

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE TECNOLOGIA

Tito Ricardo Bianchin Oliveira

Otimização do consumo de energia em terminais móveis 3G

Limeira, 2011.

Tito Ricardo Bianchin Oliveira

Otimização do consumo de energia em terminais móveis 3G

Dissertação apresentada ao Curso de Mestrado da Faculdade de Tecnologia da Universidade Estadual de Campinas, como requisito para a obtenção do título de Mestre em Tecnologia.

Área de Concentração: Tecnologia e Inovação

Orientador: Prof. Dr. Varese Salvador Timóteo
Co-orientador: Prof. Dr. Edson Luiz Ursini

Limeira
2011

FICHA CATALOGRÁFICA ELABORADA POR SILVANA MOREIRA DA SILVA SOARES –
CRB-8/3965
BIBLIOTECA UNIFICADA FT/CTL
UNICAMP

OL4 Oliveira, Tito Ricardo Bianchin, 1986-
Otimização do consumo de energia em terminais móveis
3G / Tito Ricardo Bianchin Oliveira. – Limeira, SP : [s.n.],
2011.

Orientador: Varese Salvador Timóteo.
Dissertação (mestrado) – Universidade Estadual de
Campinas, Faculdade de Tecnologia.

1. Consumo de energia. 2. Dispositivo móvel. I. Timóteo,
Varese Salvador. II. Universidade Estadual de Campinas.
Faculdade de Tecnologia. III. Título.

Informações para Biblioteca Digital

Título em inglês: Energy consumption in 3G mobile terminals

Palavras-chave em inglês (Keywords):

- 1- Energy consumption
- 2- 3G
- 3- Mobile device
- 4- Always on
- 5- Prediction model

Área de concentração: Tecnologia e Inovação

Titulação: Mestre em Tecnologia

Banca examinadora: Varese Salvador Timóteo, Rangel Arthur, Luiz Henrique Bonani

Data da Defesa: 14-12-2011

Programa de Pós-Graduação em Tecnologia

DISSERTAÇÃO DE MESTRADO ACADÊMICO

Otimização do consumo de energia em terminais móveis 3G

Autor: Tito Ricardo Bianchin Oliveira

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação:



Prof. Dr. Varese Salvador Timóteo, Presidente
FT/UNICAMP



Prof. Dr. Rangel Arthur
FT/UNICAMP



Prof. Dr. Luiz Henrique Bonani
CECS/UFABC

Dedico este trabalho à minha família.

Agradecimentos

A Deus e a todas as forças externas que me iluminaram em todos os momentos.

A meus pais pelo eterno apoio.

A minha namorada, por me aguentar nos momentos de stress de trabalho e estudo.

A Samsung Instituto de Desenvolvimento para Informática, por me disponibilizar recursos de tempo para assistir aulas de mestrado e custos com a pesquisa deste mestrado, especialmente aos gerentes que acreditaram na minha idéia e ao grupo de System.

Aos meus orientadores por acreditarem na minha idéia e dar o suporte para obtenção do mestrado e a publicação de artigos internacionais, dois dos meus grandes objetivos pessoais.

A todos que acreditaram no meu trabalho, e principalmente aos que não acreditaram, pois me fizeram superar os desafios pra conquistar meu objetivo.

*"É justamente a possibilidade
de realizar um sonho que torna
a vida interessante"
(Paulo Coelho)*

RESUMO

O crescimento das redes de terceira geração, aliado a sua alta velocidade para transmissão de dados e banda disponível fazem com que novos terminais lançados no mercado utilizem o comportamento *Always On*, no qual o dispositivo fica 100% do tempo conectado à rede para transmissão e recepção de dados. Esse comportamento, no entanto, faz com que o consumo de bateria do dispositivo seja maior devido ao uso de aplicativos que recebem informações periodicamente, e principalmente pelo recebimento de pacotes não solicitados provenientes de ataques à rede. Este trabalho tem como objetivo analisar os elementos de rede responsáveis pela transmissão de pacotes de dados, identificando os fatores responsáveis pelo aumento de consumo. Ao final, é proposto um método para melhor aproveitamento dos recursos de rádio para transmissão de dados e conseqüentemente, a diminuição do consumo de energia, utilizando um modelo de previsão.

Palavras-Chave: Consumo de energia, 3G, Dispositivo Móvel, *Always On*, Modelo de Previsão

ABSTRACT

The expansion of third generation network and its high speed of data transmission and available bandwidth, made that new designed mobile devices use the “Always On” concept, in which the device is 100% connected in packet switch network, and able for data transmission. This behavior makes the device’s energy consumption to be higher due to usage of applications that receives periodically information, and mainly due to the unsolicited packages from hacker attack. This work has as main purpose analyze the network elements responsible for data package transmission, identifying the main factors related to the energy consumption increasing. Finally, it is proposed a method to enhance the radio resources for data transmission and energy consumption decreasing, using a prevision model.

Key Words: Energy Consumption, 3G, Mobile Device, Always On, Prediction Model

LISTA DE ILUSTRAÇÕES

Figura 2.1. Estabelecimento de conexão de dados. [CHEN, 2004].....	6
Figura 2.2. Estados do contexto PMM.	7
Figura 2.3. Transição de estados do contexto PDP. [CHEN, 2004].....	9
Figura 2.4. Procedimento de estabelecimento de conexão no RRC. [HOLMA, 2010].....	10
Figura 2.5. Conexão com o RRC para envio de <i>paging</i> . [HOLMA, 2010].....	12
Figura 2.6. Máquina de estado para transmissão de dados no RRC. [TACHIKAWA, 2002]	14
Figura 2.7. Procedimento de <i>Measurement Control</i> . [HOLMA, 2010]	16
Figura 2.8. Monitoramento de tráfego via canais de medição [TS25303]	17
Figura 2.9. Transição de estados do Cell_DCH para Cell_FACH.	18
Figura 2.10. Relação entre fluxo de dados recebido, consumo e estados do RRC.....	19
Figura 2.11. Recebimento de um pacote em uma rede 3G. [HAVERINEN,2007].....	20
Figura 3.1. Distribuição de pacotes não solicitados recebidos ao longo do tempo.	22
Figura 3.2. Principais portas atacadas.	23
Figura 3.3. Implantação de <i>Firewalls</i> nas interfaces de rede.	24
Figura 3.4. (a) Rede aberta desprotegida contra ataques e (b) Rede protegida contra ataques (<i>Firewall</i> e/ou NAT).....	25
Figura 3.5. Mensagens <i>keep-alive</i> recebidas pelo dispositivo móvel. [HAVERINEN, 2007].....	26
Figura 3.6. Relação entre consumo 2G e 3G para mensagens <i>keep-alive</i> [HAVERINEN, 2007]	27
Figura 3.7. Representação do tempo de inatividade dos canais para uma operadora específica... ..	28
Figura 3.8. Desconexão do RRC requisitada pelo dispositivo.	29
Figura 3.9. Consumo gerado pelo recebimento de um pacote para um $T2$ de (a) 30 segundos (rede) e (b) 5 segundos (telefone).....	30
Figura 3.10. Consumo gerado pelo recebimento de dois pacotes a um intervalo de 15 segundos para um $T2$ de (a) 30 segundos (rede) e (b) 5 segundos (telefone).	31
Figura 3.11. Consumo gerado pelo recebimento de dois pacotes em um intervalo de 15 segundos para um tempo de inatividade dinâmico – estimado.	33
Figura 4.1. Passos para estimação utilizando filtro de Kalman.....	38
Figura 4.2. Algoritmo para cálculo do Filtro de Kalman.	38
Figura 4.3. Dedução do subfiltro.	39
Figura 4.4. Algoritmo para cálculo do Filtro de Kalman com subfiltro.	42
Figura 5.1. Função densidade de probabilidade Exponencial com diferentes valores de λ	44
Figura 5.2. Função Densidade de Probabilidade Poisson para diferentes valores de λ	45
Figura 5.3. Função Densidade de Probabilidade lognormal para diferentes valores de μ e σ	47
Figura 5.4. Função Densidade de Probabilidade Weibull para diferentes valores de α e β	48
Figura 5.5. Função Densidade de Probabilidade Pareto para diferentes valores de k e $\alpha=1$	49
Figura 5.6. Função Densidade de Probabilidade Gamma para diferentes valores de α e β	50
Figura 5.7. Função Densidade de Probabilidade Beta para diferentes valores de α e β	52
Figura 5.8. Função Densidade de Probabilidade de Erlang para $\lambda=0,5$ e diferentes valores de k	53
Figura 5.9. Transformação inversa da FDA.	55
Figura 5.10. Classificação de tráfego [PINOTTI, 2011].	57
Figura 5.11. Distribuição dos pacotes ao longo do tempo para (a) pacotes não solicitados e (b) e-mail.	59

Figura 5.12. (a) Resultados do ajuste de distribuições utilizando o “Fit All” do “Input Analyzer” e (b) Parâmetros obtidos após ajuste para determinada distribuição.....	60
Figura 5.13. Correspondência gráfica entre os dados reais (gráfico de barra) e a distribuição ajustada (gráfico de linha) gerada pelo Input Analyzer.....	61
Figura 6.1. Modelo de simulação baseado na rede real. [TITO, 2011].....	64
Figura 6.2. Relação entre diferentes valores de $T2$ para um alto tráfego e o consumo de energia. [OLIVEIRA, 2011].....	67
Figura 6.3. Modelo de simulação do Filtro de Kalman.....	68
Figura 6.4. Tráfego real (azul) e tráfego estimado (vermelho) para T_SAMPLE (a)1s (b)0.001 s.	70
Figura 6.5. Modelo de simulação baseado na rede real incluindo modelo de previsão.	72
Figura 6.6. Comparação entre o consumo obtido para o modelo utilizando filtro de Kalman (Modelo II B) e limiar de consumo sem filtro de Kalman (Modelo I) para baixo, médio e alto tráfego para (a) $T1=5s$ e $T2=30s$ e (b) $T1=5s$ e $T2=5s$	79
Figura 6.7. Comparação entre o consumo obtido para o modelo utilizando o modelo II B controlado apenas pelo filtro de Kalman e o limiar de consumo sem filtro de Kalman (Modelo I) para baixo, médio e alto tráfego para $T1=0s$ e $T2=infinito$	80
Figura 6.8. Implementações para o filtro de Kalman em diversas linguagens - validação.	81
Figura 6.9. Resultado da previsão utilizando filtro de Kalman para as três implementações: C, C simplificado e MATLAB.....	82
Figura 6.10. Aplicação “Terminal” mostrando a execução do software do Kalman para medição de desempenho.	83
Figura 6.11. Comparação de desempenho do algoritmo em um computador pessoal e nos dispositivos móveis.	84
Figura A.1. Estrutura básica do sistema UMTS. [VALE, 2006].....	91
Figura A.2. Principais interfaces e sua ligação com elementos de rede.....	96
Figura A.3. Detalhamento do núcleo da rede do sistema UMTS [VALE, 2006].....	96
Figura A.4. Estrutura vertical da interface aérea do sistema UMTS.....	97
Figura A.5. Planos vertical e horizontal do sistema UMTS.....	98

LISTA DE TABELAS

TABELA 2.1. Valores possíveis para o elemento <i>Reporting Threshold</i> . [TS25331]	15
TABELA 2.2. RRC STATE INDICATOR – Indicação de estado para o UE.[TS25331].....	18
TABELA 3.1. Tráfego de pacotes não solicitados.	22
TABELA 5.1. Caracterização de tráfego de navegação web [LEE-CHEN, 2004].	57
TABELA 5.2. Caracterização de tráfego de Vídeo Streaming [LEE-CHEN, 2004].	58
TABELA 5.3. Ajuste de Distribuição para diferentes tipos de tráfego da rede real.	62
TABELA 6.1. Resultados da simulação no Arena para Baixo, médio e Alto tráfego. [TITO,2011]	66
TABELA 6.2. Erro quadrático médio de previsão para diferentes características de tráfego.....	71
TABELA 6.3. Simulação A – Filtro de Kalman atuando junto com $T1=5s$ e $T2=30s$	74
TABELA 6.4. Simulação B – Filtro de Kalman atuando junto com $T1=5s$ e $T2=5s$	75
TABELA 6.5. Simulação C – Filtro de Kalman atuando sozinho no modelo ($T1=0s$ e $T2=infinito$).	76
TABELA 6.6. Eficiência do Filtro para os cenários simulados	76
TABELA 6.7. Comparativo de Consumo para um $T2=30s$ (Modelo I) contra um $T2=Dinâmico$ (Modelo II-B) e um $T2=5s$ (Modelo I)	77
TABELA 6.8. Configurações dos dispositivos para teste de desempenho do algoritmo	84
TABELA A.1. Canais de Transporte do sistema UMTS.	99
TABELA A.2. Canais Lógicos do Sistema UMTS	99

LISTA DE ABREVIATURAS E SIGLAS

2G	<i>Second Generation</i> (Segunda Geração)
3G	<i>Third Generation</i> (Terceira Geração)
3GPP	<i>Third Generation Partnership Project</i> (Projeto de Parcerias para Terceira Geração)
4G	<i>Fourth Generation</i> (Quarta Geração)
ADB	<i>Android Debug Bridge</i> (Ponte de Depuração Android)
AICH	<i>Acquisition Indicator Channel</i> (Canal Indicador de Aquisição)
AMOLED	<i>Active-Matrix Organic Light-Emitting Diode</i> (Matriz Ativa de Diodos Orgânicos Emissores de Luz)
APN	<i>Access Point Name</i> (Nome de Ponto de Acesso)
ARM	<i>Advanced RISC Machine</i> (Máquina RISC Avançada)
ARQ	<i>Automatic Repeat Request</i> (Requisição de Repetição Automática)
AS	<i>Access Stratum</i>
BCCH	<i>Broadcast Control Channel</i> (Canal de Controle de Difusão)
BCH	<i>Broadcast Channel</i> (Canal de Difusão)
BMC	<i>Broadcast/Multicast Control</i> (Controle de Difusão/Multidifusão)
BM-SC	<i>Broadcast-Multicast Service Center</i> (Centro de Serviço de Difusão/Multidifusão)
BO	<i>Buffer Occupation</i>
BSS	<i>Base Station Subsystem</i> (Subsistema da Estação Base)
CC	<i>Call Control</i> (Controle de Chamada)
CCCH	<i>Common Control Channel</i> (Canal de Controle Comum)
CG	<i>Charger Gateway</i> (Gateway de Tarifação)
CPICH	<i>Common Pilot Channel</i> (Canal Piloto Comum)
CRF	<i>Charging Rule Function</i> (Função de Regra de Tarifação)

C-RNTI	<i>Cell - Radio Network Temporary Identifier</i> (Identificador Temporário da Rede de Rádio - Célula)
CS	<i>Circuit Switch</i> (Comutação por Circuito)
CSCF	<i>Call State Control Function</i> (Função de Controle de Estado de Chamada)
CTCH	<i>Common Traffic Channel</i> (Canal Comum de Tráfego)
DCCH	<i>Dedicated Control Channel</i> (Canal de Controle Dedicado)
DCH	<i>Dedicated Channel</i> (Canal Dedicado)
DoS	<i>Denial of Service</i> (Negação de Serviço)
DPCCH	<i>Dedicated Physical Control Channel</i> (Canal Dedicado de Controle Físico)
DPDCH	<i>Dedicated Physical Data Channel</i> (Canal Dedicado de Dados Físicos)
DRX	<i>Discontinuous Reception</i> (Recepção Descontínua)
DTCH	<i>Dedicated Traffic Channel</i> (Canal de Tráfego Dedicado)
FACH	<i>Forward Access Channel</i> (Canal de Acesso Avançado)
FDA	<i>Função Densidade Acumulada</i>
FDP	<i>Função Densidade Probabilidade</i>
FTP	<i>File Transfer Protocol</i> (Protocolo de Transferência de Arquivos)
GGSN	<i>Gateway GPRS Support Node</i> (Nó de apoio GPRS de Acesso)
GMSC	<i>GPRS Mobile Services Switching Center</i> Centro de Comutação de Serviço Móvel GPRS
GPRS	<i>General Packet Radio Service</i> (Serviço de Rádio de Pacote Geral)
GSM	<i>Global System for Mobile Communication</i> (Sistema Global para Comunicações Móveis)
GTP	<i>GPRS Tunneling Protocol</i> (Protocolo de Tunelamento GPRS)
GTP-C	<i>GTP Control</i> (GTP - Controle)
GTP-U	<i>GTP User</i> (GTP - Usuário)
HLR	<i>Home Location Register</i> (Registro de Assinantes Locais)
HSPA	<i>High Speed Packet Access</i> (Acesso a Pacotes em Alta Velocidade)

HSS	<i>Home Subscriber Server</i> (Servidor de Assinante Local)
I-CSCF	<i>Interrogation CSCF</i>
IMS	<i>IP Multimedia Subsystem</i> (Subsistema IP Multimídia)
IMSI	<i>International Mobile Subscriber Identity</i> (Identidade Internacional do Assinante Móvel)
IP	<i>Internet Protocol</i> (Protocolo de Internet)
LA	<i>Location Area</i> (Área de Localização)
LAI	<i>Location Area Identity</i> (Identificação de Área Localização)
LCD	<i>Liquid Crystal Display</i> (Display de Cristal Líquido)
LED	<i>Light-Emitting Diode</i> (Diodo Emissor de Luz)
MAC	<i>Media Access Control</i> (Controle de Acesso à Mídia)
MAP	<i>Mobile Application Part</i> (Parte de Aplicação Móvel)
MBMS	<i>Multimedia Broadcast Multicast Service</i> (Serviço Multimídia de Difusão/Multidifusão)
ME	<i>Mobile Equipment</i> (Equipamento Móvel)
MGCF	<i>Media Gateway Control Function</i> (Função de Controle de Entrada de Mídia)
MGW	<i>Media Gateway</i>
MM	<i>Mobility Management</i> (Gerenciamento de Mobilidade)
MRF	<i>Media Resource Functions</i> (Função de Recursos de Mídia)
MSC	<i>Mobile Switching Center</i> (Centro de Comutação Móvel)
NAS	<i>Non Access Stratum</i>
NAT	<i>Network Address Translator</i> (Tradução de Endereço de Rede)
OCS	<i>Online Charging System</i> (Sistema de Cobrança de Tempo Real)
PCCH	<i>Paging Control Channel</i> (Canal de Controle de Paginação)
P-CCPCH	<i>Primary Common Control Physical Channel</i> (Canal Físico Comum de Controle Primário)
PCH	<i>Paging Channel</i> (Canal de Paginação)
P-CSCF	<i>Proxy CSCF</i>

PDCP	<i>Packet Data Convergence Protocol</i> (Protocolo de Convergência de Rede de Pacotes)
PDF	<i>Policy Decision Function</i> (Função de Política de Decisão)
PDN	<i>Packet Data Network</i> (Rede de Pacote de Dados)
PDP	<i>Packet Data Protocol</i> (Protocolo de Dados de Pacotes)
PDU	<i>Protocol Data Unit</i> (Unidade de Dados do Protocolo)
PICH	<i>Paging Indicator Channel</i> (Canal Indicador de Paginação)
PLMN	<i>Public Land Mobile Network</i> (Rede Pública de Telefonia Móvel)
PMM	<i>Packet Mobility Management</i> (Gerenciamento de Mobilidade de Pacotes)
PRACH	<i>Physical Random Access Channel</i> (Canal Físico de Acesso Aleatório)
PS	<i>Packet Switch</i> (Comutação de Pacotes)
PSTN	<i>Public Switch Telephony Network</i> (Rede Pública de Telefonia Comutada)
P-TMSI	<i>Packet-Temporary Mobile Subscriber Identity</i> (Identidade Temporária do Assinante Móvel - Pacote)
QoS	<i>Quality of Service</i> (Qualidade de Serviço)
RA	<i>Routing Area</i> (Área de Roteamento)
RAB	<i>Radio Access Bearer</i> (Portadora de Acesso de Rádio)
RACH	<i>Random Access Channel</i> (Canal de Acesso Aleatório)
RAI	<i>Routing Area Identification</i> (Identificação de Área de Roteamento)
RAN	<i>Radio Access Network</i> (Rede de Acesso de Rádio)
RISC	<i>Reduced Instruction Set Computer</i> (Computador de Conjunto de Instruções Reduzidas)
RLC	<i>Radio Link Control</i> (Controle de Enlace de Rádio)
RLC-AM	<i>Radio Link Control Acknowledged Mode</i> (Controle de Enlace de Rádio – Modo de Reconhecimento)
RLC-UM	<i>Radio Link Control Unacknowledged Mode</i> (Controle de Enlace de Rádio – Modo de Não-Reconhecimento)
RNC	<i>Radio Network Control</i> (Controladores da Rede-Rádio)

RNTI	<i>Radio Network Temporary Identifier</i> (Identificador Temporário da Rede de Rádio)
RRC	<i>Radio Resource Control</i> (Controle de Recurso de Rádio)
RTPC	<i>Rede de Telefonia Pública Comutada</i>
SAP	<i>Service Access Points</i> (Ponto de Acesso a Serviço)
S-CCPCH	<i>Secondary Common Control Physical Channel</i> (Canal Físico Comum de Controle Secundário)
SCH	<i>Synchronization Channel</i> (Canal de Sincronismo)
SCP	<i>Service Control Point</i> (Ponto de Controle de Serviço)
SDK	<i>Software Development Kit</i> (Pacote de Desenvolvimento de <i>Software</i>)
S-CSCF	<i>Server CSCF</i> (Servidor CSCF)
SGSN	<i>Serving GPRS Support Node</i> (Nó de Suporte GPRS de Serviço)
SI	<i>System Information</i> (Informação de Sistema)
SIB	<i>System Information Block</i> (Bloco de Informação de Sistema)
SIP	<i>Session Initiation Protocol</i> (Protocolo de Iniciação de Sessão)
SM	<i>Session Management</i> (Gerenciamento de Sessão)
SMS	<i>Short Message Services</i> (Serviço de Mensagens Curtas)
SRB	<i>Signaling Radio Bearer</i> (Portadora de Sinal de Rádio)
TE	<i>Terminal Equipment</i> (Equipamento do Terminal)
TMSI	<i>Temporary Mobile Subscriber Identity</i> (Identidade Temporária do Assinante Móvel)
UE	<i>User Equipment</i> (Equipamento do Usuário)
UMTS	<i>Universal Mobile Telecommunications System</i> (Sistema de Telecomunicação Móvel Universal)
URA	<i>UTRAN Registration Area</i> (Área de Registro da UTRAN)
U-RNTI	<i>User Radio Network Temporary Identifier</i> (Identificador Temporário da Rede de Rádio- Usuário)
USIM	<i>Universal Subscriber Identity Module</i> (Módulo de Identidade Universal do Assinante)
UTRAN	<i>UMTS Terrestrial Radio Access Network</i>

VoIP (Rede UMTS de Acesso à Radio Terrestre)
Voice over IP
(Voz sobre IP)

WAP *Wireless Application Protocol*
(Protocolo de Aplicação Sem Fio)

SUMÁRIO

1	INTRODUÇÃO.....	1
2	DISPOSITIVOS 3G ALWAYS ON	5
2.1	ALWAYS-ON E CONEXÃO DE DADOS	5
2.1.1	Estabelecimento de conexão e atribuição de endereço de rede.....	5
2.1.2	Conexão dos recursos de rádio para transmissão de dados.....	9
2.1.2.1	Estados do RRC	12
2.1.2.2	Informação de Medição e Troca de Estados do RRC	15
3	CONSUMO DE ENERGIA: PROBLEMAS E SOLUÇÕES	21
3.1	RECEBIMENTO DE PACOTES NÃO SOLICITADOS PELO USUÁRIO	21
3.2	CONSUMO EXTRA GERADO POR PACOTES DE SINALIZAÇÃO.....	26
3.3	TEMPO DE INATIVIDADE DO RRC.....	28
3.4	REDUÇÃO DINÂMICA DO TEMPO DE INATIVIDADE DO RRC	32
4	MODELO DE PREVISÃO	35
4.1	FILTRO DE KALMAN.....	35
4.2	SUBFILTRO DE KALMAN.....	38
5	CARACTERIZAÇÃO DE TRÁFEGO DA REDE REAL	43
5.1	DISTRIBUIÇÕES DE PROBABILIDADE	43
5.1.1	Distribuição Exponencial.....	43
5.1.2	Distribuição de Poisson.....	45
5.1.3	Distribuição Lognormal	46
5.1.4	Distribuição Weibull	47
5.1.5	Distribuição de Pareto	48
5.1.6	Distribuição Gamma	50
5.1.7	Distribuição Beta.....	51
5.1.8	Distribuição de Erlang.....	52
5.2	AJUSTE DE DISTRIBUIÇÃO E GERAÇÃO DE VARIÁVEL ALEATÓRIA	54
5.3	TIPOS DE TRÁFEGO.....	56
5.4	MONITORAMENTO DO TRÁFEGO REAL DE PACOTES NÃO SOLICITADOS E E-MAIL	58
6	MODELOS DE SIMULAÇÃO.....	63
6.1	MODELO DE SIMULAÇÃO I – FUNCIONAMENTO DA MÁQUINA DE ESTADOS DO RRC DE ACORDO COM A REDE REAL	63
6.1.1	Validação e resultados das simulações	65
6.2	MODELO DE SIMULAÇÃO II – FUNCIONAMENTO DA MÁQUINA DE ESTADOS DE ACORDO COM A REDE REAL INCLUINDO MODELO DE PREVISÃO	67
6.2.1	Modelo II-A - Filtro de Kalman	67
6.2.1.1	Simulações e Resultados.....	69
6.2.2	Modelo II-B - Modelo Completo.....	71
6.2.2.1	Simulações e Resultados.....	74
6.3	IMPLEMENTAÇÃO DO FILTRO NOS DISPOSITIVOS MÓVEIS.....	80
7	CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS	85
	REFERÊNCIAS	87

APÊNDICE A – ARQUITETURA DE REDE DA TERCEIRA GERAÇÃO CELULAR	91
A.1. Arquitetura do Sistema UMTS	91
A.1.1. Interface de Protocolos de Rádio	97
A.1.2. Canais UMTS	98

1 INTRODUÇÃO

O consumo de bateria dos celulares é um problema que acompanha os fabricantes de telefonia desde o surgimento dos primeiros aparelhos móveis. Com sua evolução, os dispositivos tornaram-se cada vez menores, fazendo com que a dimensão física da bateria também diminuísse, e conseqüentemente sua capacidade.

O número de funcionalidades e aplicativos existentes para esses dispositivos também aumentaram. Essas por sua vez, também são responsáveis por elevarem o consumo de bateria do telefone. Atualmente, os dispositivos móveis estão passando por uma evolução constante e aumentando seu poder de processamento para rodar sistemas operacionais mais complexos como Linux (Android) e *Windows Mobile*. Nesses sistemas é possível executar inúmeras aplicações diferentes, fazendo com que o dispositivo móvel seja semelhante a um computador portátil de tamanho reduzido, porém com menor capacidade de carga.

A bateria dos dispositivos móveis sofreu modificações não apenas de tamanho e capacidade, mas também no tipo de material que é constituída, que passaram das baterias de NiCd (*Níquel-Cádmio*) e NiMH (*Níquel Metal Hydrid*), consideravelmente grandes, utilizadas no início dos anos 80 e meados dos anos 90, até as de Li-Ion (*Lithium-Ion*), que possuem memória e são menores a um custo de menor carga fornecida e, atualmente evoluindo para as baterias de Li-Poly (*Lithium-Ion Polymer*), que prometem uma alta capacidade em um menor espaço físico. Nenhuma delas, porém, é capaz de se comparar a uma bateria de um computador portátil, ou tem capacidade suficiente para manter o celular ligado por vários dias, quando o dispositivo móvel está utilizando o máximo de aplicações e recursos de rede disponíveis.

