estado atual do ambiente s_t ;

Repita (para cada passo da trajetória):

Escolha uma ação a_t ;

Realize a ação a_t , observe o retorno r_{t+1} e o próximo estado s_{t+1} ;

Calcule os valores:

$$e'_t = r_{t+1} + \gamma V_t(s_{t+1}) - Q_t(s_t, a_t)$$

$$e_t = r_{t+1} + \gamma V_t(s_{t+1}) - V_t(s_t)$$

Atualize os valores cada par estado-ação (s, a) :

$$e(s, a) = \gamma \lambda e(s, a)$$

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha e(s, a) e_t$$

Atualize os valores para o par estado-ação atual (s_t, a_t) :

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha e'_t$$

$$e(s_t, a_t) = e(s_t, a_t) + 1$$

Faça $s_t = s_{t+1}$ e $a_t = a_{t+1}$;

Até que s_t seja o estado final.

Até que a trajetória seja a última.

Neste caso, a_t = ação dada pela política π para o estado s_t e $t = 0, 1, 2, 3, \dots$

Para uma política fixa não-avara, a função $Q(s, a)$ converge para uma combinação de Q_π com Q^* . Entretanto, se a política se torna gradualmente avara, o método ainda pode convergir para Q^* . Este algoritmo tem apresentado bons resultados empiricamente. Vários estudos mostram que este método tem desempenho superior àquele do algoritmo $Q(\lambda)$ de Watkins e similar ao do algoritmo Sarsa(λ), segundo Sutton & Barto (1998). Em sistemas contínuos no tempo, ou mesmo em alguns sistemas de tempo discreto, a utilização de algoritmos que propagam a informação aprendida a somente um passo anterior a cada instante de tempo é inviável. Nestes casos, o emprego do fator λ se torna necessário.