Diversas tentativas para diminuição de consumo têm sido feitas a partir de alterações em várias partes do dispositivo. A mais comum é a mudança de hardware, com a substituição de componentes que consomem menos energia. Os processadores ARM¹, que são tipicamente RISC (*Reduced Instruction Set Computer*), tem a arquitetura desenvolvida para possibilitar implementações enxutas sem abrir mão do desempenho, devido à simplicidade dos processadores. Isso acaba acarretando em um baixo consumo de energia, o que torna esses processadores

¹ Advanced RISC Machines – Processador RISC de 32 bits baseado nas tecnologias dos processadores MOS Technology 6502 e Berkeley RISC1, desenvolvido atualmente pela ARM Ltda.

interessantes para aplicações móveis. Outros exemplos são a evolução do *Bluetooth* 4.0 e soluções de conexão sem fio mais eficientes em relação ao consumo; e telas de LED (*Light-Emitting Diode*) ou AMOLED (*Active-Matrix Organic Light-Emitting Diode*) em substituição as telas de LCD (*Liquid Crystal Display*).

Apesar desses esforços apresentarem um aumento do tempo de utilização do aparelho e redução significativa do consumo de energia, existem casos que a diminuição de consumo se torna mais complexa. Esse caso está relacionado à tecnologia de rádio empregada para transmissão de dados do dispositivo. As tecnologias de segunda (2G) e terceira geração (3G), GSM (*Global System for Mobile Communication*) e UMTS (*Universal Mobile Terrestrial System*), respectivamente, possuem normas que regem esses padrões e especificam o nível de potência adequado de transmissão e recepção de dados. A conexão com os canais de rádio também é responsável pelo consumo e está presente durante todo tempo que o telefone permanece ligado. Nesses casos, qualquer método para diminuição do consumo se torna mais complicado, pois pode envolver mudanças em normas internacionais. Deve-se então procurar um meio mais eficiente do dispositivo interagir com o sistema para otimizar o consumo sem que seja necessário alterar os padrões ou equipamentos já existentes, o que o torna um grande desafio.

A expansão das redes de terceira geração celular ao redor do mundo e sua banda larga para transmissão de dados a uma alta velocidade fizeram com que novos serviços pudessem ser comercializados. O oferecimento de serviços de planos de dados para acesso a Internet pelo dispositivo móvel ou através de equipamentos denominados mini modem 3G, fazem com que a rede da operadora de telefonia móvel seja integrada a outras redes para comunicação de dados.

Inicialmente projetada com pouca ou nenhuma preocupação com segurança dos dados trafegados, as redes das operadoras de telefonia móvel ficam bastante suscetíveis a ataques de hacker [LEE-WOO, 2007]. Desse modo, um dispositivo móvel de um usuário conectado na rede de dados com um endereço IP (*Internet Protocol*) válido está sujeito a receber um ataque tanto quanto um computador conectado à Internet, por exemplo. Os computadores, no entanto, podem se proteger dos ataques com a utilização de antivírus e *Firewall*. Um dispositivo móvel por sua vez não possui esses mecanismos. Outra diferença está na infra-estrutura dos provedores de Internet, que foi projetada prevendo esse tipo de ataque, e se preveniram com a instalação de equipamentos de rede para bloquear o tráfego de pacotes maliciosos. Esse fato não foi levado em conta nas primeiras gerações de rede celular, e está ganhando foco com o surgimento das redes de

terceira geração e aumento do uso da rede de dados das operadoras devido a maior largura de banda disponível.

Estes pacotes maliciosos, ou não solicitados como apresentado em [RICCIATO, 2006], podem degradar o desempenho da rede e do dispositivo, de forma que a comunicação da rede fique impraticável devido à falta de recursos. Os tipos mais comuns de ataque são do tipo DoS (*Denial of Service*), como apresentado em [LEE-WOO, 2007]. A alocação de recursos da interface aérea para recebimento desses pacotes pelo dispositivo faz com que o dispositivo tenha um aumento no consumo de energia, diminuindo a vida útil da bateria, como pode ser visto em [PERÄLÄ, 2009] e [HAVERINEN, 2007]. Como apresentado em [LEE, 2009], o impacto do recebimento desses pacotes em uma rede de terceira geração é maior do que nas redes de segunda geração devido às características dos canais de interface aérea.

Tendo em vista os problemas apontados anteriormente, o consumo de bateria pode ser alto sem que nenhuma aplicação esteja sendo executada. Isso ocorre principalmente devido ao ataque de usuários mal intencionados, no caso da rede de dados não estar protegida com *Firewall*. O consumo pode ser ainda maior caso aplicações como que trafeguem dados, como *e-mail*, por exemplo, estejam sendo executadas em segundo plano. Em [ERONEN, 2008], e [HAVERINEN, 2007], os autores apontam o alto fluxo de mensagens de sinalização quando se utiliza NAT (*Network Address Translator*) e *push e-mail*² como um dos principais fatores para o aumento do consumo de energia dos dispositivos móveis. No entanto, outro fator de maior relevância para o consumo de energia e que está presente em todo o tipo de tráfego recebido pelo dispositivo móvel são os altos valores dos tempos de inatividade configurados para desalocação dos canais do recurso de rádio da interface aérea, como é relatado em [ARJONA, 2007], [YEH, 2004] e também apresentado em [HAVERINEN, 2007]. Em [LEE-CHEN, 2004], é apresentado um método para o cálculo do tempo ideal em que o dispositivo móvel deve permanecer com os recursos de rádio ativos, sem que haja degradação do desempenho na recepção da informação e, com isso, diminuir o consumo de energia.

Em [PENG, 2008], é proposto um método adaptativo de redução de consumo levando em consideração as características de tráfego do usuário em três condições: alto, médio e baixo

² Notação usada para descrever sistemas de e-mail que provêm a transferência de e-mail ao usuário assim que recebido pelo seus servidores, sem que haja necessidade de espera do cliente realizar uma busca no servidor por novos e-mails.

tráfego. O método proposto considerou um tráfego elástico (tratado com mais detalhes no Capítulo 5) sem distinção de serviços, utilizando um processo de chegada de Poisson.

A proposta deste trabalho é a utilização de um modelo de previsão para prever a ocorrência de chegada de um novo pacote ao dispositivo. Uma vez que o padrão 3GPP (*Third Generation Partnership Project*) permite que a desconexão do recurso de rádio seja originada pelo dispositivo, a lógica de controle de tráfego será implementada no dispositivo móvel, e a decisão de desconexão será tomada por ele e não pela rede. Com isso, tem-se um mecanismo de desconexão diferenciado para cada usuário, de acordo com o tráfego recebido. O tempo de inatividade deixa de ser um valor fixo, como é observado atualmente, e passa a ser um valor dinâmico melhor ajustado para cada usuário, pois não há necessidade de se manter recursos alocados quando for previsto que não haverá recebimento de pacote de dados pelo dispositivo. Em [LIERS, 2006] é proposto um modelo dinâmico para os tempos de inatividade para desconexão dos recursos de rádio. No entanto, a troca de estados não é decidida por meio da previsão de tráfego, mas sim por meio da capacidade de recursos da rede, onde recursos não utilizados são liberados caso outro usuário necessite utilizá-los. Esse controle, no entanto, é feito do lado do núcleo da rede e não do dispositivo móvel, e não há preocupação com o consumo de energia do dispositivo, mas com a eficiência de alocação de recursos por parte da rede.

No Capítulo 2 são apresentados o conceito *Always On* e o mecanismo de conexão de dados e troca de pacotes. Uma leitura mais detalhada sobre a arquitetura das redes de terceira geração para leitura complementar a este capítulo encontra-se no Apêndice A. Os problemas relacionados ao consumo de energia no estado *Always On* e as soluções para sua diminuição, entre elas o método proposto neste trabalho, são apresentados no capítulo 3.

No Capítulo 4 é apresentado o modelo de previsão baseado em filtro de Kalman utilizado neste estudo. As funções de distribuição de probabilidade comumente utilizadas para geração de dados e a caracterização de tráfego de rede feita pelo do monitoramento dos pacotes no dispositivo móvel e o ajuste de distribuição são tratados no Capítulo 5.

No Capítulo 6 são apresentados dois modelos: um simulando uma rede real quanto ao recebimento de pacotes, transição de estados dos recursos de rádio e seus tempos de inatividade, e outro modelo baseado na rede real, porém com o método de decisão de desconexão do recurso de rádio baseado em um modelo de previsão. O algoritmo de decisão do filtro de Kalman e sua implementação no software Arena também são apresentados com mais detalhes nesse capítulo.

2 DISPOSITIVOS 3G ALWAYS ON

2.1 ALWAYS-ON E CONEXÃO DE DADOS

Aplicações como *e-mail* que utilizam *push e-mail* e VoIP (*Voice over Internet Protocol*) são exemplos de aplicações que necessitam que o dispositivo seja alcançado a qualquer momento. Para isso, o móvel deve permanecer conectado a rede de dados com um endereço válido durante todo o tempo para que as solicitações feitas a essas aplicações possam alcançar o destino com sucesso.

Always On é o termo utilizado para um dispositivo que fica 100% do tempo conectado a rede de dados e possui um endereço válido para comunicação de dados.

O crescimento das redes 3G foi responsável pelo crescimento do uso deste conceito nos dispositivos, pois dispõem de uma largura de banda maior que as tecnologias anteriores. Isso possibilita que vários usuários estejam conectados ao mesmo tempo e trocando dados a uma taxa máxima de 28 Mbps no *downlink* (Canal Direto) e 11.5 Mbps no *uplink* (Canal Reverso) se for considerado um caso ideal para uma rede que suporte a tecnologia de rede HSPA+ (*High Speed Packet Access Plus - Release 7*), sucessora do 3G e predecessor da tecnologia de quarta geração (4G).

Neste capítulo serão abordados os procedimentos para estabelecimento de conexão e troca de pacotes de dados em uma rede 3G. Informações detalhadas sobre a arquitetura da rede de terceira geração encontram-se no Apêndice A.

2.1.1 Estabelecimento de conexão e atribuição de endereço de rede

O estabelecimento de conexão na rede comutada de pacotes (*Packet Switch - PS*) nas redes 3G é realizado como mostra a Figura 2.1. Ela tem por finalidade o registro na rede e a atribuição de um endereço válido para o móvel, para que seja possível o registro e interação no IMS (*IP Multimedia Subsystem*).

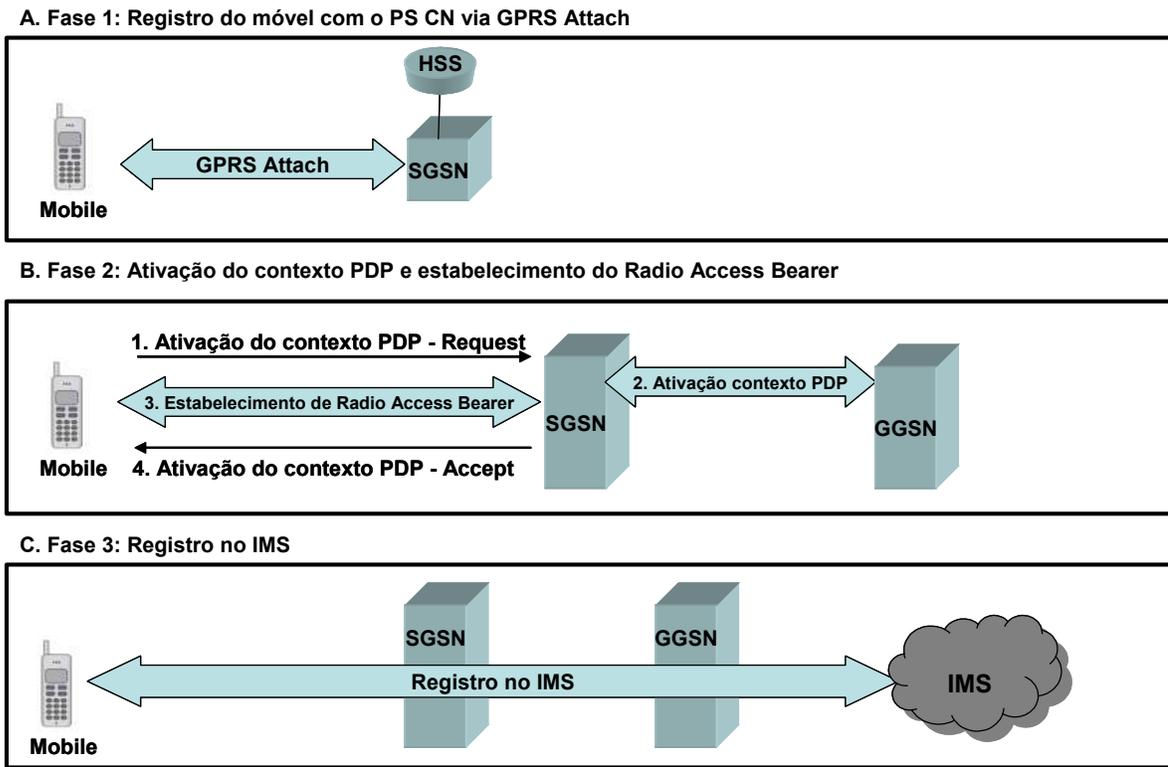


Figura 2.1. Estabelecimento de conexão de dados. [CHEN, 2004]

Para atribuição de um endereço válido para transmissão e recepção de pacotes da rede, como um endereço IP (*Internet Protocol*), que é o mais utilizado, por exemplo, é necessário a criação de dois contextos:

- Contexto PMM (*Packet Mobility Management*);
- Contexto PDP (*Packet Data Protocol*).

Um contexto PMM em um dispositivo móvel é um conjunto de informações utilizadas pela rede para rastrear a localização do móvel. O estado de um dispositivo no contexto PMM determina quais conexões de redes (portadoras) entre o dispositivo e a rede devem ser mantidas e como a rede deve localizá-los. Os SGSNs (*Serving GPRS Support Node*) são responsáveis por rastrear a localização dos móveis que estão utilizando pacotes de dados.

A máquina de estados do contexto PMM é apresentada na Figura 2.2, e possui três estados:

- **PMM-*Detached***: Neste estado não há comunicação, localização válida ou informação de rota entre o dispositivo e o SGSN. O dispositivo não reage às informações do sistema relacionadas ao SGSN que não pode alcançar o móvel.

- **PMM-Connected:** Neste estado, o SGSN e o dispositivo têm um contexto PMM estabelecido e uma sinalização de conexão dedicada é estabelecida entre eles. A localização do móvel dentro do RAN (*Radio Access Network*) é monitorada pelo RNC (*Radio Network Controller*) em um nível das células de rádio. No CN (*Core Network*) do PS, o SGSN monitora a localização do móvel pelo RNC que está servindo o móvel. Neste estado, o contexto PDP pode ou não ser ativado. Antes de ele ser ativado, o móvel não estará habilitado a enviar ou receber pacotes da rede.

- **PMM-Idle:** O SGSN e o móvel têm os contextos PMM estabelecidos e a localização do dispositivo é monitorada pelo SGSN por meio da informação de *Routing Area*. O móvel é alcançado pelo *Core Network* via informação de *Paging*. Nenhuma sinalização ou conexão de tráfego existe entre o dispositivo e o SGSN. O dispositivo entra nesse estado para conservar recursos como largura de banda.

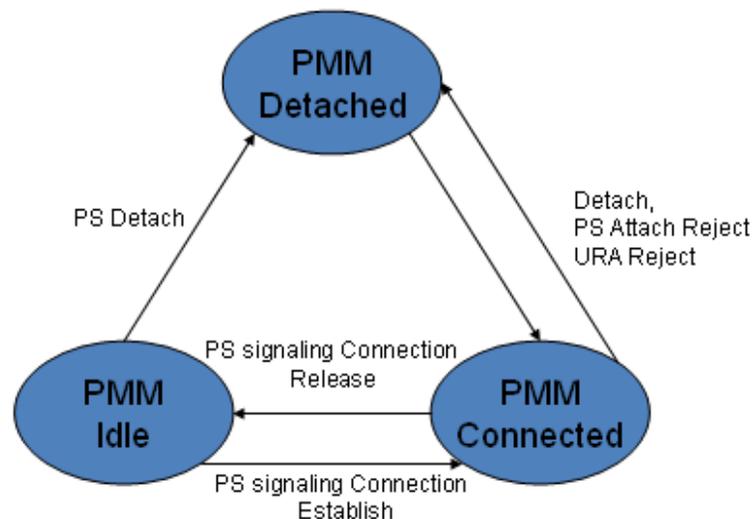


Figura 2.2. Estados do contexto PMM.

Para enviar e receber pacotes, o móvel deve estabelecer um contexto PDP (*Packet Data Protocol*) para adquirir e configurar um endereço PDP no domínio do PS do *Core Network* e no móvel, que pode ser um endereço IP ou X.25, por exemplo. Um dispositivo também pode utilizar múltiplos endereços PDP simultaneamente. Um contexto PDP mantido por um nó da rede contém um conjunto de informações que a rede utiliza para determinar como distribuir os pacotes de origem e destino de um usuário de um endereço PDP.

O contexto PDP mantido no SGSN e no GGSN contém as seguintes informações:

- **Endereço PDP:** usado pelo móvel para enviar e receber pacotes PDP.
- **Routing Information:** Informação usada pela rede para determinar onde encaminhar o pacote do usuário. Esta informação inclui informações dos túneis estabelecidos entre SGSN e GGSN para este contexto PDP e o *Access Point Name* (APN).
- **Perfis de Qualidade de Serviço (Quality of Service - QoS):** Existem três categorias:
 - Subscribed:** Descreve características de QoS subscritas pelo usuário. A rede não irá garantir QoS maior do que o usuário tenha subscrito. Os parâmetros de QoS subscritos estão armazenados no HLR e passados para o SGSN;
 - Requested:** Descreve QoS requisitada atualmente pelo usuário;
 - Negotiated:** Descreve o QoS provido no momento pela rede ao móvel.

A máquina de estados do contexto PDP é apresentada na Figura 2.3 e os estados possíveis para o contexto são:

- **PDP CONTEXT ACTIVE:** Um contexto PDP no estado ACTIVE contém informações atualizadas para encaminhamento de pacotes entre o móvel e o GGSN. No entanto, o fato de o PDP estar no estado ACTIVE não significa realmente que os *Radio Access Bearer* (RABs) requisitados para transportar pacotes sobre o RAN estão estabelecidos. Ao contrário, o RAB pode ser estabelecido apenas quando o móvel tem pacotes para enviar para rede ou a rede tem pacotes a serem enviados.
- **PDP CONTEXT INACTIVE:** Um contexto PDP no estado INACTIVE pode conter um endereço PDP válido, mas não irá conter roteamento válido e mapeamento de informação necessária para determinar como processar os pacotes PDP. Nenhum dado pode ser transferido entre o móvel e a rede. Uma mudança de localização de um móvel não irá causar uma atualização para o contexto PDP.

Se um GGSN tem pacotes a enviar ao móvel, mas o contexto PDP para o endereço de destino estiver no estado INACTIVE, o GGSN pode utilizar o procedimento de *PDP Context Activation Request* para mudar o contexto PDP para o móvel de destino para o estado ACTIVE. O GGSN pode também descartar pacotes destinados ao móvel se o contexto PDP correspondente estiver no estado INACTIVE.

As ações que podem ser realizadas por um contexto PDP para transição de estados entre ACTIVE e INACTIVE são:

- **PDP Context Activation:** Cria e ativa um contexto PDP, movendo-o do estado INACTIVE para ACTIVE. O móvel ou o GGSN pode iniciar o procedimento de PDP Context Activation, mas o GGSN pode apenas iniciar o PDP Context Activation sobre algumas limitações.
- **PDP Context Modification:** Muda as características de um PDP ativo, como endereço do PDP ou atributos de QoS para ser suportado pela rede.
- **PDP Context Deactivation:** Remove um Contexto PDP existente. Esse procedimento move um contexto PDP de ACTIVE para INACTIVE. O estado de um contexto PDP pode também ser transicionado quando o PMM muda seu estado de IDLE para DETACHED.

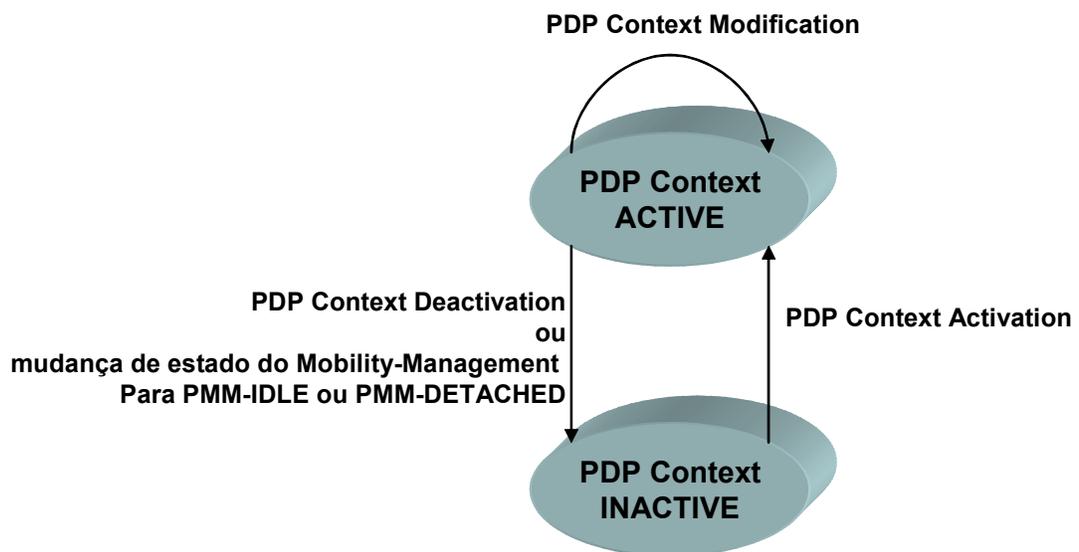


Figura 2.3. Transição de estados do contexto PDP. [CHEN, 2004]

2.1.2 Conexão dos recursos de rádio para transmissão de dados

O estabelecimento do canal de rádio, no entanto, é feito de acordo com a necessidade de envio e recebimento dos pacotes de dados. Eles não são alocados 100% do tempo apesar do dispositivo estar registrado na rede de comutação de pacotes, e sim por demanda.

A conexão de recursos de rádio é realizado pelo RRC (*Radio Resource Control*), que é o protocolo da camada 3 que oferece serviços como configuração, troca de conexão e desconexão,

bem como sinalização e controla os protocolos de camada mais baixa para prover esses serviços, e também oferece serviços para camadas superiores do NAS (*Non Access Stratum*). Todas as camadas de sinalização (*Mobility Management - MM, Call Control - CC, Session Management - SM*) são encapsuladas no RRC.

Para uma conexão iniciada do lado do UE (*User Equipment*), o estabelecimento de uma conexão RRC e SRB (*Signaling Radio Bearer*) entre o UE e a UTRAN (*UMTS Terrestrial Access Network*) é iniciada por uma requisição das camadas altas do NAS. A requisição do NAS é uma requisição para configuração de uma *Signaling Connection* entre a UE e o *Core Network*. Apenas se o UE está no modo *Idle*, ou seja, sem conexão com o RRC, o UE inicia um procedimento de estabelecimento de conexão com o RRC. Pode existir uma ou nenhuma conexão com o RRC entre a UE e a UTRAN.

Para uma conexão iniciada pela rede, o estabelecimento é precedido de uma mensagem de *RRC paging*.

A manutenção da conexão RRC refere ao funcionamento do re-estabelecimento de conexão RRC, que pode ser usado para restabelecer uma conexão após uma falha de link de rádio.

O procedimento de estabelecimento de conexão do RRC iniciada pelo móvel é apresentado na Figura 2.4.

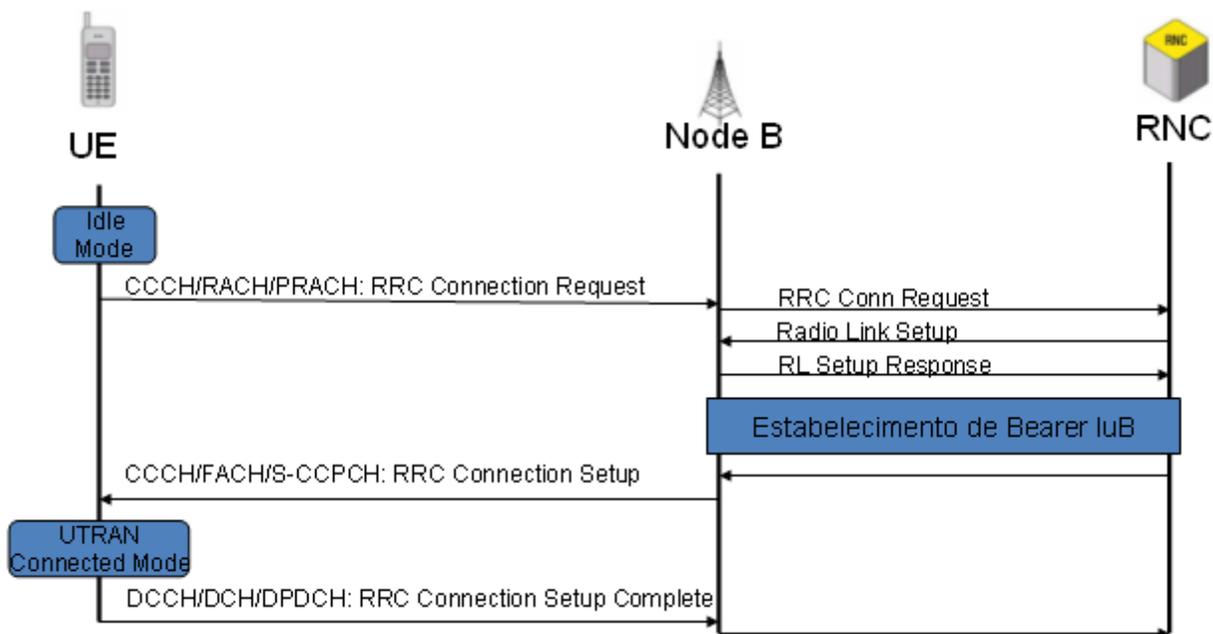


Figura 2.4. Procedimento de estabelecimento de conexão no RRC. [HOLMA, 2010]

A configuração da conexão (*Connection Setup*) pode incluir uma atribuição de um canal físico dedicado para o UE (estado CELL_DCH) ou pode comandar a UE para utilizar os canais comuns (CELL_FACH), que serão detalhados ao decorrer deste capítulo.

As funções do protocolo RRC consistem em:

- Difusão da informação relacionada ao NAS do *Core Network* por meio das mensagens de *System Informations* (SI) que incluem informações de células específicas e informações comuns de todas as células e ao *Access Stratum* (AS) para todos os UEs da rede.
- Estabelecimento, alteração e desconexão do *Radio Bearer*, requisitado pelas camadas mais altas e realizado no *U-Plane* (*User Plane*).
- Alocação, alteração e desconexão do RRC quando requisitado pela própria conexão do RRC.
- Gerenciamento de mobilidade da conexão do RRC, que incluem funções como preparação de *handover* baseado nos resultados de medição executado pelo UE, *handover* para GSM e outros sistemas de rádio, re-seleção de células e procedimentos para atualização de *paging*.
- Estabelecimento, re-estabelecimento, manutenção e desconexão da conexão do RRC entre o UE e a UTRAN. O estabelecimento é ativado das camadas altas do lado do UE para estabelecer a sinalização da conexão. O controle da conexão inclui a pré-seleção de células adicionais, controle de admissão e estabelecimento de *links* de sinalização da camada 2. A desconexão é feita quando as camadas mais altas requisitam a desconexão ou quando ocorrem falhas. Neste último caso, o UE deve requisitar seu re-estabelecimento.
- Funções de *paging* e notificações onde o RRC difunde informações de *paging* para um UE específico em resposta a uma requisição das camadas mais altas do lado da rede. Também faz o *paging* para o UE com o qual a conexão do RRC é estabelecida, como mostra a Figura 2.5. Pode ser utilizado para os seguintes propósitos [HOLMA, 2010]:
 - Em uma chamada originada pelo *Core Network* ou configuração de sessão. Nesse caso a requisição vem do *Core Network* pela Iu.

- Para mudar o estado do UE de CELL_PCH ou URA_PCH para CELL_FACH. Pode ser inicializado pela atividade de dados do *downlink*.
- Para indicar mudanças de informações do sistema. Este tipo de *paging* é direcionado a todos os UEs da célula.
- Controle de qualidade de serviço da portadora e alocações de recurso de rádio e controle de potência.
- Informações de medição: *Measurement Control* e *Measurement Reporting* para o UE.

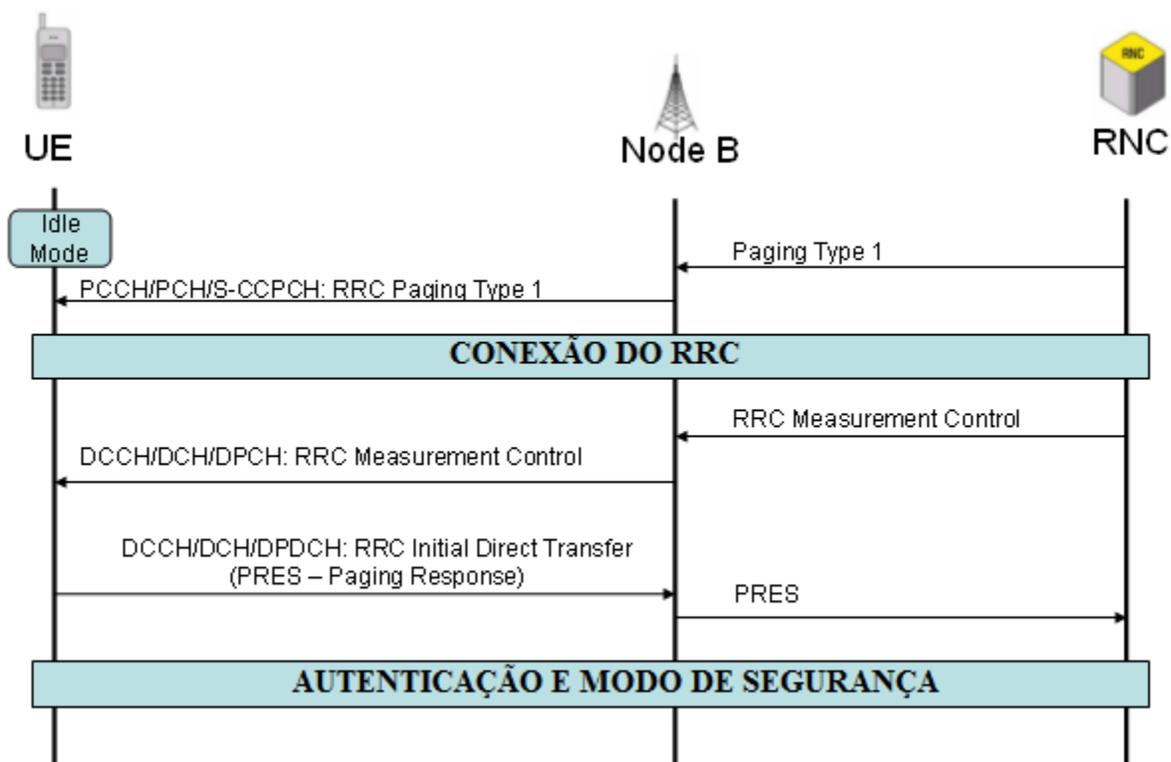


Figura 2.5. Conexão com o RRC para envio de *paging*. [HOLMA, 2010]

2.1.2.1 Estados do RRC

A seguir são apresentados os modos e estados do RRC:

- **IDLE:** O modo *Idle* compreende o estado em que o UE mantém informações limitadas ao *Access Stratum* e a UTRAN não tem informação do UE. A identificação do móvel pelo

NAS é feita pelo IMSI (*International Mobile Subscriber Identity*), TMSI (*Temporary Mobile Subscriber Identity*) e P-TMSI (*Packet TMSI*). O estabelecimento da conexão com o RRC só será efetuado mediante a uma resposta a uma requisição do UE ou uma requisição de *paging* efetuada pela rede. Quando recebida a mensagem de confirmação de conexão com o RRC pela UTRAN, o UE vai para o modo *UTRAN-connected*.

- **UTRAN-Connected:** Após a conexão de o RRC ser confirmado, o móvel vai para o estado CELL_FACH ou CELL_DCH que fazem parte do modo *UTRAN-connected*. Um RNTI (*Radio Network Temporary Identity*) é atribuído ao modo, o qual é utilizado para identificar o UE no CTCH (*Common Traffic Channel*). A máquina de estado do RRC vai depender do nível do canal de transporte que poderá ser utilizado pelo UE. Quando está no modo *UTRAN-connected*, pode variar nos seguintes estados: CELL_PCH, CELL_FACH, CELL_DCH e URA_PCH.

- **CELL_DCH (*Dedicated Channel*):** Estado o qual é alocado um canal dedicado ao terminal. Este estado é utilizado para transmissão de dados quando a quantidade de dados transferida é alta, garantindo uma máxima vazão e mínimo atraso a um custo de alto consumo de energia do terminal (superior a 200 mA). Este canal é utilizado tanto para o *uplink* quanto para o *downlink*.
- **CELL_FACH (*Forward Access Channel*):** Canal comum utilizado quando há baixo volume de dados a ser transmitido, ou o fluxo de dados ultrapassar um limiar configurado pelo *Core Network*. No *downlink* a transmissão fica por conta do canal FACH (*Forward Access Channel*) e no *uplink* por um canal comum ou compartilhado como o RACH (*Random Access Channel*) ou CPCH (*Common Package Channel*). Este último consome menos energia e utiliza um espectro mais eficiente para quantidades pequenas de dados. O consumo de energia do UE neste caso chega à metade do estado anterior (em torno de 100 mA).
- **CELL_PCH (*Paging Channel*):** Estado opcional no qual o telefone pode receber um sinal de *paging* da rede se existir pacotes provenientes do *downlink* e então entrar no estado CELL_DCH ou CELL_FACH para iniciar a transição de pacotes, mas enquanto estiver em CELL_PCH não está habilitado a receber ou enviar pacotes. Algumas literaturas utilizam a nomenclatura PCH_IDLE

para este estado. O consumo de energia neste estado é o mais baixo, entre 1% e 2% do valor do CELL_DCH. Neste estado o UE tem rastreabilidade em nível de célula. Neste estado o UE também é capaz de receber mensagens do BMC (*Broadcast/Multicast Control*). No caso de re-seleção de célula o móvel pode ir para CELL_FACH e retornar para CELL_PCH se nenhuma atividade for iniciada durante o *Cell Update*.

- **URA_PCH (UTRAN Registration Area - PCH)**: Semelhante ao estado CELL_PCH, no entanto o rastreamento do UE tem visibilidade pela UTRAN. Algumas literaturas utilizam a nomenclatura IDLE_URA.

Em ambos estados CELL_PCH e URA_PCH, como não há transmissão no *uplink*, há a possibilidade de economia de energia. Há ainda a opção de se utilizar *Discontinuous Reception* (DRX) que é suportado apenas por estes dois estados e, quando habilitado, o UE monitora um curto período do ciclo de *Paging Indicator Channel* (PICH) no canal físico. O contexto PDP continua ativo e, caso haja necessidade de transmissão de mensagens para rede, a sessão pode ser restabelecida rapidamente.

A Figura 2.6 ilustra os estados e as trocas possíveis entre os estados do RRC.

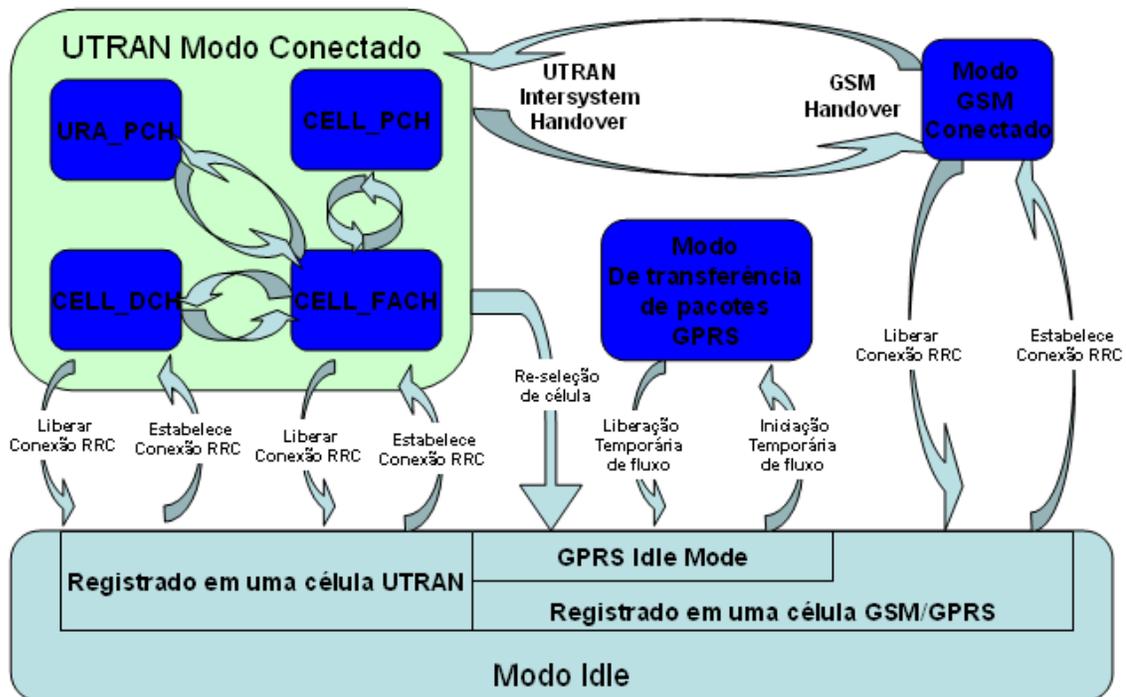


Figura 2.6. Máquina de estado para transmissão de dados no RRC. [TACHIKAWA, 2002]

2.1.2.2 Informação de Medição e Troca de Estados do RRC

Para que ocorra a troca de estados durante a transmissão e envio de pacotes, o UE deve monitorar o fluxo de dados, e enviar um relatório de medição de tráfego para rede, que tomará a decisão de trocar de estado ou não.

As mensagens que estão relacionadas com a medição de tráfego e troca de estados são:

- *Measurement Control*

- *Measurement Report*

A mensagem de *Measurement Control* indica o tipo de medição a ser feita. O RNC pode iniciar, parar ou modificar um número de medições paralelas no UE e cada uma das medições pode ser controlada independentemente. A informação de controle de medição é incluída no SIB (*System Information Block*) tipo 12 e SIB tipo 11, quando o telefone estiver em Idle ou quando não estiver utilizando o canal dedicado. Se o UE estiver no estado CEL_DCH, um *Measurement Control* dedicado pode ser utilizado, como mostra a Figura 2.7. Nessa mensagem, o parâmetro *Measurement Type* pode indicar diferentes tipos de medição, entre elas o *Traffic Volume Measurement* que indica os parâmetros de medição de volume de tráfego. Essa mensagem, por sua vez, contém os parâmetros *Traffic Volume Event Identity* (evento) e o *Reporting Threshold* (indica o valor do limiar de tráfego para o evento indicado).

Na Tabela 2.1 podem ser visualizados os valores possíveis para os limiares de tráfego.

TABELA 2.1. VALORES POSSÍVEIS PARA O ELEMENTO *REPORTING THRESHOLD*. [TS25331]

Informação Elemento/Nome do Grupo	Necessidade	Multi	Tipo e Referência	Descrições Semânticas
>> <i>Reporting Threshold</i>	MP (Parâmetro Obrigatório)	-	Enumerado(8, 16, 32, 64, 128, 256, 512, 1024, 2K, 3K, 4K, 6K, 8K, 12K, 16K, 24K, 32K, 48K, 64K, 96K, 128K, 192K, 256K, 384K, 512K, 768K)	Limiar em Bytes sendo N Kbytes = N*1024 bytes

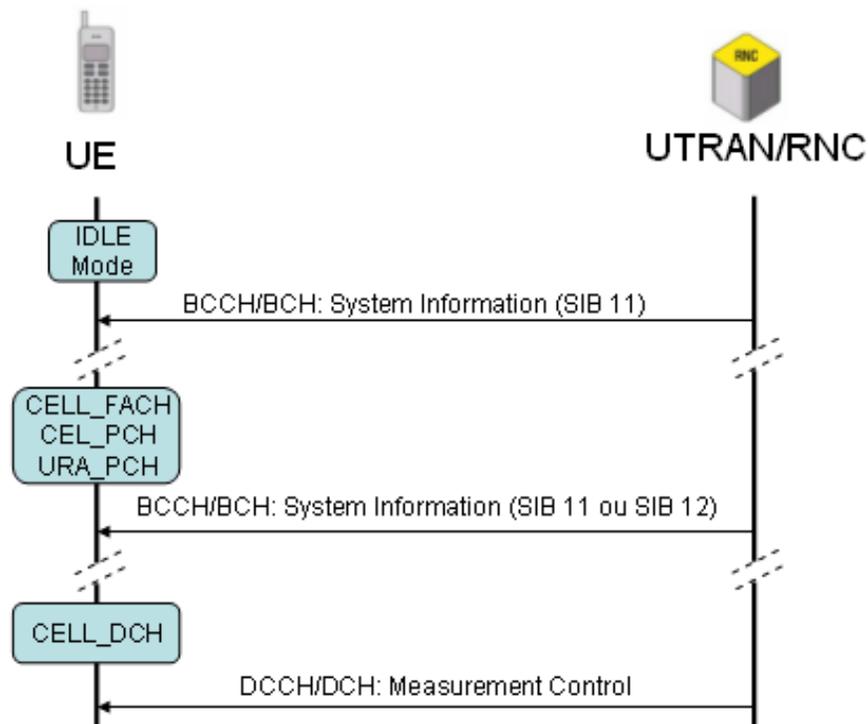


Figura 2.7. Procedimento de *Measurement Control*. [HOLMA, 2010]

A mensagem de *Measurement Report* é utilizada para enviar os resultados das medições requisitadas pela rede por meio das mensagens de *Measurement Control*. Ela é enviada pelo UE quando determinados critérios são satisfeitos. O UE envia a mensagem de *Measurement Report* com o número da identificação da medição e os resultados da medição, que podem ser:

Evento 4a: Volume de tráfego do canal excedeu o limiar

Evento 4b: Volume de tráfego do canal ficou abaixo do limiar

A Figura 2.8 ilustra o procedimento de monitoramento de volume de tráfego. As informações necessárias para se realizar as medições de tráfego são recebidas pelo móvel por meio da mensagem de *Measurement Control*. Essas informações são então enviadas à camada MAC (*Medium Access Control*) do móvel. Existem duas maneiras de o MAC indicar o volume de tráfego para o RRC: eventos periódicos e eventos disparados. Se o tipo de medição for periódico, a camada MAC reporta os resultados da medição periodicamente. Caso seja disparado por evento, a camada MAC só irá avisar a camada RRC quando o resultado estiver próximo do limiar descrito no parâmetro da mensagem *Reporting Threshold*. As medições periódicas e disparadas por evento podem ocorrer simultaneamente como mostra a Figura 2.8.

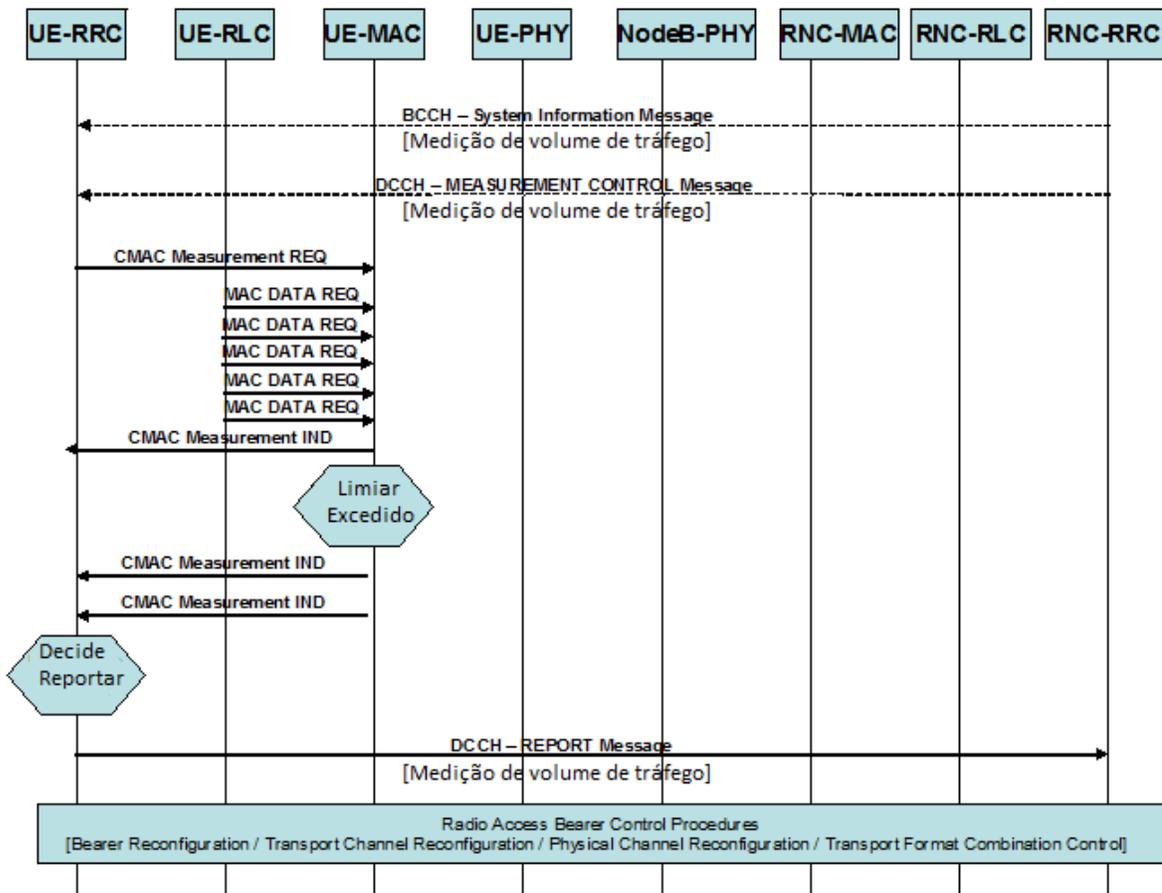


Figura 2.8. Monitoramento de tráfego via canais de medição [TS25303]

Feito isto, baseado nas medições da camada MAC e nos critérios recebidos pela UTRAN, a camada RRC do móvel decide quando deve mandar a mensagem de *Measurement Report* a UTRAN. Quando a mensagem é recebida, a UTRAN toma a ação apropriada baseada no resultado da medição enviada pelo móvel, que pode ser:

- Reconfiguração de portadora
- Reconfiguração de canal de transporte
- Reconfiguração do canal físico

O critério para transição de estados é o nível de ocupação de *Buffer* (*Buffer Occupation - BO*) no terminal do usuário que armazena os dados a serem enviados. A medição de ocupação do *Buffer* é enviada do UE para a rede que toma a decisão de mudar ou não de estado. A rede então envia uma mensagem *RADIO BEARER RECONFIGURATION*. Nessa mensagem está presente

o campo *RRC State Indicator*, como mostra a Tabela 2.2, informando qual é o novo estado o qual o dispositivo móvel deverá fazer a transição.

A transição de Cell_DCH para Cell_FACH e vice-versa é dada quando determinados limiares de fluxo de dados sejam excedidos. Caso BO atinja um valor maior que esse limiar, ocorre a transição de Cell_FACH para Cell_DCH, ou caso esteja menor que o limiar, deverá ocorrer a transição de Cell_DCH para Cell_FACH, como ilustra a Figura 2.9.

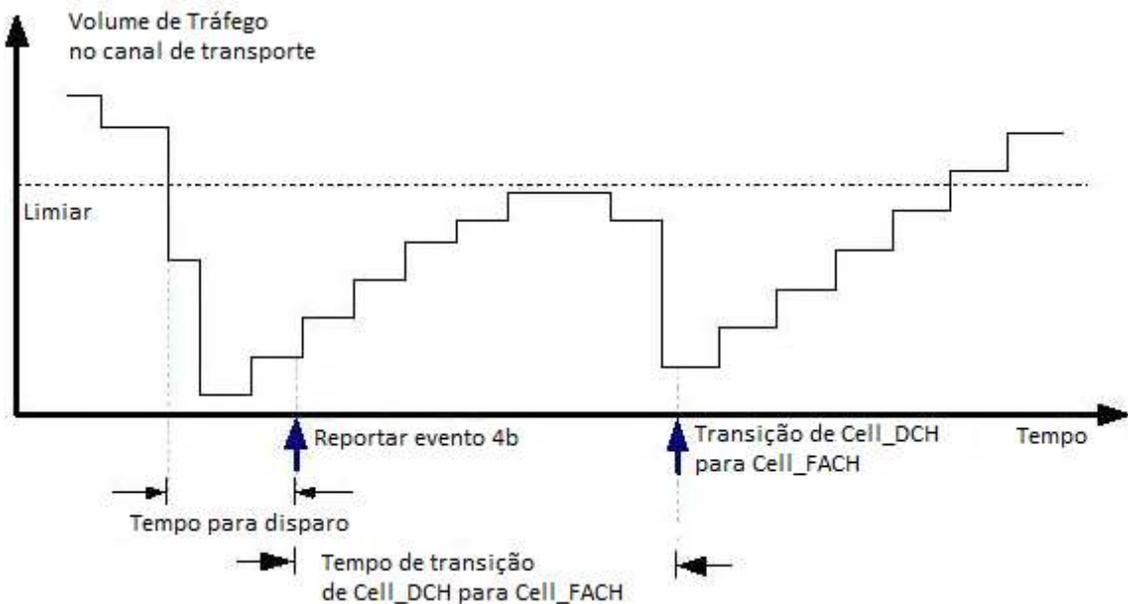


Figura 2.9. Transição de estados do Cell_DCH para Cell_FACH.

TABELA 2.2. RRC STATE INDICATOR – INDICAÇÃO DE ESTADO PARA O UE.[TS25331]

Informação Elemento/Nome do Grupo	Necessidade	Multi	Tipo e Referência	Descrições Semânticas
<i>RRC State Indicator</i>	MP (Parâmetro Obrigatório)	-	Enumerado(CEL_DCH, CELL_FACH, CELL_PCH, URA_PCH)	-

Quando o envio e recebimento de pacotes são cessados, outro mecanismo de troca de estados e desconexão é utilizado. O mecanismo utilizado para transição de estados e desconexão é o tempo de inatividade dos canais do RRC, e que são controlados pelo equipamento RNC do

núcleo da rede. Na Figura 2.10, $T1$, $T2$ e $T3$ representam, respectivamente, os tempos de inatividade dos estados Cell_DCH, Cell_FACH e URA_Idle/Cell_PCH. Caso seja detectado que não há mais tráfego de dados no estado atual e após decorrido determinado tempo, a rede realiza a troca de estado utilizando a mensagem RADIO BEARER RECONFIGURATION. A nomenclatura $T1$, $T2$ e $T3$ não é a oficial do padrão 3GPP (que utiliza apenas o termo tempo de inatividade), no entanto, é uma forma muito comum utilizada em artigos da área e na linguagem técnica. Esse tempo varia de operadora para operadora, e pode também variar de RNC para RNC, ou seja, uma mesma operadora atuando com duas RNCs diferentes, uma no Rio de Janeiro e outra em Campinas podem ter valores de temporização diferentes.

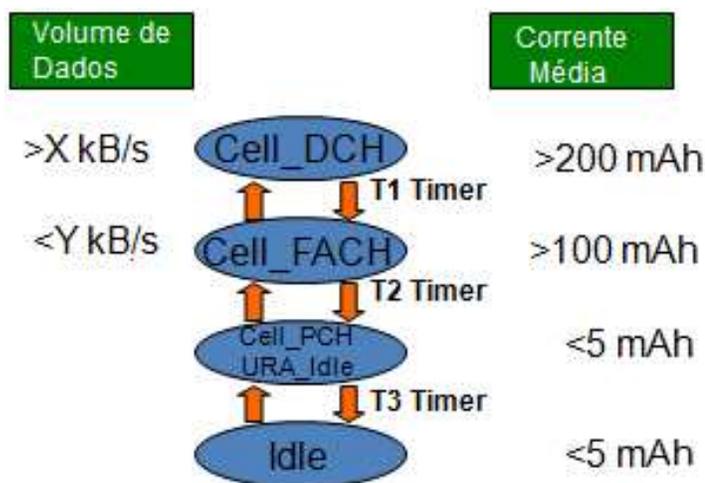


Figura 2.10. Relação entre fluxo de dados recebido, consumo e estados do RRC.

Caso o período em Cell_FACH sem recebimento de pacotes seja grande o suficiente a ponto de ocorrer um estouro do tempo de inatividade, a rede pode decidir transicionar para os estados Cell_PCH ou URA_Idle (ainda no estado *Connected*), ou simplesmente enviar uma mensagem de RRC RELEASE para fechar a conexão do RRC (*Idle*). A desconexão também pode ser originada do dispositivo móvel, e esse mecanismo é utilizado para redução de consumo de energia nos terminais como será visto com maiores detalhes no Capítulo 3.

Na Figura 2.10 também é possível ver o consumo médio de corrente para cada estado. A curva característica de recebimento de um simples pacote é identificada na Figura 2.11. O nível mais alto do gráfico representa o móvel no estado Cell_DCH, enquanto o nível intermediário

representa o estado Cell_FACH. O nível mais baixo representa o estado Idle. Os estados Cell_PCH e URA_Idle possuem o mesmo comportamento apresentado em Idle.

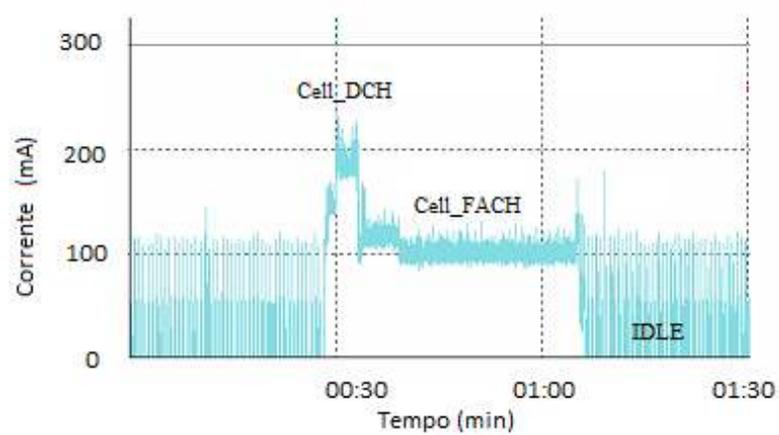


Figura 2.11. Recebimento de um pacote em uma rede 3G. [HAVERINEN,2007]

3 CONSUMO DE ENERGIA: PROBLEMAS E SOLUÇÕES

Conforme descrito anteriormente, este trabalho tem por objetivo a identificação de problemas relacionados ao consumo de energia com relação à utilização de canais da interface aérea com relação ao recebimento de pacotes e sua otimização para redução do consumo, e não a elementos de hardware e sua otimização.

Com a expansão das redes 3G e o surgimento do conceito *Always On*, o alto tráfego de pacotes, principalmente quando o telefone está em *Stand By*, demonstrou ser um dos principais responsáveis por um consumo alto de energia pelo dispositivo móvel.

O método proposto para solução do problema do consumo de energia leva em consideração implementações a serem feitas do lado do dispositivo móvel, porém outras soluções apresentadas no decorrer do capítulo podem ser empregadas no lado da operadora de telefonia móvel.

3.1 RECEBIMENTO DE PACOTES NÃO SOLICITADOS PELO USUÁRIO

Partindo de [RICCIATO, 2006], os pacotes de dados que chegam ao dispositivo móvel podem ser classificados em:

- **Pacotes solicitados:** Tráfego no qual os dados que chegam trazem alguma informação relevante ao usuário, proveniente de uma aplicação, como e-mail, por exemplo.

- **Pacotes não-solicitados:** Tráfego de pacotes que não são provenientes de resposta a uma solicitação pelo usuário, como os ataques de hacker, por exemplo.

Um dispositivo móvel no modo *Always On* possui um endereço válido por onde pode ser localizado. Dependendo do tipo servidor o qual foi conectado, pode ter seu endereço privado ou público. Caso esteja em uma rede privada, qualquer pedido de serviço a um elemento de fora dessa rede será feita por um NAT (*Network Address Translator*), e o dispositivo só terá visibilidade a quem pertencer esta rede e ficará menos suscetível a ataques de pacotes não solicitados. No caso de realizar sua conexão com um servidor que lhe retorne um endereço público de rede, ele estará mais suscetível ao recebimento de pacotes não solicitados, a menos que a rede esteja protegida por algum mecanismo de filtragem de bloqueio de pacotes, como

Firewall. Esses pacotes não solicitados estão relacionados aos ataques do tipo DoS sofrido pelo dispositivo, como apresentado em [LEE-WOO, 2007]. A seguir são apresentados resultados de medições realizadas durante este estudo em uma determinada operadora, que apresentam a quantidade e fluxo de pacotes recebidos pelo dispositivo provenientes de dados não solicitados para uma conexão feita por um servidor o qual forneceu um endereço de IP público ao dispositivo móvel. A Tabela 3.1 apresenta o resumo do tráfego recebido pelo dispositivo.

TABELA 3.1. TRÁFEGO DE PACOTES NÃO SOLICITADOS.

Tempo medido (segundo)	309.746
Pacotes	2.312
Tamanho médio dos Pacotes (bytes)	51,5
Total trafegado (bytes)	119.102
Fluxo (bytes/segundo)	0,385

Na Figura 3.1 é apresentada a distribuição do tráfego ao longo do tempo, e na Figura 3.2 as principais portas atacadas durante o monitoramento realizado.

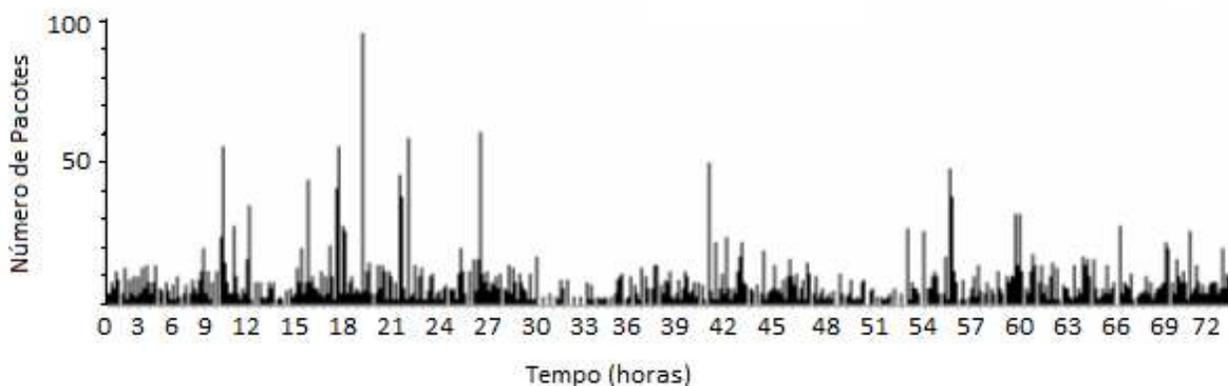


Figura 3.1. Distribuição de pacotes não solicitados recebidos ao longo do tempo.

Os pacotes não solicitados são recebidos pelo dispositivo mesmo que nenhuma aplicação esteja ativa. Para isso, basta o dispositivo manter uma conexão de dados e um endereço válido.

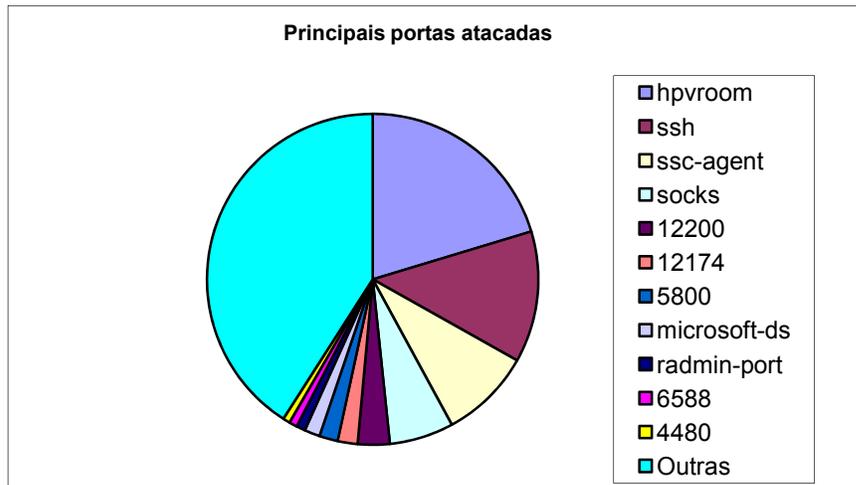


Figura 3.2. Principais portas atacadas.

Para evitar o recebimento desse tráfego, *Firewalls* devem ser aplicados em lugares estratégicos da rede a fim de bloquear o tráfego de pacotes mal intencionados de forma eficiente. Existem diversas tentativas de ataques à rede e para cada uma delas, uma maneira específica de ser tratada. Neste estudo em específico, o foco são os pacotes não solicitados que tem por destino o usuário final, e não ataque a elementos da rede. Para esse tipo de ataque, deve-se priorizar a filtragem de pacotes nas interfaces Gi, Gp e Gn, que são pontos estratégicos de entrada e saída de dados para as redes externa e interna. Isso faz com que ele não chegue a ser recebido pelo usuário final e, com isso, evita que recursos de rádio sejam utilizados, economizando o consumo de energia do dispositivo. A disposição dos *Firewalls* e das interfaces são mostradas na Figura 3.3.

Para esse tipo de bloqueio podem ser utilizadas as seguintes técnicas, como apresentado em [BAVOSA, 2004]:

- **Filtragem de pacotes ingressantes e enviados:** No caso da interface Gp, previne a rede da operadora de telefonia móvel de ser usada como fonte de ataque a outros parceiros de *Roaming*. Na interface Gi, previne a possibilidade de dados não solicitados de dispositivo a dispositivo bloqueando o tráfego entrante cujos endereços de origem são os mesmos dos atribuídos a um dispositivo com acesso à rede pública.
- **Filtragem de pacotes no GTP:** Permite apenas tráfego requisitado e apenas de origem e destino de parceiros de *Roaming*. Isto previne outras redes de operadoras conectadas a

parceira de *Roaming* de iniciar qualquer tipo de ataque. Um *Firewall* que suporta inspeção no GTP assegura que o GGSN ou o SGSN não processem pacotes GTP que são mal formados, tem cabeçalhos mal formados ou não estão no estado correto.

- **Inspeção de pacotes:** Usa uma política de segurança que permite apenas ao dispositivo iniciar conexões à rede pública de dados, e implementa a filtragem de pacotes de forma que o dispositivo nunca enxergue o tráfego que é iniciado da rede pública. Caso necessário, aplicações confiáveis devem ser instaladas no servidor para permitir entrada de dados originadas da rede pública.

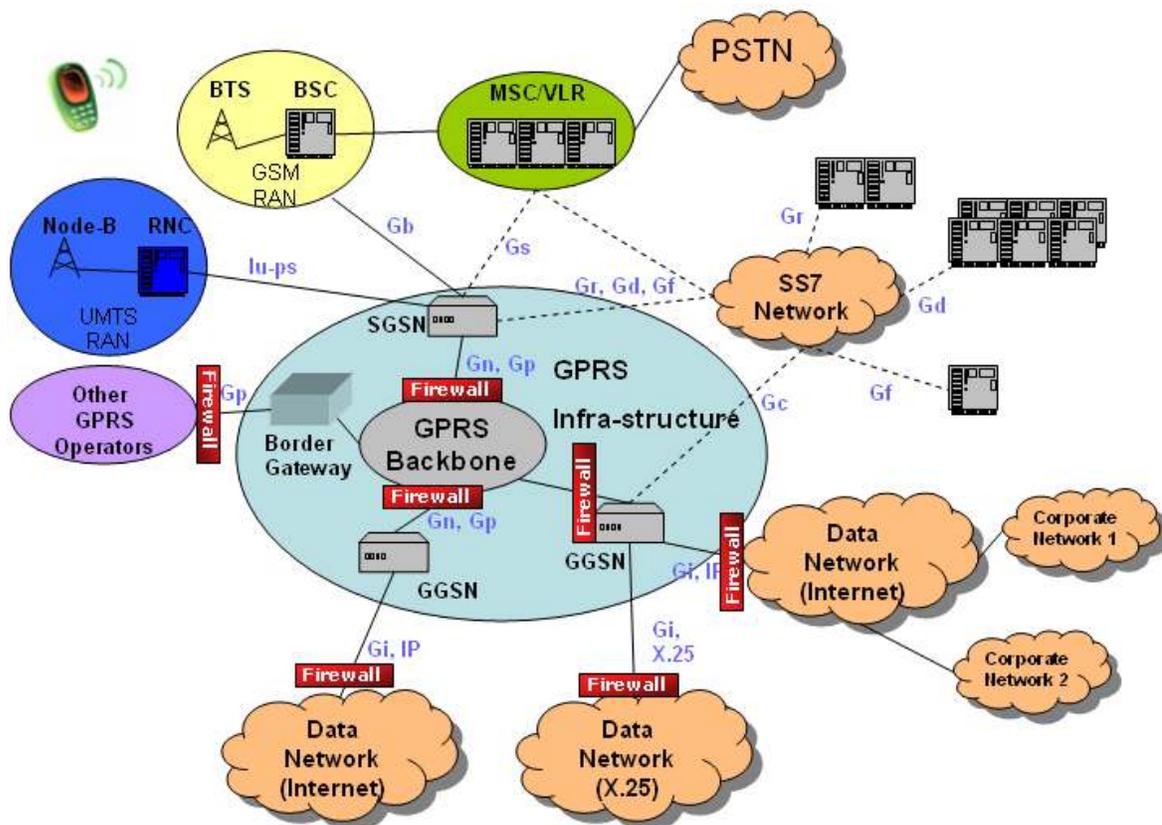


Figura 3.3. Implantação de *Firewalls* nas interfaces de rede.

No caso do uso de NAT, a entrada de pacotes pela interface com a rede externa só pode ser realizada através de pedidos dos usuários de dentro da rede privada. Sendo assim, um pacote que chega ao roteador vindo de fora e que não tenha sido gerado em resposta a um pedido da rede não encontrará nenhuma entrada no NAT e esse pacote será automaticamente descartado, não sendo entregue a nenhum computador da rede. Isso impossibilita a entrada de pacotes mal

intencionados e o NAT acaba funcionando como um *Firewall*. No entanto ele não é um *Firewall* perfeito, pois ainda pode sofrer ataques provenientes da rede interna, apesar das medições realizadas neste trabalho mostrarem que esse tipo de ataque é inexistente, pois usuários mal intencionados utilizam faixas de IP público para realizar o ataque. Outro problema do NAT é o alto custo de implementação quando se tem milhares de usuários conectados à rede. É necessário fazer o mapeamento de todos os endereços e guardá-los em registros, o que acaba sendo inviável para um grande volume de tráfego.

A implantação de *Firewall* GTP é uma solução que exige altos investimentos e deve ser implantada em diversos pontos da rede, porém é eficaz ao bloqueio de pacotes a um baixo custo de processamento extra da rede em oposição ao NAT. A Figura 3.4 mostra o comportamento do consumo de corrente para uma rede não filtrada (a), que possui um consumo de 117,2 mWh em comparação com uma rede filtrada (b), que possui um consumo de 11,92 mWh, gerando uma redução de 89,8% no consumo de energia pois não realiza a alocação de recursos de rádio para recebimento de pacotes não solicitados.

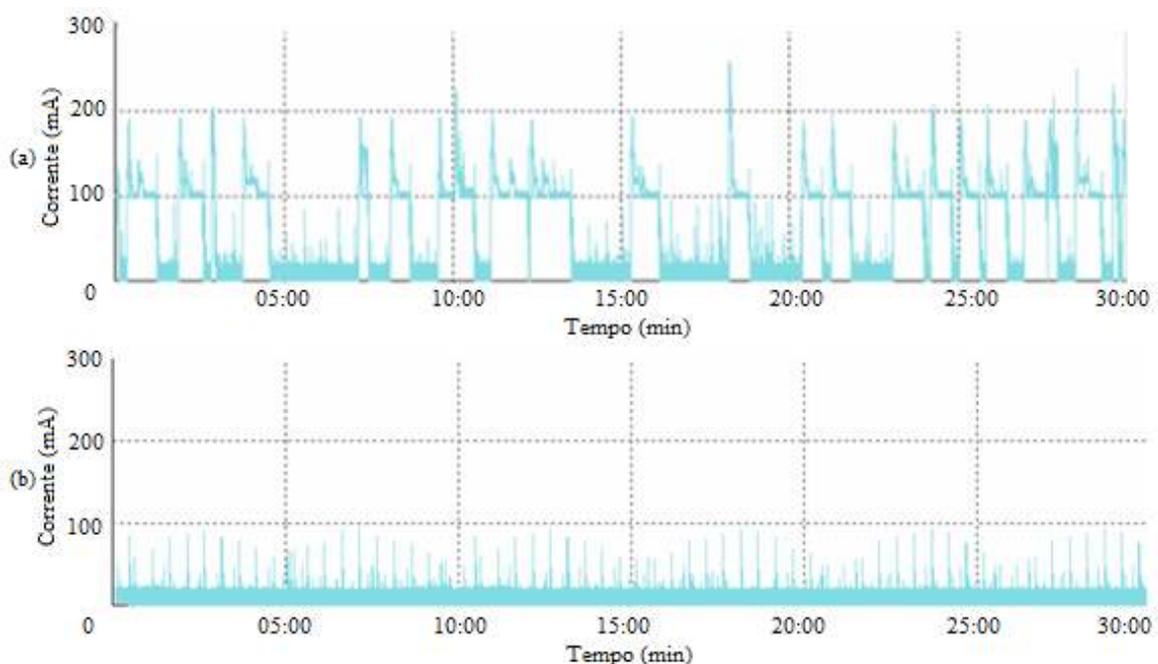


Figura 3.4. (a) Rede aberta desprotegida contra ataques e (b) Rede protegida contra ataques (*Firewall* e/ou NAT).

3.2 CONSUMO EXTRA GERADO POR PACOTES DE SINALIZAÇÃO

O dispositivo permanece no estado *Always On*, pois necessita ser alcançado a qualquer momento. Porém, muitas vezes, não há dados sendo trafegados durante um longo período. A expiração do tempo de conexão por período de inatividade faz com que muitos protocolos incluam um mecanismo de *keep-alive*, que consiste no envio de dados para reiniciar os temporizadores nos *middlebox*³ com determinada periodicidade. Esse tráfego extra de pacotes acaba gerando um consumo extra de energia que depende do fluxo de dados recebidos ou enviados.

Em uma rede em que não haja a presença de pacotes não solicitados, provenientes de ataques de *hacker* ou troca de e-mails, pode-se isolar as mensagens de *keep-alive* para analisar seu impacto no consumo de energia, como mostra a Figura 3.5.

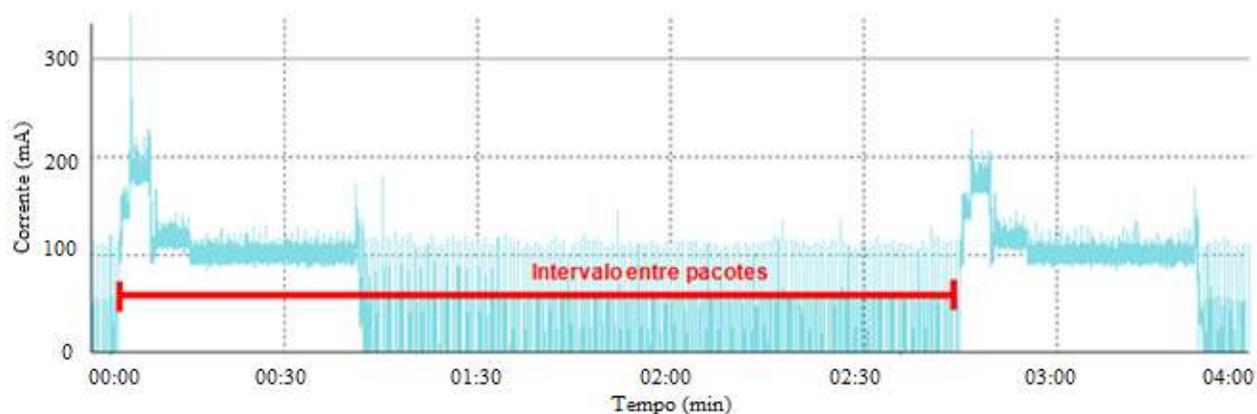


Figura 3.5. Mensagens *keep-alive* recebidas pelo dispositivo móvel. [HAVERINEN, 2007]

Em [HAVERINEN,2007], o autor relaciona o intervalo de chegada de mensagens *keep-alive* e o consumo. O consumo diminui à medida que o intervalo de chegada entre mensagens *keep-alive* aumenta. O consumo é mais significativo em redes 3G do que nas redes 2G devido às características dos canais de rádio. Para os casos extremos o qual o intervalo entre mensagens é muito pequeno (menor que 20s) ou muito grande (tendendo ao infinito), o valor de consumo entre

³ Equipamento que provê aplicação de políticas de transporte de dados em uma rede, gerenciamento de sistemas de detecção de intrusos e gerenciamento de *buffer* multimídia.

as redes 2G e 3G são bem próximos. No entanto, para os outros valores desse intervalo, que são mais comuns em um tráfego de rede, o consumo em uma rede 3G é maior do que em uma rede 2G, como mostra a Figura 3.6.

A solução de aumento do intervalo de mensagens *keep-alive* é exclusiva para casos onde a conexão é feita por meio de um servidor que necessite de atualização periódica de status do dispositivo como conexões que utilizam NAT, o qual esse tipo de tráfego é aplicado. O ajuste do intervalo entre mensagens é configurado do lado do servidor e um intervalo ideal que leve em conta um baixo consumo de energia está fora do alcance dos fabricantes de dispositivo. Com isso, eles acabam sofrendo as consequências do mau dimensionamento dos servidores da infraestrutura da rede.

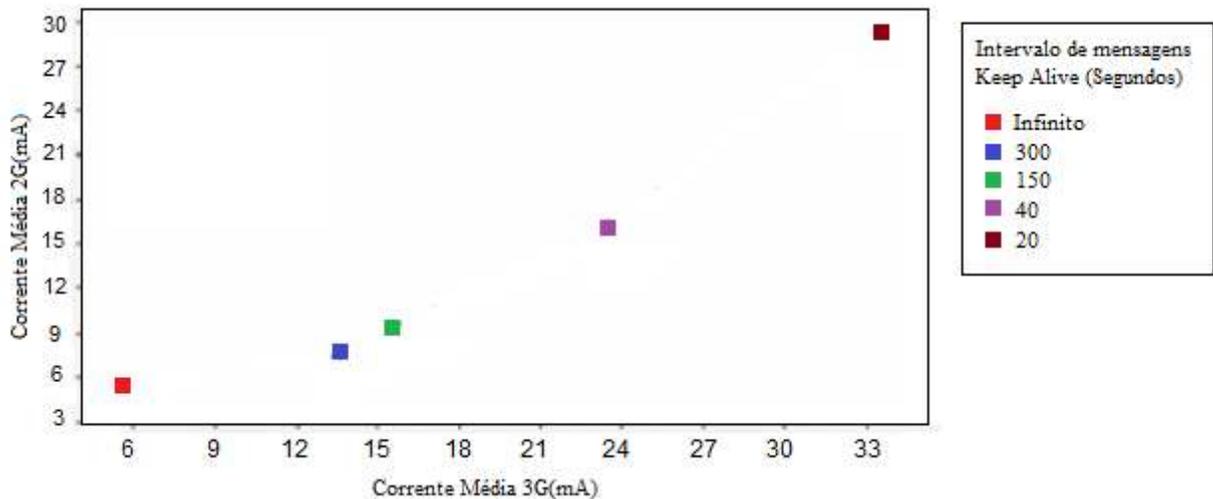


Figura 3.6. Relação entre consumo 2G e 3G para mensagens *keep-alive* [HAVERINEN, 2007]

Outro problema que tem características semelhantes aos das mensagens de *keep-alive* são as mensagens de *polling* trocadas entre cliente (dispositivo móvel) e servidor (servidor de aplicação de e-mail). Essas mensagens de sinalização têm por finalidade manter a conexão ativa entre eles, e são trocadas a um intervalo que pode ser configurado pelo cliente ou servidor. Quando for possível a configuração do lado do cliente, os fabricantes de dispositivo podem realizar um dimensionamento adequado para um consumo eficiente de energia, respeitando as características do servidor. Caso contrário, essa mudança tem que ser feita no servidor, o que muitas vezes não é possível devido a restrições da operadora.

3.3 TEMPO DE INATIVIDADE DO RRC

Esse problema não está intrínseco apenas no recebimento de pacotes não solicitados ou mensagens de *keep-alive*, mas em qualquer tipo de pacote que seja trafegado pelo dispositivo móvel em qualquer aplicação. Esse tempo é determinado pela rede, e indica quanto tempo após o recebimento do último pacote de dados o dispositivo deve permanecer em determinado estado antes de fazer a transição para outro estado do RRC.

Manter os recursos de rádio ativos durante um longo período é uma solução para a operadora economizar recursos de rede. Com isso, novas conexões de rádio são evitadas caso vários pacotes sejam enviados em um curto intervalo de tempo. Por outro lado, o dispositivo móvel é prejudicado, pois deve se manter nesse estado consumindo energia sem que nenhuma informação seja recebida.

Um exemplo de uma operadora real é ilustrado na Figura 3.7, que tem esse valor configurado para 30 segundos, gerando um consumo extra de 841 uWh nesse período.

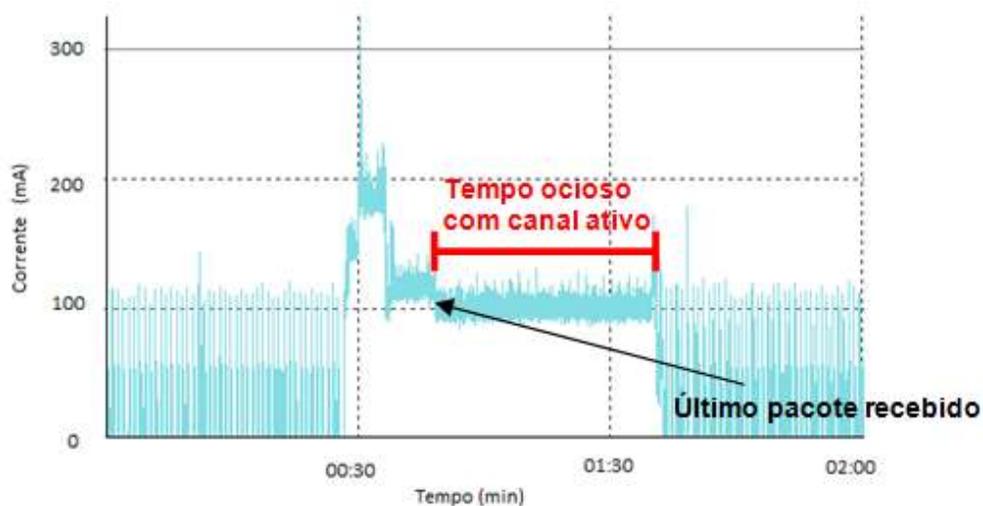


Figura 3.7. Representação do tempo de inatividade dos canais para uma operadora específica.

Esse problema deve ser tratado com maior atenção, uma vez que está presente no recebimento de qualquer tipo de pacote, seja ele solicitado ou não. Como esse tempo é configurado pelo lado da rede da operadora, e não do dispositivo, deve-se recorrer a uma solução alternativa para que os recursos de rádio não permaneçam ativos durante longos períodos, e

desligá-los quando não necessários. O padrão 3GPP estabelece na norma [TS25331] uma maneira para que a desconexão do RRC seja requisitada pelo dispositivo quando detectar inatividade e, dessa maneira, evitar o consumo excessivo de energia. A Figura 3.8 ilustra o procedimento de desconexão do RRC com requisição originada pelo dispositivo. Este procedimento causa sinalização extra de uma mensagem do que se a desconexão partisse da operadora. O procedimento de desconexão do RRC ser realizado pelo dispositivo móvel antes do término do tempo de inatividade da rede é chamado *Fast Dormancy* [QIAN, 2010], e é comumente utilizado por vários fabricantes de dispositivos móveis.

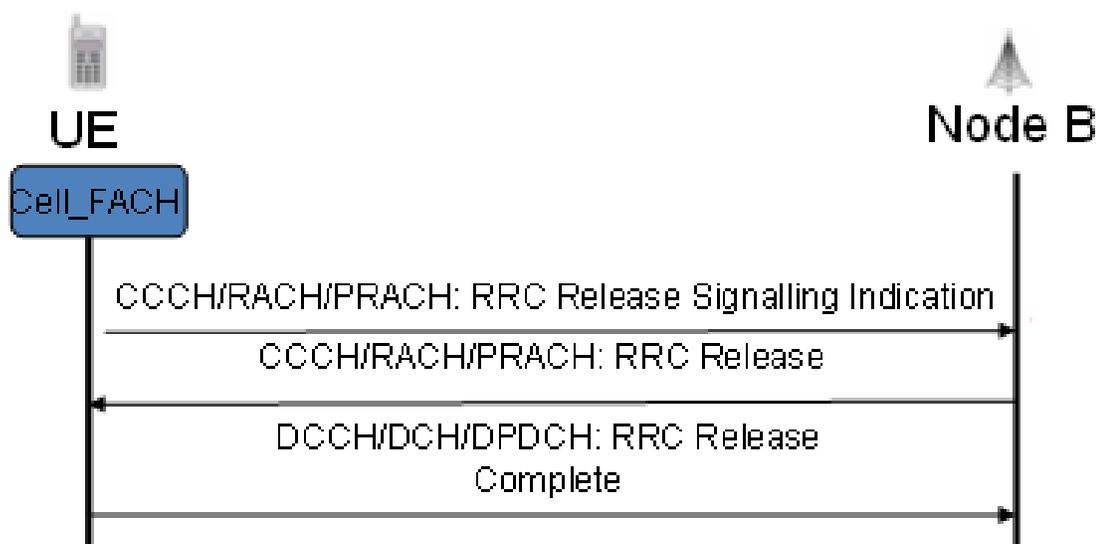
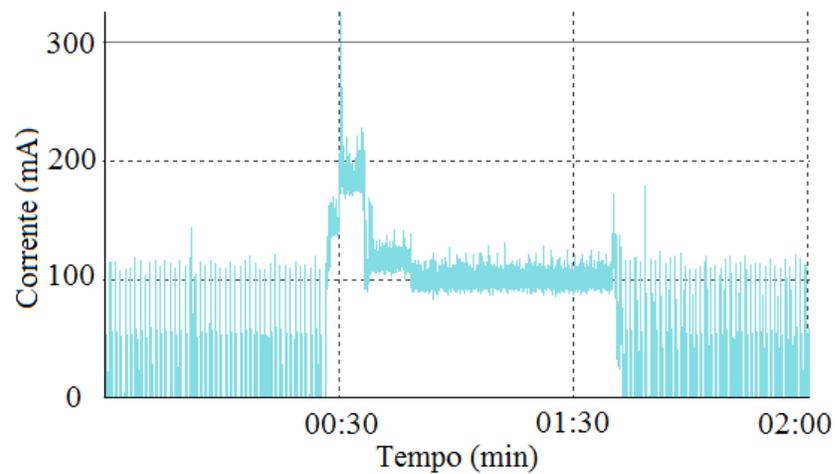


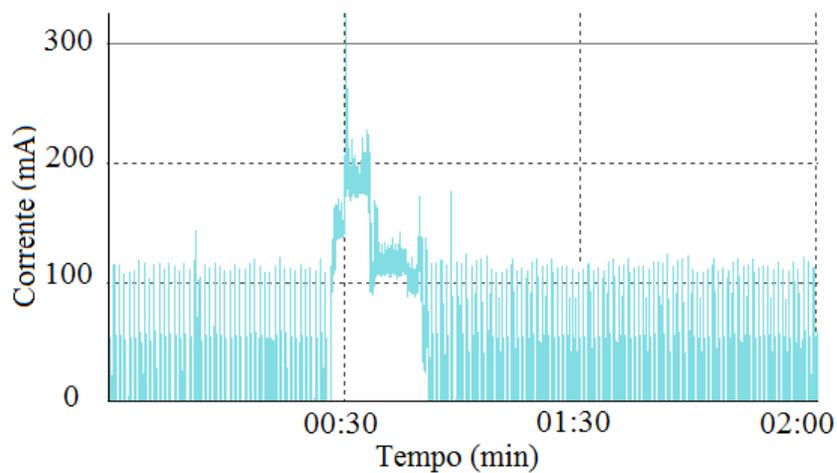
Figura 3.8. Desconexão do RRC requisitada pelo dispositivo.

A Figura 3.9 ilustra o consumo gerado para os casos onde o $T2$ da rede é mantido inalterado em 30 segundos e para o caso onde esse tempo é reduzido para 5 segundos, tendo a desconexão originada pelo telefone. Com a redução do intervalo, o consumo passa de 4,43mWh para 1,78mWh, gerando uma redução no consumo de 59,7%.

O consumo de energia será tão menor quanto menor for o tempo de inatividade atribuído ao Cell_FACH [HAVERINEN, 2007]. Embora essa solução seja eficiente em termos de diminuição do consumo, alguns problemas podem ocorrer, como a diminuição de desempenho para acesso de dados e aumento de mensagens de sinalização.



(a)



(b)

Figura 3.9. Consumo gerado pelo recebimento de um pacote para um $T2$ de (a) 30 segundos (rede) e (b) 5 segundos (telefone).

A diminuição de desempenho está ligada à acessibilidade do usuário. Se o tempo de inatividade for muito curto, e uma nova requisição para acesso de um conteúdo (como uma página *Web*, por exemplo) for realizada logo após a desconexão do RRC, a resposta irá demorar mais devido ao tempo extra de um novo estabelecimento de conexão do RRC que deve ser realizado. Conseqüentemente, há um aumento do número de sinalizações entre o dispositivo e a rede para estabelecimento dessa nova conexão, o que não aconteceria para um tempo de inatividade um pouco menor.

A Figura 3.10 ilustra o recebimento de dois pacotes com um intervalo de 15 segundos entre eles. Em (a) tem-se um $T2$ configurado para 30 segundos (rede), e em (b) 5 segundos (dispositivo). Para ambos, $T1$ é de 5 segundos.

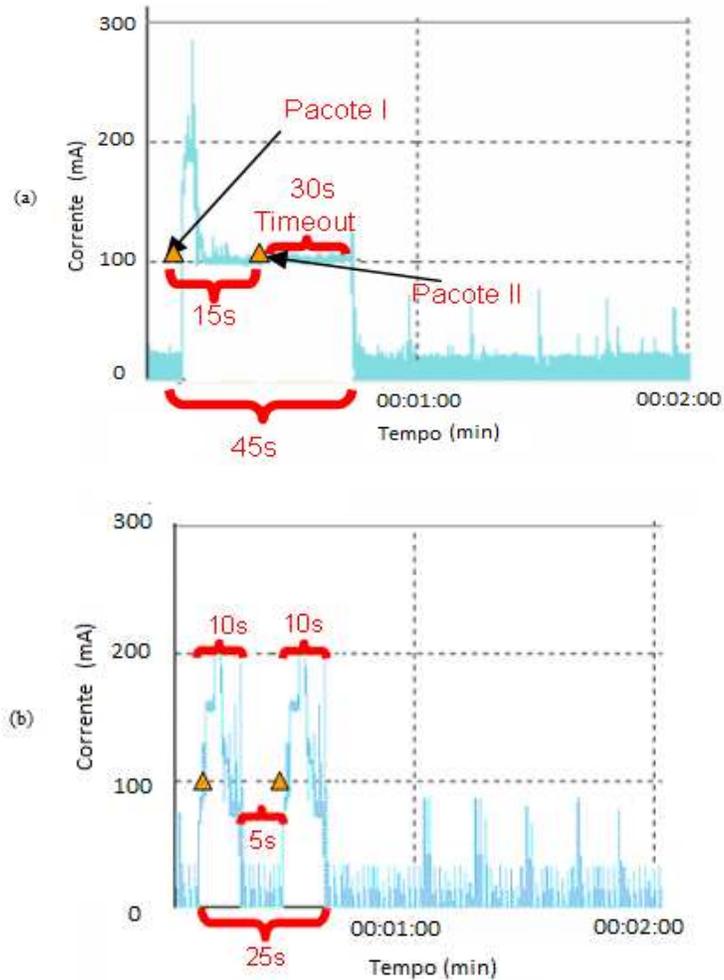


Figura 3.10. Consumo gerado pelo recebimento de dois pacotes a um intervalo de 15 segundos para um $T2$ de (a) 30 segundos (rede) e (b) 5 segundos (telefone).

Para o tempo de inatividade de maior valor, passados 15 segundos do recebimento do primeiro pacote, o dispositivo ainda está no estado Cell_FACH utilizando o canal comum com conexão ativa no RRC, e o pacote é recebido sem que haja necessidade de uma nova conexão. Um temporizador de 30 segundos logo após o recebimento do segundo pacote é então iniciado, fazendo com que a transação de recebimento dos dois pacotes demore 45 segundos, consumindo 5,35 mWh.

Para o tempo de inatividade de menor valor, passados 10 segundos após o recebimento do primeiro pacote, a conexão com o RRC é desfeita. Cinco segundos após a desconexão, o novo pacote é recebido. Nesse caso, uma nova conexão com o RRC deve ser refeita, aumentando o número de mensagens de sinalização e um atraso maior no recebimento do pacote para o usuário. O pacote é então recebido, e após transicionar para Cell_FACH aguarda os 5 segundos de inatividade para desfazer a conexão do RRC. A transação nesse caso tem duração de 25 segundos, consumindo 3,337 mWh. Nesse caso, a redução do consumo chega em torno de 37,6%.

3.4 REDUÇÃO DINÂMICA DO TEMPO DE INATIVIDADE DO RRC

Em [LIERS, 2006], é apresentada uma proposta de redução dinâmica, controlada pelo núcleo da rede, onde o dinamismo se baseia em um melhor controle para atribuição de recursos a usuários que realmente necessitem, utilizando um método adaptativo, que se baseia além da verificação dos estados do RRC, em características como controle de admissão de chamadas e controle de congestionamento. Em [QIAN, 2010], no entanto, são citados os problemas em se utilizar um tempo de inatividade dinâmico controlado pela rede. Além do aumento de sobrecarga de processamento para calcular os tempos adequados para cada usuário, diferentes padrões de acessos a diferentes aplicações tornam mais difícil de o controle ser feito pelo RNC.

Um controle dinâmico realizado pelo dispositivo móvel surge como uma alternativa ao método dinâmico controlado pela rede.

Neste trabalho é proposto um método alternativo para solucionar o problema do consumo extra de energia dos dispositivos móveis baseado na redução do tempo de inatividade do RRC de uma forma dinâmica, como mostrado em [LEE-CHEN, 2004], porém utilizando modelos de previsão.

O método da redução de $T2$ mostra-se bastante eficiente quando o intervalo de recebimento de pacotes é grande, o que acontece em períodos de baixo fluxo. Para o Período de Maior Movimento (PMM), onde o fluxo de recebimento de pacotes é maior e o intervalo de chegada entre pacotes menor, o método de redução de $T2$ ainda apresenta uma redução no consumo, porém muito inferior ao de período de baixo fluxo.

Para otimizar o consumo em ambos os casos, é necessário um modelo de previsão que verifique quando irá ocorrer a chegada de um novo pacote, e dessa forma tomar a melhor decisão

se deve haver ou não a desconexão do RRC. Com isso, desconexões indevidas são evitadas caso o próximo pacote chegue em um curto intervalo de tempo e evitaria o tempo de inatividade que consome energia desnecessária quando o próximo pacote a ser recebido estiver em um intervalo grande.

Além de prever quando um novo pacote irá chegar, deve-se também saber a estimativa do fluxo de pacotes a serem recebidos. Com isso, é possível verificar qual o estado mais adequado, por exemplo, se o dispositivo estiver em Cell_FACH, e em um curto intervalo de tempo for receber um fluxo de pacotes muito grande a ponto de ter que mudar para o estado Cell_DCH, pode-se desfazer a conexão do RRC e estabelecê-la novamente, pois esse procedimento economiza energia em alguns segundos de inatividade em Cell_FACH e não ocasiona atrasos de recebimento, pois o tempo para sinalização de troca de estados é semelhante ao de estabelecimento de uma nova conexão.

No entanto, se o dispositivo estiver em Cell_FACH e o novo fluxo previsto para ser recebido for tão baixo a ponto de permanecer nesse estado, deve-se abdicar de poucos segundos de redução de consumo e permanecer conectado no RRC para o recebimento dos pacotes, evitando atrasos de recebimento devido a desconexões e novas conexões, e, além disso, se uma nova conexão fosse feita ao invés de permanecer em Cell_FACH, o dispositivo entraria em Cell_DCH e o consumo extra gerado por este estado seria maior do que se permanecesse em Cell_FACH.

A Figura 3.11 ilustra o comportamento e estimativa de consumo caso fosse utilizado modelo de previsão.

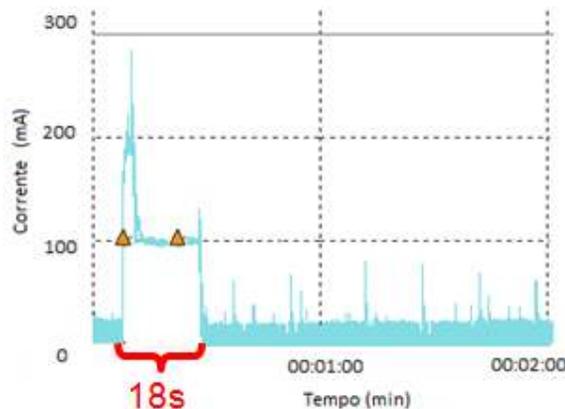


Figura 3.11. Consumo gerado pelo recebimento de dois pacotes em um intervalo de 15 segundos para um tempo de inatividade dinâmico – estimado.

Após o recebimento do primeiro pacote, o dispositivo permaneceria conectado em Cell_FACH, uma vez que o resultado de sua previsão apontaria a chegada de um pacote nos próximos segundos. Após o recebimento do segundo pacote, uma nova previsão é feita e nenhum pacote é apontado para ser recebido. Nesse exemplo é suposto um intervalo de 3 segundos para que seja efetuada uma nova previsão, totalizando 18 segundos para a operação desde o recebimento do primeiro pacote até sua desconexão do RRC, consumindo 2,331 mWh. Se comparado com o caso da Figura 3.10 (a) em que o $T2$ é da ordem de 30 segundos, tem-se uma redução de 57,4%, o que corresponde a aproximadamente 20% a mais de economia do que se fosse utilizado um $T2$ de 5 segundos da Figura 3.10(b).

4 MODELO DE PREVISÃO

Para mudança dos tempos de inatividade de forma dinâmica é necessário um modelo de previsão que consiga projetar da forma mais precisa possível o fluxo de dados que será recebido pelo dispositivo móvel nos próximos instantes e, desse modo, fazer com os recursos de rádio permaneçam alocados ou não.

O modelo de previsão escolhido foi o filtro de Kalman [OKUTANI, 1984] devido a sua excelente aceitação para previsões em tempo real e utilização em previsões de tráfego.

Outros modelos de previsão também podem ser utilizados, como Redes Neurais [SMITH, 1994], e Lógica Fuzzy [LI, 2006], por exemplo. No entanto, a eficiência na redução do consumo ficará dependente da eficiência e desempenho do filtro.

4.1 FILTRO DE KALMAN

Em [Kalman, 1960], foi descrita pela primeira vez uma solução recursiva para o problema de filtragem de dados discretos em um sistema linear. Dados alguns valores iniciais, pode-se prever e ajustar os parâmetros do modelo a partir de cada nova medição, obtendo a estimativa do erro em cada atualização. A sua habilidade para incorporar os efeitos de erros e sua estrutura computacional fez com que o filtro de Kalman tivesse um amplo campo de aplicações, especialmente no que se refere à análise de trajetórias em visão computacional.

O filtro de Kalman estima um processo usando uma forma de controle por realimentação: o filtro estima o estado do processo em dado instante e obtém então a realimentação na forma de medidas ruidosas. Assim, as equações para o filtro de Kalman podem ser representadas por dois grupos: atualização das equações de tempo e a atualização das equações de medição.

As equações de atualização do tempo são responsáveis para projetar com antecedência (no tempo) as estimativas da covariância do estado atual e do erro, para obter as estimativas a priori para a próxima etapa.

As equações de atualização de medição são responsáveis pela realimentação, isto é, incorpora-se uma nova medida na estimativa a priori para obter posteriormente uma estimativa melhorada.

Em resumo, as equações de atualização do tempo são equações de predição, enquanto as equações de atualização de medição são equações de correção.

O filtro de Kalman trata do problema de estimação recursiva do vetor de estados $x \in \mathfrak{R}^n$ de um sistema dinâmico estocástico representado no espaço pelas equações (4.1) e (4.2):

$$X_{k+1} = \Phi X_k + LU_k + \Gamma W_k \quad (4.1)$$

$$Y_k = HX_k + V_k \quad (4.2)$$

onde X_k representa o vetor de estado no tempo k e Y_k o vetor de saída. As variáveis aleatórias independentes W_k e V_k são vetores de erro de modelagem e erro de observação, respectivamente, e $Q_k = E[W_k \cdot W_k^T]$ e $R_k = E[V_k \cdot V_k^T]$ representam, respectivamente, a covariância do vetor erro de modelagem e a covariância do vetor erro de observação. As matrizes de covariância refletem o nível de incerteza em relação à estimativa de X .

A matriz $\Phi_{n \times n}$ na equação diferencial descreve o estado no tempo anterior k para o estado atual $k+1$, na falta de uma função de referência ou ruído processado. Na prática muda em cada passo, porém assume-se que ele é constante, com valor $\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. A matriz L é o ganho da entrada determinística e Γ é a matriz de ruído de modelagem.

A matriz H descreve o estado para a medida Y_k (matriz de observação). Na prática pode mudar em cada nova medição ou temporalmente, porém assume-se que ela é constante.

Seja X_k^* a predição do estado X_k a partir do estado estimado do instante anterior \bar{X}_k , e \bar{X}_k a atualização dessa estimativa com base na leitura efetuada no instante k (filtragem), dados pelas equações (4.3) e (4.4).

$$\bar{X}_k = X_k^* + K_k r_k \quad (4.3)$$

$$X_k^* = \Phi \bar{X}_k \quad (4.4)$$

sendo que $\bar{X}_k = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix}$ e $X_k^* = \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix}$, onde \bar{X}_1 e X_1^* são componentes correspondentes ao nível, e

\bar{X}_2 e X_2^* componentes relacionadas ao incremento de crescimento.

O termo r_k dado pela equação $r_k = Y_k - HX_k^*$ (4.5) representa o erro de predição (também chamado como termo de inovação) entre a saída Y_k e sua previsão $y^* = HX_k^*$, e a matriz K_k é denominada ganho do filtro, que é obtida pela minimização da matriz de covariância do erro de predição P_k^* , dada por:

$$P_k^* = \Phi \bar{P}_k \Phi^T + Q_k \quad (4.6)$$

$$K_k = P_k^* H^T (\sigma_k^2)^{-1} \quad (4.7)$$

onde a covariância do termo de inovação é dada por:

$$\sigma_k^2 = E[r_k^2] = H P_k^* H^T + R_k \quad (4.8)$$

A covariância do erro de estimação é dada por:

$$\bar{P}_k = (I - K_k H) P_k^* \quad (4.9)$$

Pode-se inferir pelo ganho de Kalman que à medida que a variância do ruído sobre a saída R diminui, o ganho de Kalman aumenta de modo a influenciar mais sobre o resíduo.

À medida que a matriz de covariância do erro de predição P_k^* diminui, o ganho de Kalman vai influenciar cada vez menos no resíduo.

Quando a covariância do ruído sobre a saída diminui, a leitura atual Y_k é cada vez mais confiável e menos confiável é a predição da saída.

A diminuição de P_k^* implica numa confiança maior sobre a estimativa, porém menor confiança sobre a leitura.

A primeira tarefa durante a atualização da medida é computar o ganho de Kalman. O valor filtrado é então obtido levando em conta o valor estimado e o ganho de Kalman. A próxima previsão (valor estimado) é feita baseada no valor filtrado, como mostra a Figura 4.1. O filtro de Kalman condiciona recursivamente a estimativa atual em todas as medidas passadas. Essa natureza recursiva é um dos diferenciais do filtro.

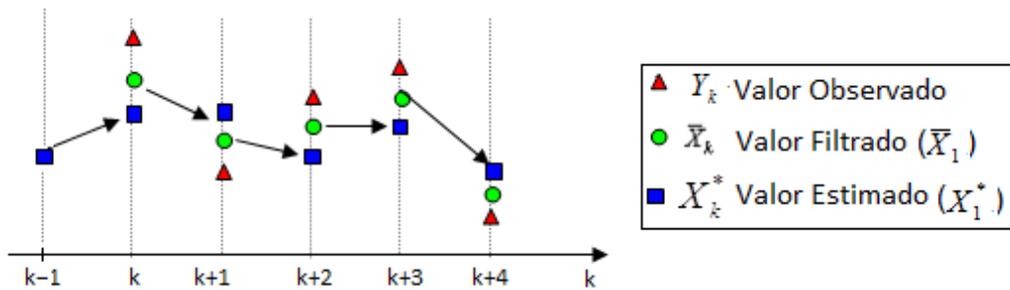


Figura 4.1. Passos para estimação utilizando filtro de Kalman.

No algoritmo discreto do filtro, são utilizados os passos apresentados na Figura 4.2:

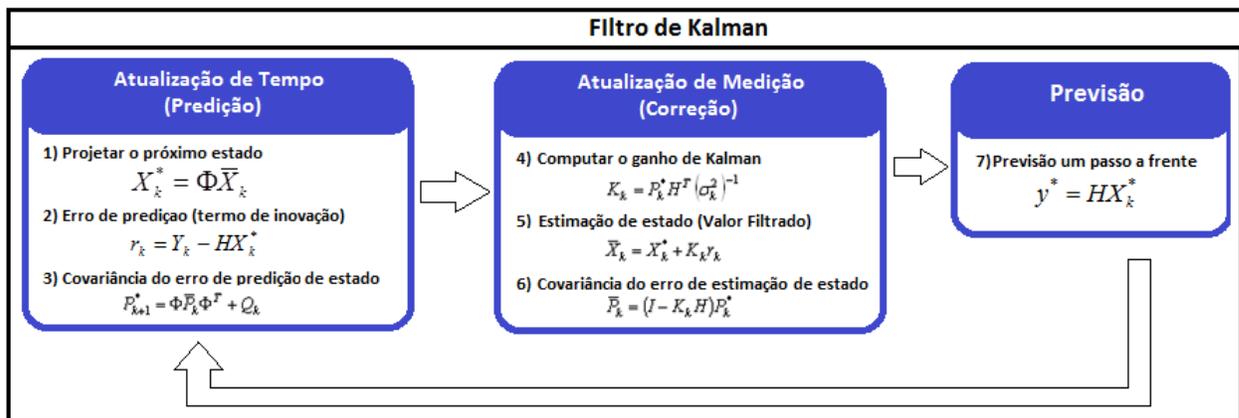


Figura 4.2. Algoritmo para cálculo do Filtro de Kalman.

4.2 SUBFILTRO DE KALMAN

Para implementação do filtro de Kalman é necessário o conhecimento prévio das variâncias do ruído de modelagem e do ruído de observação. A obtenção da covariância do ruído sobre a saída pode ser efetuada a partir de uma sucessão de leituras obtidas separadamente.

Quando $W_k = W$ e $V_k = V$ são constantes, o ganho de Kalman e a respectiva matriz de covariância convergem para um valor constante.

Na execução real do filtro, a covariância do ruído da medida R_k é medida antes da operação do filtro.

A determinação do processo de covariância do ruído Q_k exige mais esforços porque não se tem a facilidade de observar diretamente o processo que se está estimando. Um processo do modelo relativamente simples pode produzir resultados aceitáveis se forem incorporadas muitas incertezas ao processo via seleção de Q_k . Quando não se tem uma base racional para escolher os parâmetros, frequentemente se mede o melhor desempenho do filtro (estatisticamente falando), que pode ser obtido ajustando os parâmetros do filtro Q_k e R_k . Esse ajuste é feito de modo *off-line*, ou seja, com a ajuda de outro filtro, separadamente ao filtro de Kalman, a fim de que não ocorra a divergência do sistema devido às covariâncias de erro calculadas.

Para o cálculo da matriz Q_k , um filtro adicional chamado subfiltro de Kalman foi proposto por [RIOS, 1982].

Para que o método de Kalman convirja, ele necessita que a relação de covariâncias Q_k/R_k mantenha-se estável, dentro de uma margem de tolerância, durante o procedimento de cálculo. A solução proposta é efetuar continuamente o recálculo dos valores das matrizes Q_k e R_k a cada passo de interação, de modo a manter a relação Q_k/R_k estável no entorno do respectivo estado. Para tanto, as matrizes Q_k e R_k também são recalculadas. Esse recálculo é efetuado pelo subfiltro. Para isso, o subfiltro calcula o erro de predição r_k^v , supondo uma curva real, dada pela relação descrita na Figura 4.3.

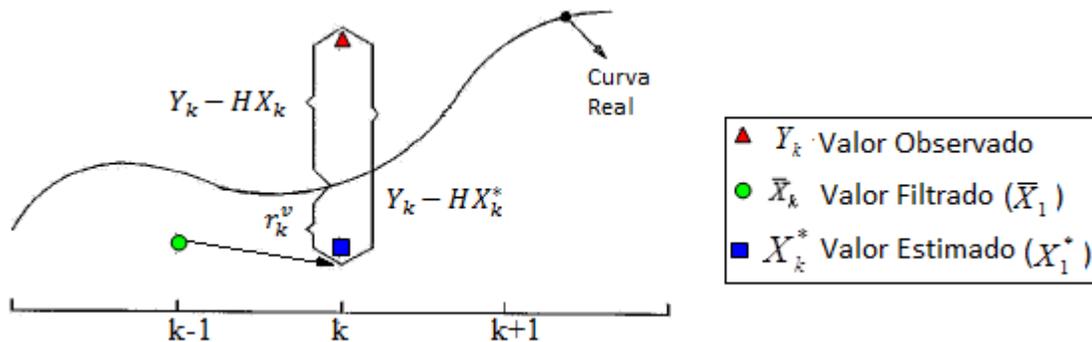


Figura 4.3. Dedução do subfiltro.

A fim de obter um modelo especial para os elementos da matriz Q_k , o erro de predição é definido como:

$$r_{k+1}^v \triangleq (Y_{k+1} - HX_{k+1}^*)(Y_{k+1} - HX_{k+1}) \quad (4.10)$$

Das equações (4.1) e (4.5) pode-se dizer que:

$$r_{k+1}^v = r_{k+1} - V_{k+1} \quad (4.11)$$

A equação (4.12) é obtida pelo cancelamento dos valores Y_{k+1} em (4.10), seguido pela substituição de X_{k+1}^* e X_{k+1} por suas respectivas expressões dada pelo sistema de equações (4.1) e (4.2), e a equação de predição (4.4):

$$r_{k+1}^v = H[\Phi(X_k - \bar{X}_k) + \Gamma W_k] \quad (4.12)$$

Deste modo:

$$\left. \begin{aligned} E[(r_{k+1}^v)^2] &= H(\Phi \bar{P}_k \Phi^T + \Gamma Q_k \Gamma^T) H^T \\ &= H \Phi \bar{P}_k \Phi^T H^T + H \Gamma Q_k \Gamma^T H^T \end{aligned} \right\} \quad (4.13)$$

Partindo da equação (4.10):

$$(r_{k+1}^v)^2 = r_{k+1}^2 - 2r_{k+1}V_{k+1} + V_{k+1}^2 \quad (4.14)$$

Usando a consistência estatística, $E[r_{k+1}V_{k+1}] = 0$ e $R_{k+1} = E[V_{k+1}^2]$, segue que:

$$E[(r_{k+1}^v)^2] = r_{k+1}^2 + R_{k+1} \quad (4.15)$$

Definindo o erro de observação esperado do erro de covariância da predição real, chamado de pseudo-observação, como:

$$z_{k+1} \triangleq E[(r_{k+1}^v)^2] - E[(r_{k+1}^v)^2 | Q_k = 0] \quad (4.16)$$

Usando (4.13) e (4.15), tem-se que:

$$z_{k+1} = r_{k+1}^2 + R_{k+1} - H \Phi \bar{P}_k \Phi^T H^T \quad (4.17)$$

A equação de pseudo-observação é dada por:

$$z_{k+1} = M q_{k+1} + \eta_{k+1} \quad (4.18)$$

Levando em consideração (4.13) e (4.14):

$$M q_{k+1} = H \Gamma Q_k \Gamma^T H^T \quad (4.19)$$

e

$$\eta_{k+1} = 2r_{k+1}V_{k+1} - V_{k+1}^2 + R_{k+1} \quad (4.20)$$

O modelo de estado para este subsistema é dado por:

$$q_{k+1} = \Phi^q q_k + W_k^q \quad (4.21)$$

onde Φ^q é a matriz de transição com $\Phi^q \in \mathfrak{R}^{(r+p) \times (r+p)}$, e p o número de elementos da matriz triangular superior de Q_k (ou inferior, uma vez que Q_k é simétrica), sem a diagonal principal (que possui os elementos r). Como a dinâmica do sistema é desconhecida, a matriz identidade é usada para esta matriz de transição. W_k^q é um vetor de ruído branco com média zero e covariância σ_q^2 .

Os valores dos elementos da matriz estão relacionados à precisão do equipamento o qual está sendo calculado. Por exemplo, um computador que tem precisão de d dígitos, a covariância é dada por:

$$\sigma_q^2 = 10^{-2d} I, I \in \mathfrak{R}^{(r+p) \times (r+p)} \quad (4.22)$$

onde I é a matriz identidade. Para estimar a matriz Q_k a partir do vetor de estados q_k , o subfiltro de Kalman é utilizado da seguinte forma:

- **Correção**
$$K_{k+1}^q = P_{k+1}^{q*} M^T \left(M P_{k+1}^{q*} M^T + E[\eta_{k+1}^2] \right)^{-1} \quad (4.23)$$

$$\bar{q}_{k+1} = q_{k+1}^* + K_{k+1}^q (z_{k+1} - M q_{k+1}^*) \quad (4.24)$$

$$\bar{P}_{k+1}^q = P_{k+1}^{q*} - K_{k+1}^q M P_{k+1}^{q*} \quad (4.25)$$

- **Predição**
$$q_{k+1}^* = \Phi^q \bar{q}_k \quad (4.26)$$

$$P_{k+1}^{q*} = \Phi^q \bar{P}_k^q (\Phi^q)^T + \sigma_q^2 \quad (4.27)$$

onde \bar{P}_k^q é o erro de estimação da covariância. Este valor é comumente inicializado como $\bar{P}_0^q = 10^6 I$. O valor de \bar{q}_0 depende do valor inicial do filtro de Kalman. O erro de covariância da pseudo-observação é dado por:

$$E[\eta_{k+1}^2] = 4r_{k+1}^2 R_{k+1} + 2R_{k+1}^2 \quad (4.28)$$

O vetor q_k é dado por:

$$q_k^T = [Q_{11}, Q_{22}, \dots, Q_{rr}, Q_{12}, Q_{13}, \dots, Q_{(r-1)r}] \quad (4.29)$$

A matriz Q_k após a aplicação de cada subfiltro é mostrada a seguir:

$$\left. \begin{array}{l} q_1 \rightarrow Q_{11} \\ q_2 \rightarrow Q_{22} \\ \vdots \\ q_r \rightarrow Q_{rr} \end{array} \right\} \begin{array}{l} \text{elementos} \\ \text{da} \\ \text{diagonal} \\ \text{principal} \end{array} \quad \left. \begin{array}{l} q_{r+1} \rightarrow Q_{12} \text{ and } Q_{21} \\ q_{r+2} \rightarrow Q_{13} \text{ and } Q_{31} \\ \vdots \\ q_{r+p} \rightarrow zQ_{(r-1)r} \text{ and } Q_{r(r-1)} \end{array} \right\} \begin{array}{l} \text{elementos} \\ \text{fora da} \\ \text{diagonal} \\ \text{principal} \end{array}$$

Após a estimação destes valores, a matriz Q_k é testada de modo a evitar que Q_k se torne negativa:

$$j = 1, 2 \begin{cases} \text{se } q_k^*(j, 1) \geq 0 \rightarrow Q(j, j) = q_k^*(j, 1) \\ \text{senão, } Q(j, j) = 0 \end{cases}$$

Colocando as equações obtidas acima de forma prática, pode-se levar em consideração o algoritmo para predição utilizando o filtro e subfiltro de Kalman apresentado na Figura 4.4.

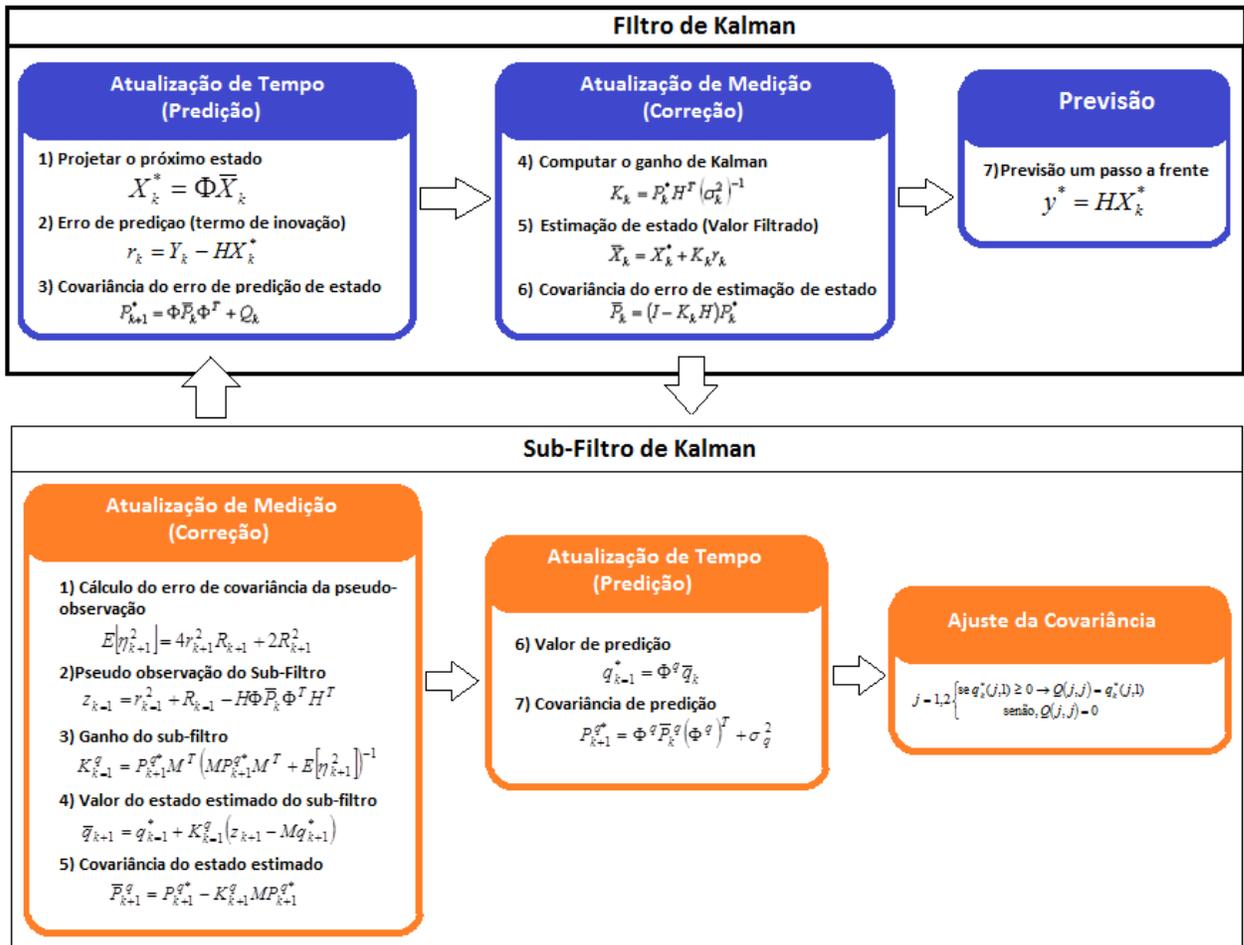


Figura 4.4. Algoritmo para cálculo do Filtro de Kalman com subfiltro.

Dessa forma, em vez de se delimitar um espaço estático dentro do qual a relação $Q_k R_k$, inicialmente calculada é válida, o método utilizando o subfiltro efetua uma reavaliação contínua de Q_k e R_k , realizando eventualmente ajustes nos valores que compõem as matrizes, de modo a manter a equivalência matemática da relação entre elas e, a cada novo passo essa relação seja sempre válida para os pontos do respectivo entorno.

5 CARACTERIZAÇÃO DE TRÁFEGO DA REDE REAL

Neste capítulo são abordadas as distribuições de probabilidade frequentemente utilizadas para representação de tráfego de rede de pacotes e o procedimento utilizado para obtenção das distribuições que caracterizam o tráfego de pacotes da rede real em estudo para ser utilizada no modelo de simulação.

5.1 DISTRIBUIÇÕES DE PROBABILIDADE

O comportamento de cada serviço pode diferir de uma rede para outra apresentando diferentes distribuições.

Para realizar simulações para validar o método de redução do consumo de energia é necessário obter previamente o comportamento da rede real, e conseguir gerar novamente o mesmo comportamento do tráfego observado por uma distribuição de probabilidade adequada que melhor represente o tráfego em questão.

A seguir são apresentadas as distribuições mais utilizadas em uma rede de dados e suas aplicações.

5.1.1 Distribuição Exponencial

A distribuição exponencial é utilizada quando se deseja analisar o espaço ou intervalo de acontecimento de um evento, utilizada para eventos contínuos. Ela calcula a probabilidade para determinado tempo e espaço entre eventos sucessivos, ocorrendo em um processo de Poisson.

Uma variável aleatória X é dita ter uma distribuição exponencial com parâmetro $\lambda > 0$ se sua Função Densidade Probabilidade (FDP) (Figura 5.1) é dada por [SPIEGEL, 2003]:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (5.1)$$

Sua Função Densidade Acumulada (FDA) é

$$F(x) = \begin{cases} 1 - \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (5.2)$$

A média e a variância são dadas pelas Equações (5.3) e (5.4), respectivamente:

$$E[x] = \frac{1}{\lambda} \quad (5.3)$$

$$Var(x) = \frac{1}{\lambda^2} \quad (5.4)$$

É comumente utilizado para tempos entre chegadas de serviço.

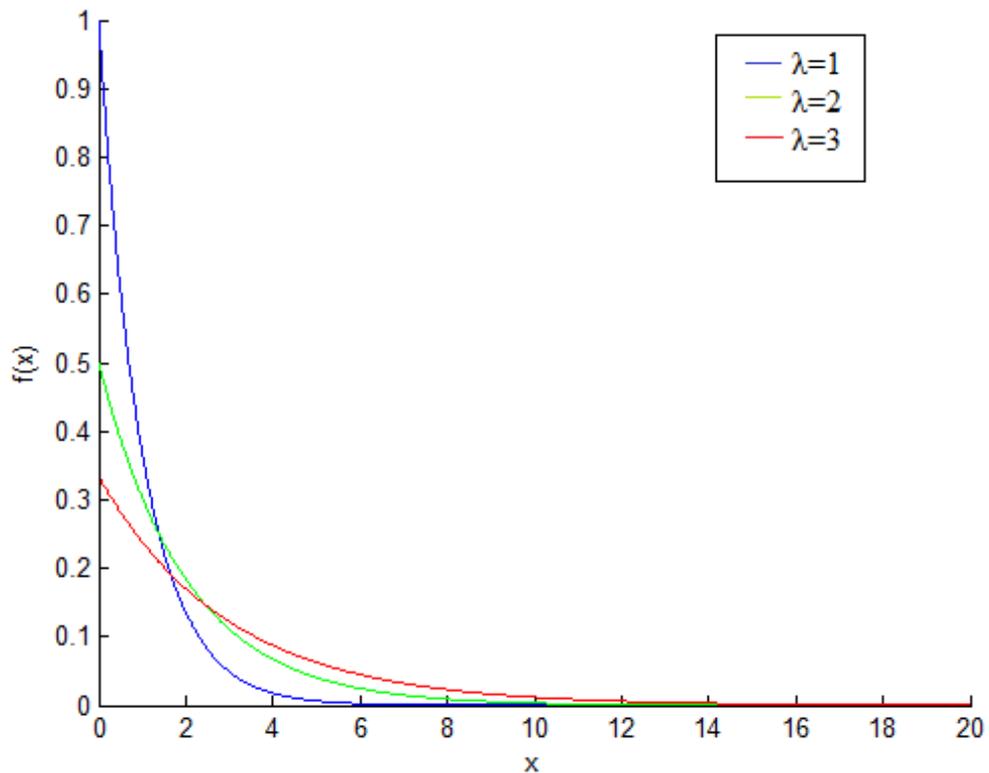


Figura 5.1. Função densidade de probabilidade Exponencial com diferentes valores de λ .

5.1.2 Distribuição de Poisson

A distribuição de Poisson é uma das mais usadas para variáveis aleatórias discretas. Ela determina a quantidade de ocorrências em um determinado intervalo de tempo. Os eventos devem ocorrer em determinado intervalo de tempo ou espaço.

A distribuição de Poisson (Figura 5.2) é dada por [SPIEGEL, 2003]:

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{para } x = 0, 1, 2, \dots \quad (5.5)$$

Sua Função Densidade Acumulada é:

$$F(x) = \begin{cases} e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} \quad (5.6)$$

Onde λ é a taxa média do processo.

O valor esperado e a variância de uma distribuição de Poisson são dados por:

$$E(X) = \lambda \quad (5.7)$$

$$Var(X) = \lambda \quad (5.8)$$

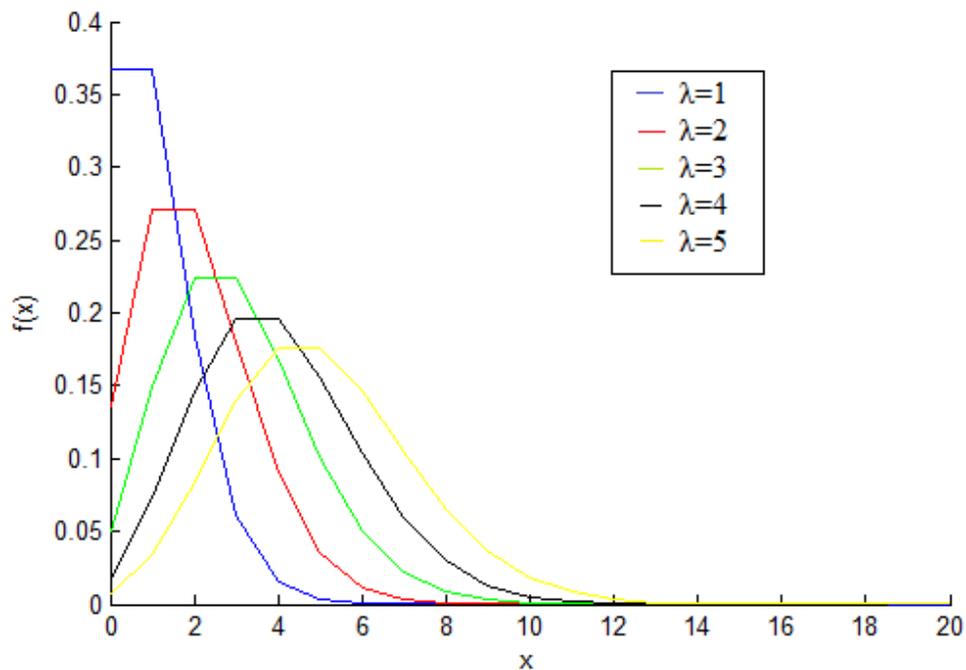


Figura 5.2. Função Densidade de Probabilidade Poisson para diferentes valores de λ .

Ela possui as seguintes características:

- Não possui memória: O número de eventos ocorrendo em um segmento de tempo ou espaço é independente do número de eventos ocorridos no segmento anterior.
- A taxa média do processo (λ) deve permanecer constante durante o período de tempo e espaço considerados.
- Quanto menor o segmento de tempo e espaço, menor a probabilidade de ocorrer mais de um evento naquele segmento. A probabilidade de ocorrência de dois ou mais eventos se aproxima de zero, quando o tamanho do segmento se aproxima de zero.

5.1.3 Distribuição lognormal

Sua função densidade de probabilidade (Figura 5.3) é expressa por [NETO, 2005]:

$$f(x) = \begin{cases} \frac{1}{(x-a)\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x-a)-\mu)^2}{2\sigma^2}\right) & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (5.9)$$

onde a é o limite inferior da amostra, μ é a média do logaritmo do tempo de falha e σ^2 é o desvio padrão, com valor esperado e variância determinados por:

$$E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad (5.10)$$

$$Var(x) = \exp(2\mu + \sigma^2)(e(\sigma^2) - 1) \quad (5.11)$$

Esta distribuição não apresenta Função Distribuição Acumulada definida.

Esta distribuição é muito utilizada para estimativa de tempo de reparo de um processo, ou seja, o tempo entre ocorrer uma falha e o tempo para que seja corrigida.

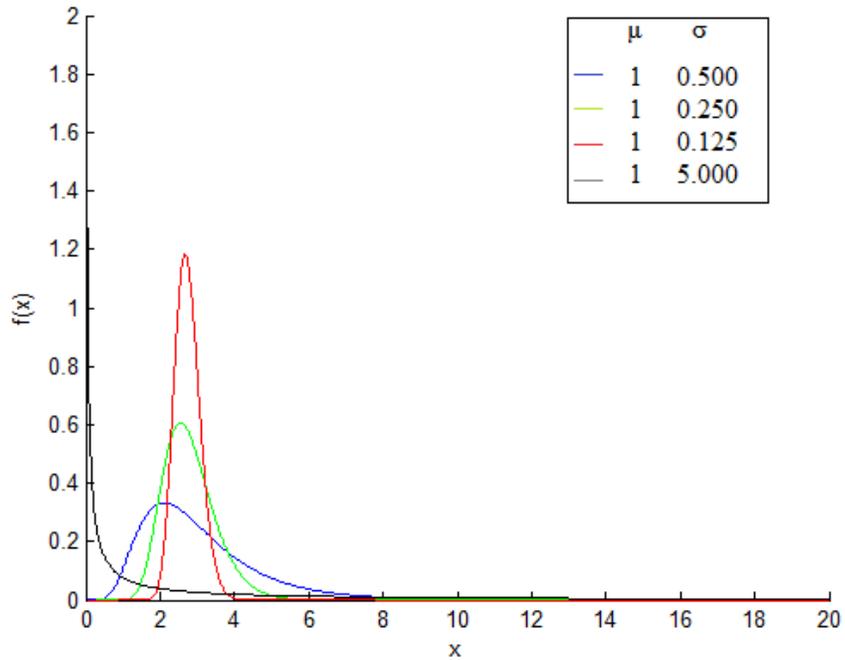


Figura 5.3. Função Densidade de Probabilidade lognormal para diferentes valores de μ e σ .

5.1.4 Distribuição Weibull

Sua função densidade de probabilidade (Figura 5.4) é expressa por [SPIEGEL, 2003]:

$$f(x) = \begin{cases} \beta \alpha x^{\beta-1} e^{-\alpha x^\beta} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (5.12)$$

onde α e β correspondem a taxa de falha, e são constantes positivas.

A função distribuição acumulada é dada por:

$$F(x) = \begin{cases} 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right] & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad (5.13)$$

O valor esperado e a variância são calculados por:

$$E(X) = \alpha^{-\frac{1}{\beta}} \Gamma\left[1 + \left(\frac{1}{\beta}\right)\right] \quad (5.14)$$

$$Var(X) = \alpha^{-2} \left[\Gamma \left[1 + \left(\frac{2}{\beta} \right) \right] - \Gamma^2 \left[1 + \left(\frac{1}{\beta} \right) \right] \right] \quad (5.15)$$

Ela é uma distribuição de probabilidade contínua, usada em estudos de tempo de vida de equipamentos e estimativa de falhas. Na área de redes de computadores, pode ser utilizada para representar o período *on* (ativo), de uma transmissão *web*, por exemplo.

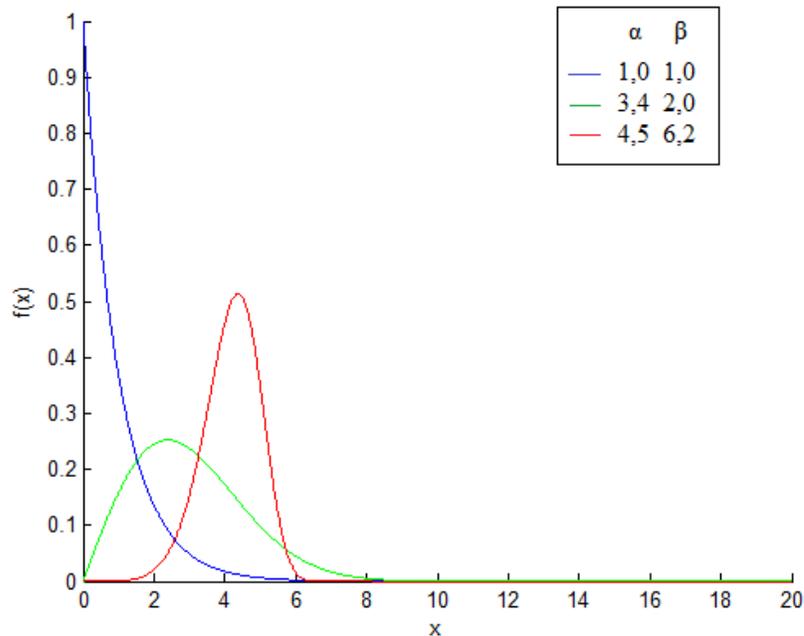


Figura 5.4. Função Densidade de Probabilidade Weibull para diferentes valores de α e β .

5.1.5 Distribuição de Pareto

Ela é uma distribuição de cauda longa, com variância infinita e valores grandes com probabilidade não desprezível.

Sua função de probabilidade (Figura 5.5) é dada pela Equação (5.16) [ASRABADI, 1990].

$$f(x) = \frac{\alpha K^\alpha}{x^{(\alpha+1)}} \text{ onde } \alpha > 0, K > 0, x \geq K. \quad (5.16)$$

A função densidade acumulada, esperança e variância são dadas pelas equações (5.17), (5.18) e (5.19), respectivamente.

$$F(x) = 1 - \left(\frac{K}{x} \right)^\alpha \text{ onde } \alpha > 0, K > 0, K \leq x < \infty \quad (5.17)$$

$$E(X) = \frac{\alpha K}{(\alpha - 1)} \text{ com } \alpha > 1 \quad (5.18)$$

$$\text{Var}(X) = \frac{\alpha K^2}{(\alpha - 1)^2(\alpha - 2)} \text{ com } \alpha > 2 \quad (5.19)$$

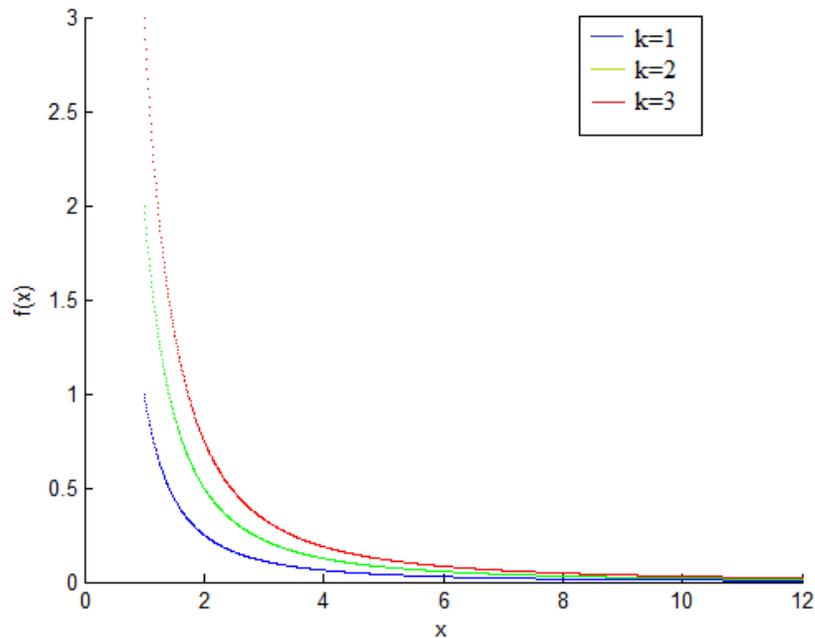


Figura 5.5. Função Densidade de Probabilidade Pareto para diferentes valores de k e $\alpha=1$.

Dentre as variáveis que assumem comportamento de distribuição Pareto, e que estão ligadas à análise de rede podem ser citados:

- Número de conexões em uma rajada de uma seção FTP
- Tamanho das rajadas em bytes
- Tamanho de arquivos
- Tempo de CPU consumido por processos
- Período ocioso (enquanto não há tráfego de dados)
- Tamanho de uma conexão WWW

Assim como as distribuições Weibull e lognormal, a distribuição de Pareto é um exemplo de distribuição de cauda longa (*heavy tailed*), o que indica o comportamento de tráfego com rajada de dados.

5.1.6 Distribuição Gamma

A distribuição Gamma tem sua função densidade (Figura 5.6) dada pela Equação (5.20) [SPIEGEL, 2003].

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases} \quad \text{onde } (\alpha, \beta) > 0 \quad (5.20)$$

Os parâmetros α e β correspondem à forma e a escala, respectivamente, e $\Gamma(\alpha)$ é a função Gamma, dada pela integral apresentada na Equação (5.21). A distribuição Gamma não apresenta função de distribuição acumulada [KAY, 2006].

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \quad (5.21)$$

A esperança e a variância são dadas pelas equações (5.22) e (5.23), respectivamente:

$$E(X) = \alpha\beta \quad (5.22)$$

$$Var(X) = \alpha\beta^2 \quad (5.23)$$

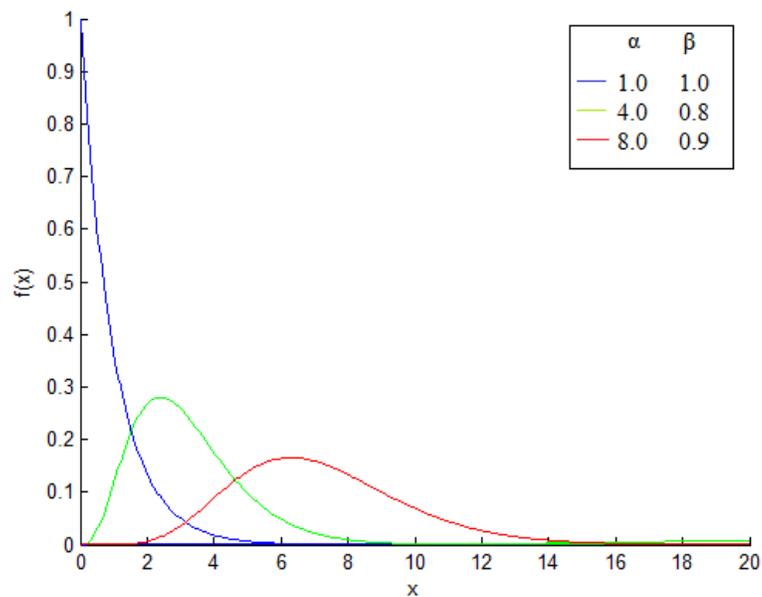


Figura 5.6. Função Densidade de Probabilidade Gamma para diferentes valores de α e β .

Aplica-se a distribuição Gamma à análise de tempo de vida de equipamentos, de tempo de retorno de mercadorias com falhas e a testes de confiabilidade e para modelar tempos de serviço e de reparo.

5.1.7 Distribuição Beta

A distribuição beta é uma distribuição de probabilidade contínua, função de densidade de probabilidade (Figura 5.7) é dada pela Equação (5.24) [SPIEGEL, 2003].

$$f(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{se } 0 < x < 1 \\ 0 & \text{caso contrário} \end{cases} \quad \text{e } (\alpha, \beta) > 0 \quad (5.24)$$

Os parâmetros α e β definem a forma da distribuição, respectivamente, e $B(\alpha, \beta)$ é a função Beta, dada pela integral da Equação (5.25):

$$B(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \quad (5.25)$$

Se $\alpha = \beta$, a distribuição é simétrica. Se $\alpha > \beta$, a assimetria é negativa e no caso de $\alpha < \beta$, sua assimetria é positiva.

A distribuição Beta não apresenta função de distribuição acumulada. Quando $\alpha = \beta = 1$, a densidade de Beta se reduz à Uniforme no intervalo (0,1).

A esperança e a variância da distribuição beta são dadas, respectivamente pelas equações (5.26) e (5.27):

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (5.26)$$

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad (5.27)$$

Ela é apropriada para modelar proporções aleatórias, por causa do seu domínio (o intervalo (0,1)) e também pela variedade de formas as quais pode assumir, de acordo com os valores especificados de α e β . Exemplos de proporções aleatórias são:

- Fração de pacotes que necessitam de retransmissões.
- Fração de Chamadas de Procedimentos Remotos (RPC) que levam mais do que um tempo determinado.

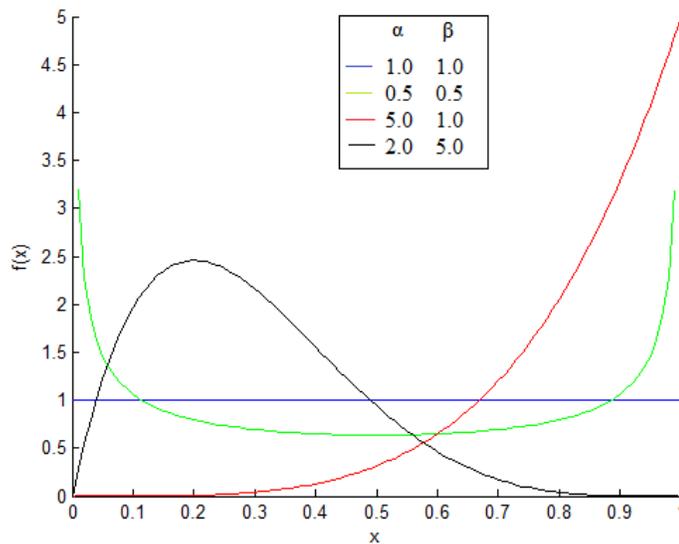


Figura 5.7. Função Densidade de Probabilidade Beta para diferentes valores de α e β .

5.1.8 Distribuição de Erlang

A função densidade probabilidade (Figura 5.8) da variável aleatória Erlang é dada por [KAY, 2006]:

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \text{ para } x, \lambda \geq 0 \quad (5.28)$$

Onde k representa a forma e λ a taxa, ambos os números são reais não negativos.

Uma alternativa para parametrização, porém equivalente, usa o parâmetro de escala μ , que é o inverso do parâmetro da taxa:

$$\mu = E(x) = \frac{1}{\lambda} \quad (5.29)$$

e a nova função densidade de probabilidade é dada por:

$$f(x) = \frac{x^{k-1} e^{-\frac{x}{\mu}}}{\mu^k (k-1)!} \text{ para } x, \mu \geq 0 \quad (5.30)$$

Quando o parâmetro k é igual a 1, a distribuição transforma-se na distribuição exponencial. A distribuição Erlang é um caso particular da distribuição Gamma, onde o parâmetro k é um inteiro.

A média e a variância de uma variável aleatória de Erlang são dadas respectivamente por:

$$E(x) = \frac{k}{\lambda} \quad (5.31)$$

$$Var(x) = \frac{k}{\lambda^2} \quad (5.32)$$

Esta distribuição pode ser utilizada para modelar tempos de serviço num modelo de rede de filas e tempo de reparo e tempo entre falhas. Também mede o tempo entre as chamadas recebidas e pode ser usado em conjunto com a duração prevista de chamadas telefônicas para produzir informações sobre o tráfego medido em Erlang unidades. Pode ser usado para determinar a probabilidade de perda de pacotes ou atrasos.

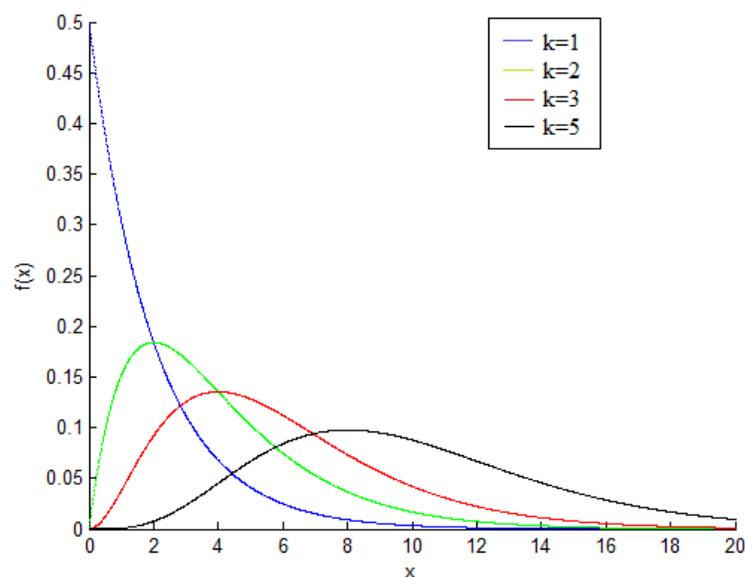


Figura 5.8. Função Densidade de Probabilidade de Erlang para $\lambda=0,5$ e diferentes valores de k .

5.2 AJUSTE DE DISTRIBUIÇÃO E GERAÇÃO DE VARIÁVEL ALEATÓRIA

Uma vez detectado o comportamento do tráfego, deve-se identificar a distribuição que melhor se ajusta a esses dados.

O ajuste de distribuições consiste em encontrar uma função matemática que represente de forma adequada uma variável estatística representada numa amostra: Tem-se em mãos uma observação de característica quantitativa X_1, X_2, \dots, X_n , como o intervalo de chegada de pacotes, ou tamanho dos pacotes recebidos, que são utilizados neste trabalho, e deseja-se testar se essas observações pertencem a uma população com uma Função Densidade de Probabilidade conhecida, como exemplo as distribuições citadas na seção anterior.

Os principais passos para ajuste de distribuição são:

- (1) Escolher o modelo/função: fazer hipóteses sobre uma família de distribuições
- (2) Estimar os parâmetros
- (3) Avaliação da qualidade do ajuste

Para estimar se os dados estão bem ajustados pela distribuição teórica, deve-se recorrer ao teste de aderência. Esse teste consiste basicamente de conduzir um teste de hipótese estatístico [SPIEGEL, 2003] para verificar se a amostra provém da distribuição teórica. Em geral, os testes podem ser enunciados da seguinte maneira:

H_0 : Os dados da amostra são provenientes da distribuição especificada

H_1 : Os dados da amostra não são provenientes da distribuição especificada

Os testes de aderência devem ser usados em complemento a análise gráfica, pois:

- Alguns testes pressupõem “amostras grandes” e, portanto, apresentam resultados aproximados;

- Para qualquer teste, não rejeitar a hipótese nula (H_0) não significa que ela seja verdadeira; pode ser que o tamanho da amostra seja muito pequeno para detectar diferenças entre a verdadeira distribuição (desconhecida) e aquela especificada no teste;

- Rejeitar a hipótese nula não significa que o modelo seja inadequado para propósitos práticos.

Para a análise gráfica, pode-se utilizar histogramas, estimativas de densidade, distribuição empírica, que podem sugerir o tipo de distribuição para ser utilizado no modelo. Os histogramas

podem fornecer uma intuição sobre assimetria, comportamento das caudas, presença de comportamento multimodal e *outliers*⁴.

Em [NETO, 2005] é apresentado um exemplo de ajuste de distribuição dos dados para distribuição Weibull. Foi utilizado o método da máxima verossimilhança [QUEIROZ, 2004] para determinar os valores de λ e β , e para teste de aderência foi utilizado o teste de hipótese pelo teste de qui-quadrado.

Com a distribuição e os parâmetros em mãos, o próximo passo é a geração dos dados.

Há uma série de métodos usados para gerar valores para variáveis aleatórias não-uniformes. Cada método é aplicável somente para uma parte da distribuição em questão. Ainda, para uma distribuição particular, um determinado método pode ser mais eficiente do que outros. Um dos métodos utilizados é a transformação inversa, por exemplo.

Este método baseia-se na observação de que, para uma dada variável aleatória x , com uma FDA $F(x)$, a variável $u = F(x)$ é uniformemente distribuída entre 0 e 1. Assim, os valores de x podem ser obtidos gerando-se números aleatórios uniformemente distribuídos e calculando-se $x = F^{-1}(u)$, conforme pode ser visto na Figura 5.9. A prova dessa observação é demonstrada em [JAIN, 1991].

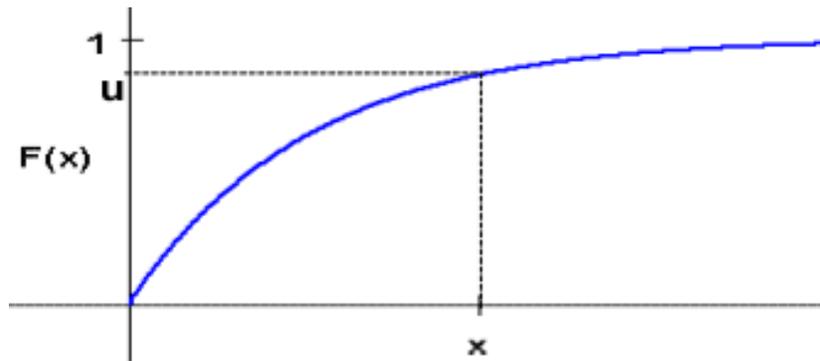


Figura 5.9. Transformação inversa da FDA.

Tomando como exemplo a distribuição exponencial, dado um determinado valor x , e uma taxa média de chegada λ , calcula-se a probabilidade de ocorrência de x ou ainda a probabilidade acumulada $P(X \leq x)$.

⁴ Uma observação que está fora do padrão de uma distribuição [MOORE, 1999]. A presença de *outliers* indica algum tipo de problema, como não ajustar a distribuição em estudo ou um erro na medição.

Dada a probabilidade $P(X \leq x) = F(x)$, pode-se calcular o valor correspondente de x usando a transformação inversa.

Partindo da FDA da distribuição exponencial tem-se que:

$$F(x) = 1 - \lambda e^{-\lambda x} = u \quad (5.33)$$

Desse modo:
$$x = -\frac{1}{\lambda} \ln(1 - u) \quad (5.34)$$

Assim, valores para a variável aleatória x podem ser produzidos através da geração de uma variável u , uniforme, e usando a equação (5.34) para determinar x .

O software ARENA, utilizado neste trabalho para realização das simulações dispõe de ferramentas que facilitam a análise dos dados, fazendo os ajustes de distribuição automaticamente, como será visto na seção seguinte. Ele também realiza a geração aleatória de valores de acordo com cada distribuição, onde é necessário passar apenas os parâmetros das distribuições.

5.3 TIPOS DE TRÁFEGO

No início das redes de computadores, o tráfego era muito baixo e era constituído basicamente de troca de arquivos e e-mail, chamado tráfego elástico. Com o avanço da tecnologia das redes, processadores, e surgimento de novos serviços multimídia, o tráfego predominante passou a ser de voz, vídeo e áudio *streaming*⁵, constituindo o tráfego *stream*.

A partir disso, pode-se dividir o tráfego de dados em dois principais tipos:

- Tráfego elástico: é a denominação utilizada para aplicações não interativas (como navegação *web*, *FTP*(*File Transfer Protocol*) e *e-mail*), onde o ponto crítico está na integridade da informação e não no atraso da entrega dos pacotes, demandando apenas requisitos de vazão.

- Tráfego *Stream*: é uma prática usual para denominar aplicações interativas que são exigentes quanto ao tempo (tempo real). Essas aplicações demandam de garantia de vazão e

⁵ Conhecido por tráfego inelástico, pois não se adaptam as variações de atrasos (elasticidade), que é a principal característica do tráfego do tipo elástico. A denominação *Streaming* é proveniente da característica do fluxo de dados presentes no tipo de transmissão dos pacotes que são praticamente constantes durante todo tempo de envio.

atraso. É levada em consideração para seu dimensionamento a probabilidade de perda de pacotes e probabilidade de bloqueio.

A Figura 5.10 ilustra os tipos de tráfego e exemplos de aplicações.

Com a evolução dos dispositivos móveis e crescimento do número de aplicações, os dois tipos de tráfego relacionados acima são suportados e podem ser trafegados entre rede e dispositivo.

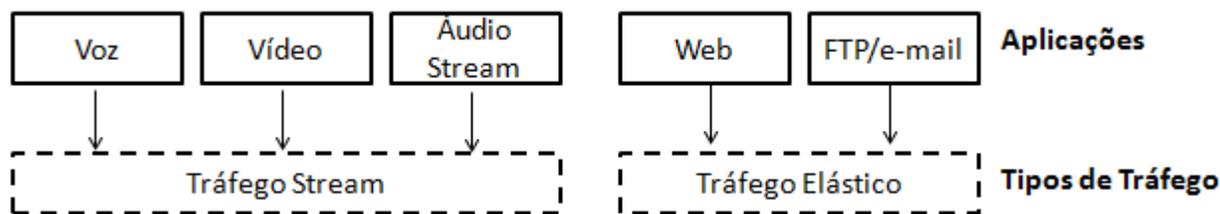


Figura 5.10. Classificação de tráfego [PINOTTI, 2011].

Para um tráfego de tempo real (*stream*), o dispositivo permanece em Cell_DCH, pois necessita de um canal dedicado devido ao alto tráfego gerado. Para um tráfego elástico, devido à característica de recebimento de pacotes a determinado intervalo de tempo, o canal predominante nesse estado é o Cell_FACH, onde é utilizado um canal compartilhado para transmissão de dados [LEE-CHEN, 2004].

Em [LEE-CHEN, 2004] é apresentada a caracterização de tráfego para um tráfego *web* (elástico), que pode ser visto na Tabela 5.1, e vídeo (*stream*), que está apresentado na Tabela 5.2.

TABELA 5.1. CARACTERIZAÇÃO DE TRÁFEGO DE NAVEGAÇÃO WEB [LEE-CHEN, 2004].

Parâmetros Observados	Distribuição
Processo de chegadas entre sessões	Geométrica com média 600 segundos
Número de requisições de pacotes por sessão	Geométrica com média 5 segundos
Tempo de leitura entre dois pacotes consecutivos em uma sessão	Geométrica com média 120 segundos
Numero de pacotes em uma requisição	Geométrica com média 25 segundos
Intervalo de tempo entre 2 pacotes consecutivos	Geométrica com média 0,0625 segundos
Tamanho do pacote	Pareto com $\alpha = 81,5$ e $K = 1.1$.

TABELA 5.2. CARACTERIZAÇÃO DE TRÁFEGO DE VÍDEO STREAMING [LEE-CHEN, 2004].

Parâmetros Observados	Distribuição
Duração	Exponencial com média 30 segundos
Intervalo de acesso entre dois vídeos sucessivos	Exponencial com média 120 segundos

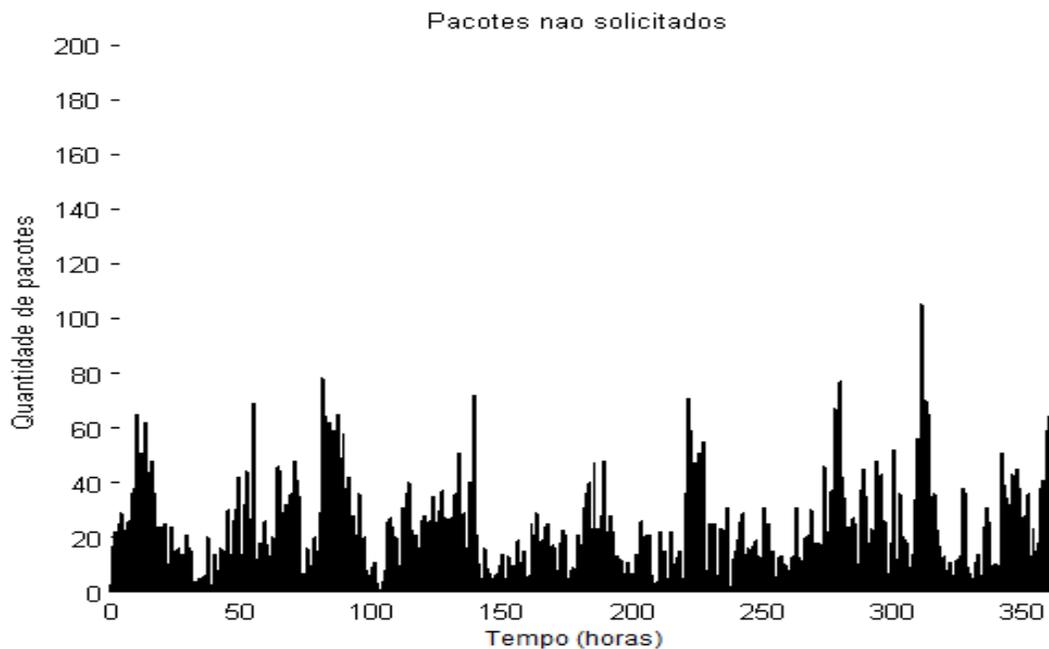
No entanto, quando o dispositivo móvel está no modo *Stand By*, o maior tráfego responsável pelo aumento do consumo no dispositivo nesse período é o tráfego elástico, principalmente os serviços de *e-mail* e pacotes não solicitados. Para verificar o comportamento desses dois tipos de tráfego foi feito o monitoramento de chegada de pacotes de um dispositivo móvel. Utilizando a ferramenta *Input Analyzer* presente no software *Arena*, foi possível fazer o ajuste da melhor distribuição e parâmetros para os processos de intervalo de chegada de pacotes, tamanho dos pacotes e distribuição do fluxo de pacotes. No caso das mensagens *Keep-Alive*, quando existente, o intervalo é constante e não há necessidade de se fazer nenhum ajuste de distribuição.

5.4 MONITORAMENTO DO TRÁFEGO REAL DE PACOTES NÃO SOLICITADOS E E-MAIL

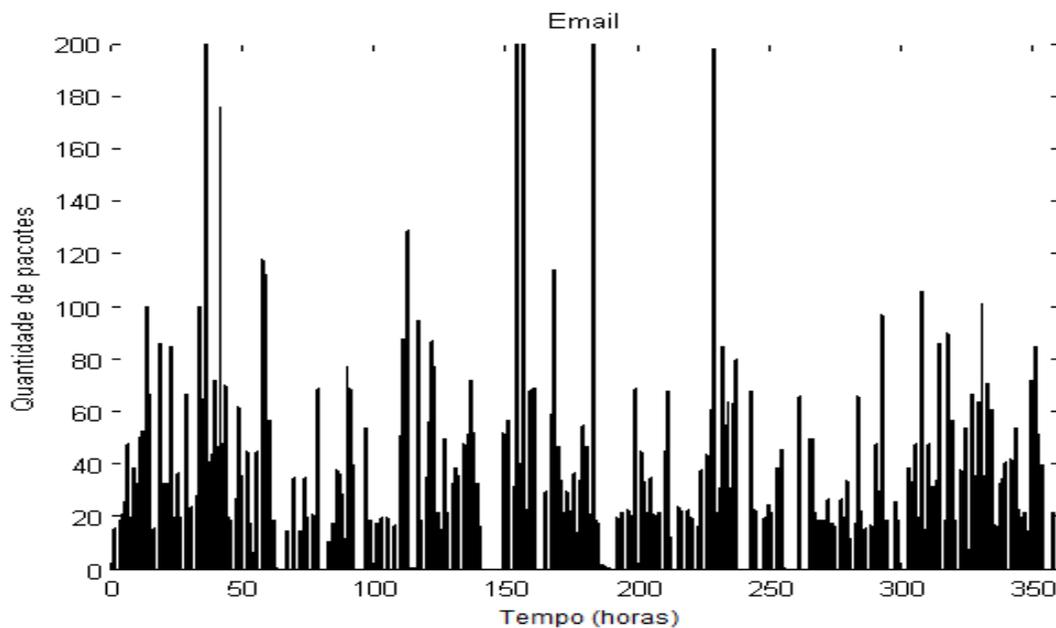
O monitoramento dos pacotes não solicitados foi feito levando em consideração a operadora que apresenta o maior recebimento desse tipo de informação. Para o tráfego de e-mails foi utilizada uma conta de e-mail de um usuário escolhido aleatoriamente.

Os dados de interesse para obtenção das distribuições são o intervalo de chegada de pacotes e o tamanho dos pacotes. Os dados foram coletados utilizando um dispositivo móvel que roda o sistema operacional *Android Always On*. Para obtenção dos pacotes trafegados, foram utilizados os recursos da biblioteca *TcpDump* para Linux, porém adaptada para o sistema operacional *Android* [TCPDMP]. O procedimento utilizado para coleta dos pacotes no dispositivo móvel é apresentado em [ADBTCPDMP]. Os dados obtidos são salvos na extensão “.pcap” que podem ser visualizados pelo do programa *Wireshark* [WSHK] .

A Figura 5.11 apresenta os gráficos para o período de 360 horas de coleta de pacotes para e-mail e pacotes não solicitados, respectivamente.



(a)



(b)

Figura 5.11. Distribuição dos pacotes ao longo do tempo para (a) pacotes não solicitados e (b) e-mail.

Feito a coleta dos dados, eles foram inseridos na ferramenta Input Analyzer. Utilizando o comando “Fit All”, foi possível ver todas as tentativas de ajustes de distribuição feita pelo programa e o erro de ajuste, como mostra a Figura 5.12(a), que apresenta o exemplo de ajuste de distribuição para o intervalo de chegada de pacotes não solicitados.

Quanto menor o erro quadrático, mais representativo é a distribuição para essa característica de tráfego. Nesse exemplo, as distribuições que melhor se adequaram ao tráfego de pacotes não solicitados foram a Weibull e Beta. A Figura 5.12(b) apresenta os parâmetros obtidos para uma determinada distribuição após o ajuste.

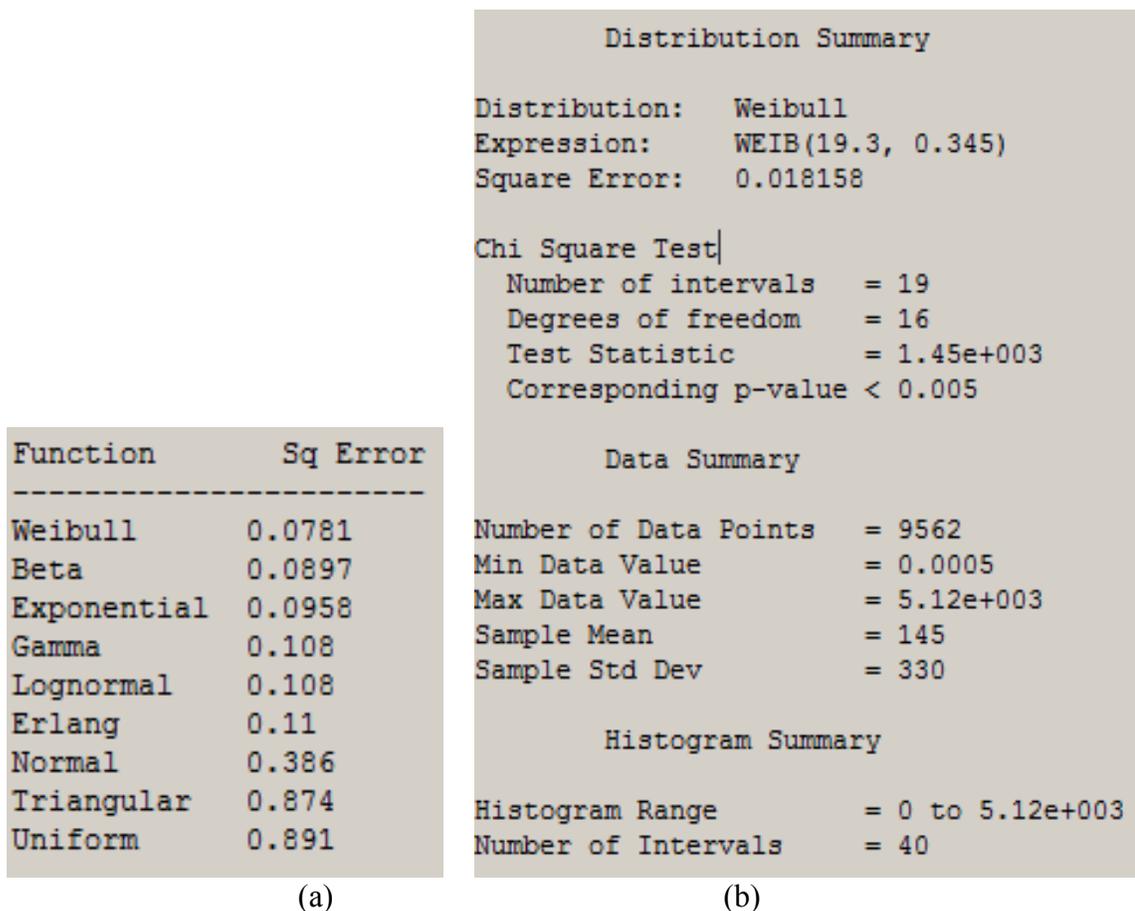


Figura 5.12. (a) Resultados do ajuste de distribuições utilizando o “Fit All” do “Input Analyzer” e (b) Parâmetros obtidos após ajuste para determinada distribuição.

O parâmetro “Expression” apresenta a expressão que representa o comportamento dos dados. Essa expressão será utilizada no bloco “ARRIVE” do modelo ARENA (tratado com mais detalhes no Capítulo 6), que irá gerar a entrada de dados de acordo com o comportamento dos dados reais obtidos.

Em *Chi-SquareTest* é possível ver o teste de aderência utilizando o método do qui-quadrado, o qual *Corresponding p-value* corresponde ao nível de significância do teste de hipótese.

A Figura 5.13 apresenta a correspondência gráfica entre os dados reais (gráfico de barras) e a distribuição ajustada (gráfico linear), correspondente a análise gráfica abordada na seção anterior.

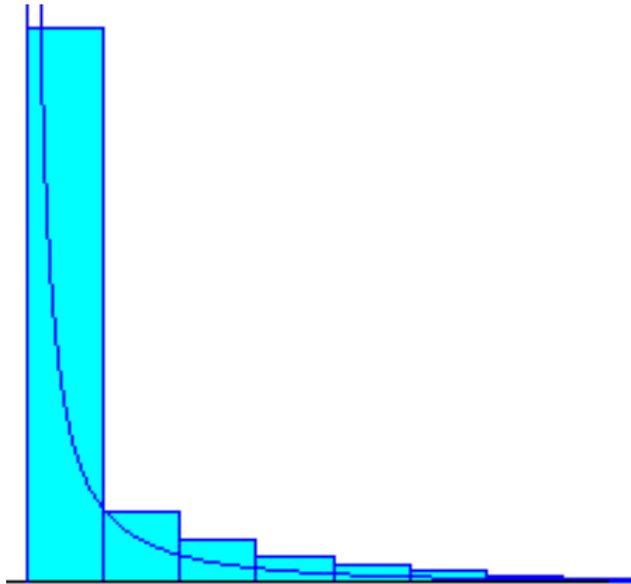


Figura 5.13. Correspondência gráfica entre os dados reais (gráfico de barra) e a distribuição ajustada (gráfico de linha) gerada pelo Input Analyzer.

O ajuste foi feito para o período total de 360 horas de observação, e também para diferentes períodos do dia, divididos de quatro em quatro horas, como mostra a Tabela 5.3.

TABELA 5.3. AJUSTE DE DISTRIBUIÇÃO PARA DIFERENTES TIPOS DE TRÁFEGO DA REDE REAL.

Período	Pacotes	Serviço	
		E-mail	Pacotes não solicitados
Período Completo	Intervalo de Chegada	Distribution: Weibull Expression: $-1 + WEIB(8.89, 0.361)$ Square Error: 0.000341	Distribution: Weibull Expression: $WEIB(19.3, 0.345)$ Square Error: 0.018158
	Tamanho	Distribution: Weibull Expression: $40 + WEIB(102, 0.377)$ Square Error: 0.050158	Distribution: Weibull Expression: $28 + WEIB(29.1, 1.63)$ Square Error: 0.078086
0h às 4h	Intervalo de Chegada	Distribution: Weibull Expression: $WEIB(1.53, 0.245)$ Square Error: 0.000470	Distribution: Weibull Expression: $WEIB(38.7, 0.307)$ Square Error: 0.006827
	Tamanho	Distribution: Weibull Expression: $40 + WEIB(97.6, 0.351)$ Square Error: 0.024874	Distribution: Erlang Expression: $28 + ERLA(7.43, 4)$ Square Error: 0.063990
4h às 8h	Intervalo de Chegada	Distribution: Weibull Expression: $WEIB(1.53, 0.246)$ Square Error: 0.000656	Distribution: Weibull Expression: $WEIB(31.9, 0.313)$ Square Error: 0.006852
	Tamanho	Distribution: Weibull Expression: $40 + WEIB(130, 0.393)$ Square Error: 0.028614	Distribution: Gamma Expression: $28 + GAMM(12.2, 2.48)$ Square Error: 0.200602
8h às 12h	Intervalo de Chegada	Distribution: Weibull Expression: $WEIB(1.51, 0.275)$ Square Error: 0.000060	Distribution: Beta Expression: $40 + 1.36e+003 * BETA(0.169, 0.129)$ Square Error: 0.068810
	Tamanho	Distribution: Weibull Expression: $WEIB(26.8, 0.364)$ Square Error: 0.007983	Distribution: Weibull Expression: $28 + WEIB(28.7, 1.61)$ Square Error: 0.053604
12h às 16h	Intervalo de Chegada	Distribution: Weibull Expression: $WEIB(2.02, 0.271)$ Square Error: 0.000214	Distribution: Weibull Expression: $WEIB(31.2, 0.382)$ Square Error: 0.007208
	Tamanho	Distribution: Weibull Expression: $40 + WEIB(90.2, 0.388)$ Square Error: 0.035811	Distribution: Erlang Expression: $28 + ERLA(9.91, 3)$ Square Error: 0.154214
16h às 20h	Intervalo de Chegada	Distribution: Exponential Expression: $-1 + EXPO(118)$ Square Error: 0.000500	Distribution: Beta Expression: $3.78e+003 * BETA(0, 0)$ Square Error: 0.012465
	Tamanho	Distribution: Weibull Expression: $40 + WEIB(234, 0.421)$ Square Error: 0.063797	Distribution: Lognormal Expression: $28 + LOGN(30.6, 23.7)$ Square Error: 0.116273
20h às 0h	Intervalo de Chegada	Distribution: Weibull Expression: $WEIB(2.3, 0.271)$ Square Error: 0.000432	Distribution: Weibull Expression: $40 + WEIB(89.1, 0.397)$ Square Error: 0.032630
	Tamanho	Distribution: Weibull Expression: $40 + WEIB(89.1, 0.397)$ Square Error: 0.032630	Distribution: Lognormal Expression: $28 + LOGN(29.5, 22.2)$ Square Error: 0.124055

6 MODELOS DE SIMULAÇÃO

Para criação do modelo que simule o recebimento de pacotes e a troca de estados do RRC foi utilizado o software de simulação de eventos discretos ARENA.

Foram construídos dois modelos. O primeiro representa o funcionamento da máquina de estados dos recursos de rádio de uma rede real, e o segundo modelo é baseado no primeiro, porém possui a implementação do filtro de Kalman para decisão de troca de estados.

Os modelos foram feitos inicialmente separados para facilitar o processo de validação, mas no modelo completo, todas as funcionalidades estão integradas.

6.1 MODELO DE SIMULAÇÃO I – FUNCIONAMENTO DA MÁQUINA DE ESTADOS DO RRC DE ACORDO COM A REDE REAL

O modelo é formado por dois blocos principais e é apresentado na Figura 6.1. O primeiro está relacionado à geração de pacotes de acordo com as características de tráfego desejadas. Neste bloco é possível configurar a distribuição de probabilidade que melhor se ajusta com o tráfego monitorado, configurando parâmetros como intervalo de chegada e tamanho médio dos pacotes, representado pelos blocos ARRIVE.

O segundo bloco está relacionado à máquina de estados do RRC, e contém toda a inteligência de troca de estados de acordo com o fluxo de pacotes, e valores de $T1$ e $T2$. O fluxo gerado aleatoriamente para cada tipo de serviço é inserido em um único canal de tráfego, representado pelo bloco ENTER, e é direcionado para o bloco PROCESS, que representa a pilha de pacotes equivalente ao *Buffer Occupation*. O bloco CHOOSE direciona os pacotes para o estado adequado de acordo com o volume de dados, equivalente aos eventos 4a e 4b do *Measurement Report*. Nesse modelo são implementadas as transições de Cell_DCH para Cell_FACH, Cell_FACH para Cell_DCH e Cell_FACH para Idle, que foi o comportamento observado nas três principais operadoras do Brasil quando ocorre a chegada de um pacote. No entanto, esse modelo também cobre os casos dos estados Cell_PCH e URA_Idle em termos de consumo, uma vez que possuem o mesmo valor que o estado Idle.

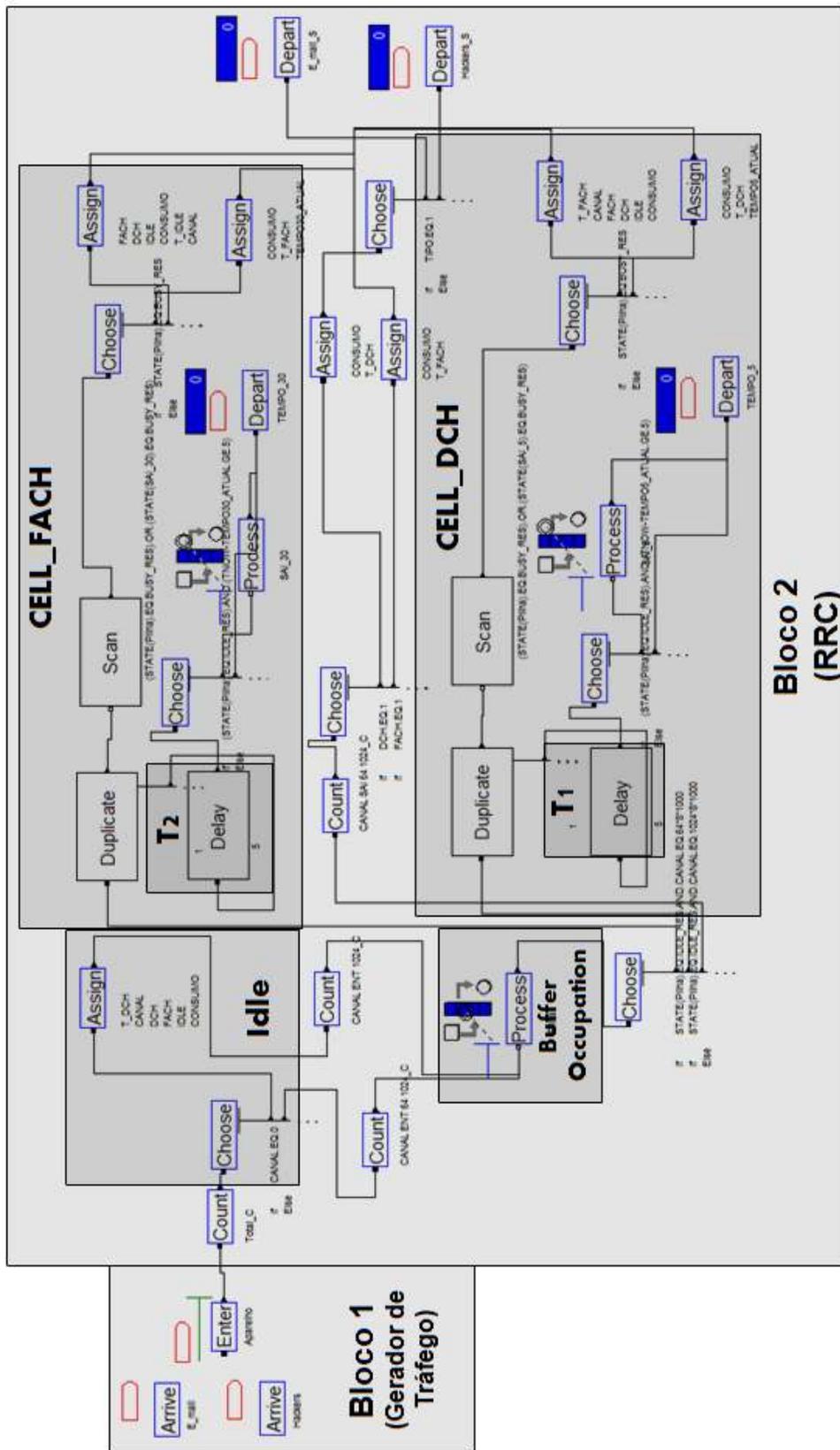


Figura 6.1. Modelo de simulação baseado na rede real. [TITO, 2011]

A cada transição de estado é iniciado o temporizador $T1$ ou $T2$, de acordo com a transição efetuada, quando o tráfego é cessado. Os blocos DELAY são responsáveis pela contabilização dos tempos de inatividade. Caso um novo tráfego chegue antes do término do temporizador ele é zerado e reiniciado assim que o fluxo for cessado novamente. Caso contrário, quando o temporizador atinge seu valor limite, a troca de estado acontece.

Para criação de duas instâncias paralelas o qual se contabiliza $T1$ ou $T2$ e, ao mesmo tempo, monitora o fluxo de pacotes que chegam ao sistema (*Buffer Occupation*), para o caso de pacotes que cheguem antes de cessar o temporizador, foi utilizado o bloco DUPLICATE. Ele cria uma cópia do processo, e caso seja detectado um novo fluxo entrante pelo bloco SCAN, ele cessa o temporizador para $T1$ ou $T2$ que foi iniciado, e, de acordo com o fluxo de dados recebido, pode permanecer no mesmo estado ou ir para outro estado (eventos 4a e 4b).

Os blocos ASSIGN são responsáveis por atribuir o estado adequado no sistema, e também por calcular o consumo de energia, de acordo com a máquina de estado. Os blocos DEPART contabilizam o número de serviços gerados aleatoriamente no sistema pelos blocos ARRIVE.

O cálculo do consumo irá obedecer fórmula dada pela Equação 6.1 [TITO, 2011]:

$$\text{Consumo} = \frac{V_{\text{bat}}}{T_{\text{sim}}} (5 \cdot T_{\text{Idle}} + 100 \cdot T_{\text{Cell_FACH}} + 200 \cdot T_{\text{Cell_DCH}}) \text{ [mWh]} \quad (6.1)$$

sendo que:

5, 100 e 200 :	valores médios de corrente para cada estado [mA]
T_{Idle} :	tempo de permanência no estado <i>Idle</i> [s]
$T_{\text{Cell_FACH}}$:	tempo de permanência no estado <i>Cell_FACH</i> [s]
$T_{\text{Cell_DCH}}$:	tempo de permanência no estado <i>Cell_DCH</i> [s]
T_{Sim}	tempo de simulação [s]
V_{bat}	tensão aplicada da bateria do dispositivo móvel [V]

6.1.1 Validação e resultados das simulações

Para validação do modelo de simulação, foi injetado um tráfego no sistema cujo intervalo de chegada de pacotes segue uma distribuição exponencial. O tamanho dos pacotes também foi

considerado fixo por hora, ou seja, o fluxo de dados varia apenas de acordo com o intervalo de chegada, para simplificação do modelo de validação.

Foram simuladas quatro condições de tráfego: baixa, média, alta e muito alta, para valores de $T2$ igual a 5 segundos e 30 segundos, como mostra a Tabela 6.1 . O critério utilizado para classificação do tráfego é o intervalo de chegada em relação ao maior tempo de inatividade configurado na rede.

Para um tráfego baixo, médio, e alto pode-se notar uma redução de consumo, enquanto para um tráfego muito alto, o qual o intervalo de chegada é bem menor que $T2$, o canal permanece alocado quase 100% do tempo em Cell_DCH, pois o recebimento de pacotes é praticamente constante.

No caso de tráfego alto, outros valores de tempo de inatividade foram utilizados a fim de validar o simulador. Os resultados obtidos se mostraram bem condizentes com a literatura [HAVERINEN, 2004] e [ARJONA, 2007], onde o consumo aumenta à medida que $T2$ aumenta. Desta forma, é possível validar a máquina de estado implementada no simulador ARENA, que está de acordo com o funcionamento da rede real. No entanto, para um alto tráfego, nota-se que nem sempre um valor menor de $T2$ gera um consumo menor. Pode acontecer casos em que um $T2$ menor resulte num consumo maior, se compararmos por exemplo um $T2=9s$ e um $T2=8s$ na Figura 6.2.

Na Figura 6.2 são apresentadas as curvas características de consumo para diferentes valores de $T2$ para baixo e alto tráfego. Os consumos obtidos para um $T2$ de 30s para baixo tráfego e 50s para alto tráfego corresponderam aos obtidos nas medições reais para duas operadoras distintas.

TABELA 6.1. RESULTADOS DA SIMULAÇÃO NO ARENA PARA BAIXO, MÉDIO E ALTO TRÁFEGO.

[TITO,2011]

Tráfego	Intervalo de chegada de pacotes	Consumo (mWh)		Redução (%)
		T2 = 5s	T2 = 30s	
Baixo	e-mail: 1 a cada 1500 s hacker: 1 a cada 500 s	163,63	173,71	5,81%
Médio	e-mail: 1 a cada 360 s hacker: 1 a cada 120 s	206,69	217,81	5,11%
Alto	e-mail: 1 a cada 60 s hacker: 1 a cada 20 s	289,31	341,96	15,40%
Muito Alto	e-mail: 1 a cada 0.5 s hacker: 1 a cada 1.5 s	739,99	739,99	0,00%

O período simulado foi de 14400s (4 horas) para cada valor de $T2$, sendo realizadas 50 replicações para cada simulação. Se fosse realizado em um ambiente controlado na rede real, o tempo mínimo para realizar todas as medições seria de aproximadamente 10.000h, sem levar em consideração o tempo para preparar o dispositivo móvel com o carregamento de novo software com $T2$ modificado. Utilizando o modelo de simulação, foi possível estimar o consumo para um intervalo de $T2$ de 1s a 60s em aproximadamente 40 minutos.

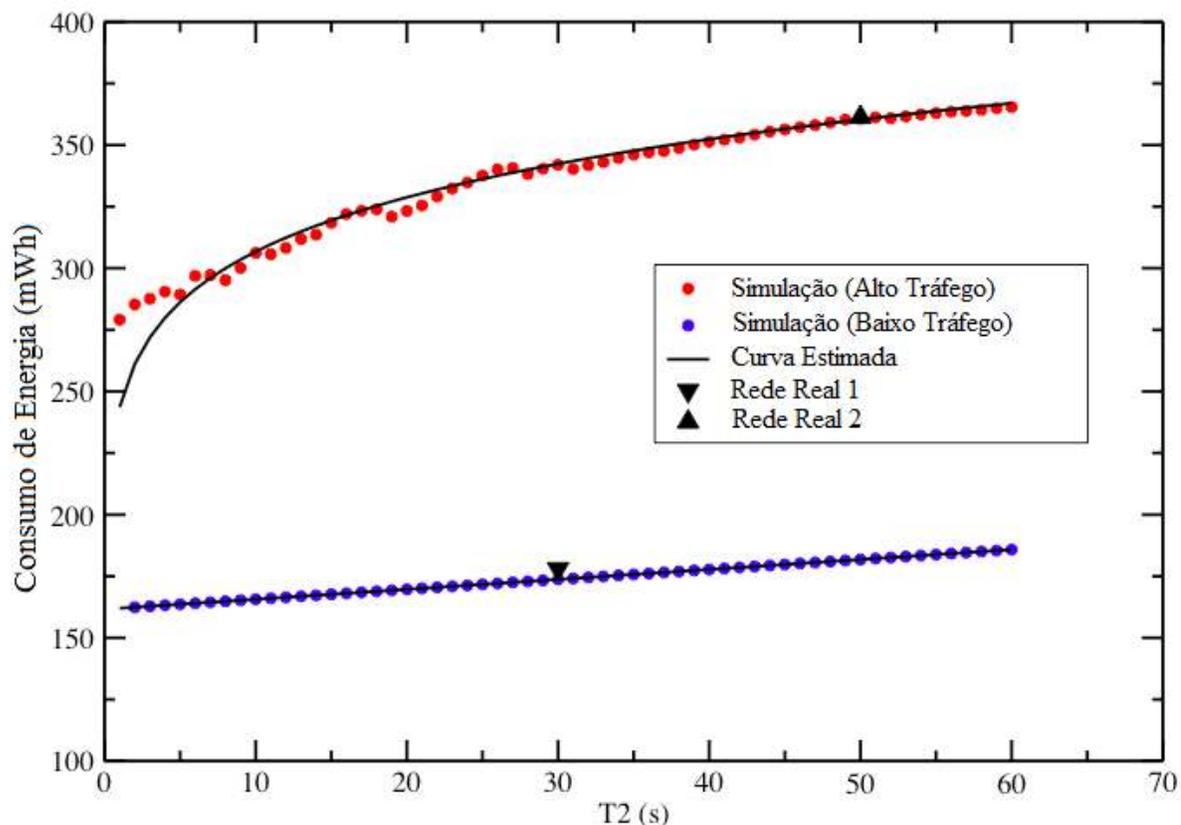


Figura 6.2. Relação entre diferentes valores de $T2$ para um alto tráfego e o consumo de energia.

[OLIVEIRA, 2011]

6.2 MODELO DE SIMULAÇÃO II – FUNCIONAMENTO DA MAQUINA DE ESTADOS DE ACORDO COM A REDE REAL INCLUINDO MODELO DE PREVISÃO

6.2.1 Modelo II-A - Filtro de Kalman

Na Figura 6.3 é apresentado um módulo referente à implementação do modelo do filtro de Kalman.

Esse modelo tem por objetivo a validação do modelo do filtro e subfiltro de Kalman antes de ser incluído no modelo de simulação da rede real e testar a eficiência do filtro para previsão de tráfego e pode ser dividido em dois grandes blocos:

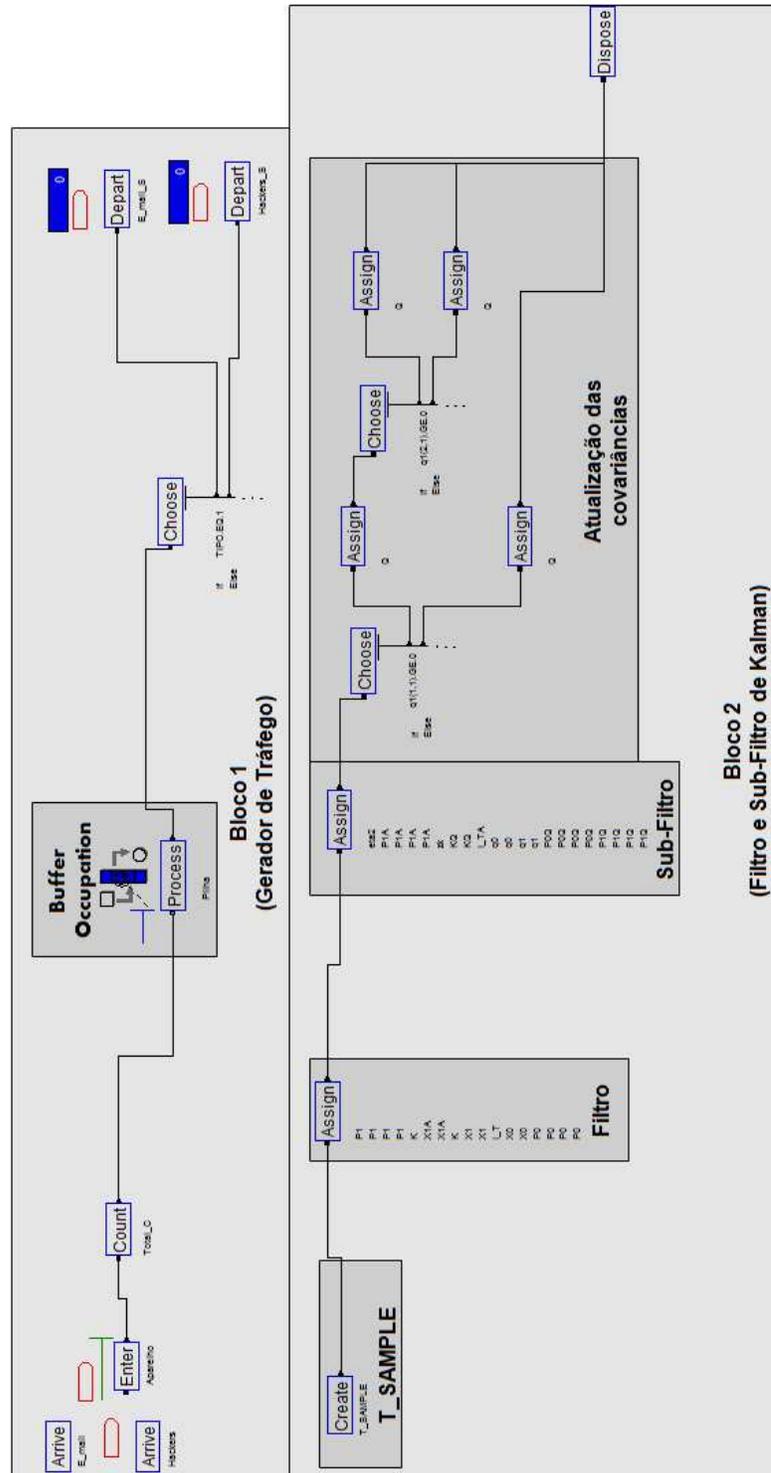


Figura 6.3. Modelo de simulação do Filtro de Kalman

O bloco 1 apresenta a chegada do fluxo de pacotes de acordo com uma distribuição pré-determinada, representando os serviços de e-mail e pacotes não solicitados. Os fluxos gerados são, por sua vez, inseridos em um mesmo canal, para que seja computada a fila do processo (*Buffer Occupation*), representado pelo bloco PROCESS, e os blocos DEPART contabilizam os pacotes gerados para cada tipo de tráfego. Esse é o processo de geração de pacotes similar ao utilizado no Modelo I.

O bloco 2 apresenta a implementação do filtro e subfiltro de Kalman. O bloco CREATE indica o tempo de amostragem (T_SAMPLE) da pilha dos pacotes recebidos, ou seja, de quanto em quanto tempo se deve monitorar a fila a fim de se ter o valor de observação Y_k . Esse tempo de amostragem é responsável também por indicar o tempo da próxima previsão, ou seja, a previsão será feita para o próximo instante $t_{atual} + T_SAMPLE$ segundos. O primeiro bloco ASSIGN é responsável pela implementação do filtro de Kalman, e consiste da máquina de estado apresentada na Figura 4.4 (Filtro de Kalman). O segundo bloco ASSIGN é responsável pelo cálculo do subfiltro de Kalman e ajuste dos erros de observação para evitar divergência do modelo, de acordo com a máquina de estado apresentada na Figura 4.4 (subfiltro de Kalman). Por último, é realizada a atualização da covariância do vetor do erro de modelagem Q_k .

Três principais variáveis são utilizadas no modelo:

- NQ(Pilha), que corresponde à observação (Y_k)
- X1A, que corresponde à previsão no instante k, feita a partir de k-1 (X_k^*)
- X0, que é o valor da filtragem feita no instante k (\bar{X}_k)

6.2.1.1 Simulações e Resultados

Uma distribuição exponencial foi configurada no modelo a fim de gerar um perfil de tráfego. Foi observada a saída gerada pelo bloco ASSIGN e o valor predito pelo filtro, a fim de se comparar sua eficiência. O filtro de Kalman se mostrou muito eficiente para previsão como pode ser observado na Figura 6.4. Os valores de previsão (vermelho) ficaram muito próximos do valor real medido (azul). No entanto, a previsão será mais precisa quanto menor for o valor de T_SAMPLE , como mostra as figuras 6.4(a) e 6.4(b). Para um valor maior, apesar de não ser tão

precisa, a previsão fornece o padrão de comportamento da rede, o que é suficiente para saber se haverá um aumento ou diminuição do tráfego.

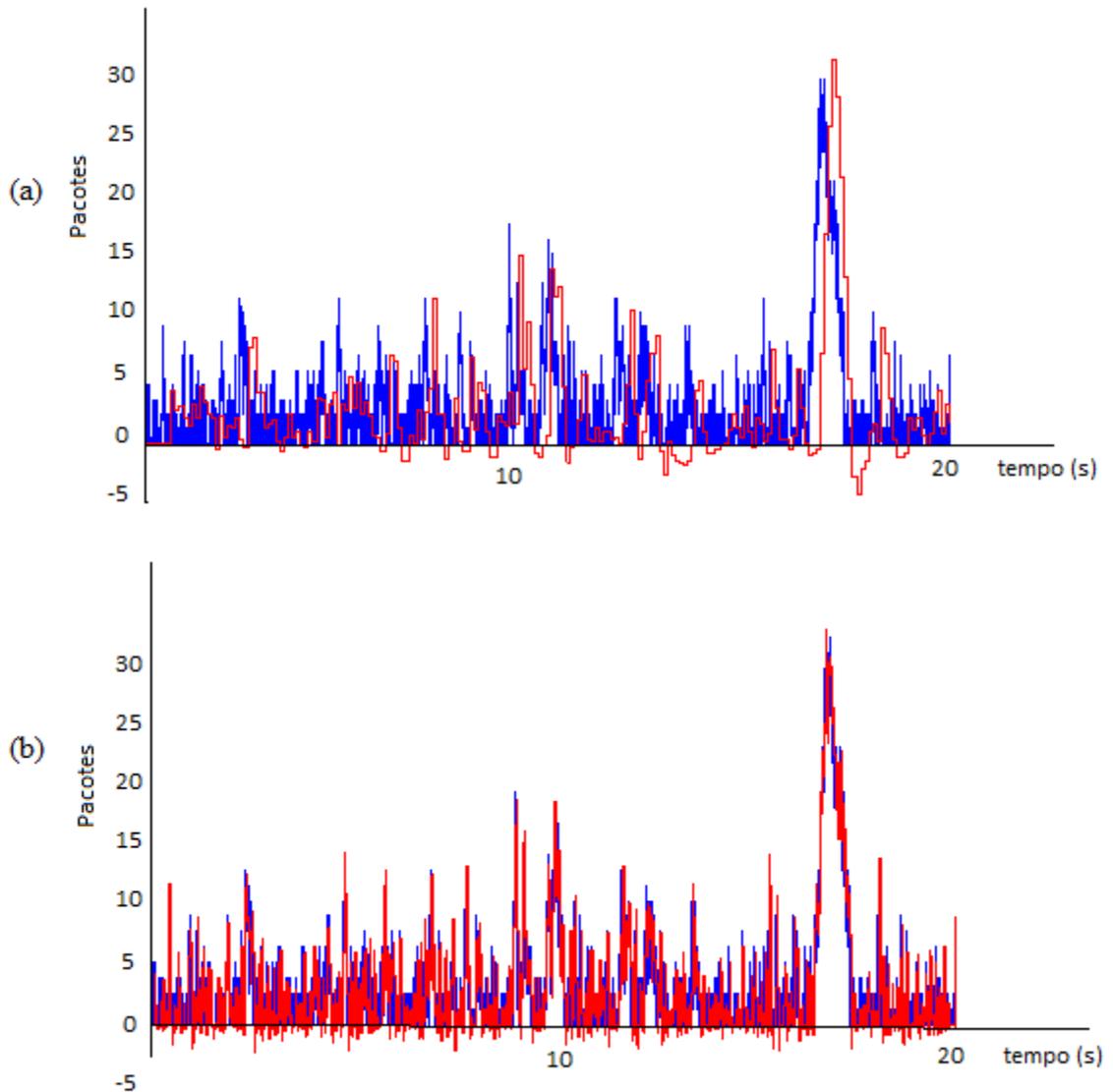


Figura 6.4. Tráfego real (azul) e tráfego estimado (vermelho) para T_SAMPLE (a)1s (b)0.001 s.

Na Tabela 6.2 são apresentados os erros de estimação utilizando o filtro de Kalman para diferentes valores de T_SAMPLE e para os diferentes tipos de tráfego apresentados na seção anterior. O erro é maior quanto maior for o valor de T_SAMPLE , como discutido anteriormente, no entanto o erro de predição é considerado baixo, mesmo para o maior valor de T_SAMPLE . O período simulado para cada valor de T_SAMPLE foi de 14400s (4 horas), e foram realizadas 50 replicações da simulação para cada valor.

TABELA 6.2. ERRO QUADRÁTICO MÉDIO DE PREVISÃO PARA DIFERENTES CARACTERÍSTICAS DE TRÁFEGO.

	Erro Quadrático Médio			
T_SAMPLE	Tráfego			
	Baixo	Médio	Alto	Muito alto
0.01	.00396	.00576	.01948	.09026
0.1	.00441	.00608	.01962	.09026
1	.00682	.00786	.02032	.08981
3	.01018	.01069	.02130	.09054
5	.01275	.01357	.02284	.09014
10	.01732	.01800	.02481	.09025
20	.02398	.02452	.02997	.09117
30	.02929	.02974	.03524	.09090

O valor de T_SAMPLE deve ser um compromisso entre desempenho do dispositivo móvel e consumo de energia, para que a previsão seja a mais próxima do real possível, porém, sem degradar a usabilidade do terminal para o usuário. Para um valor pequeno de T_SAMPLE, um maior poder de processamento é exigido, o que pode degradar o desempenho de alguns terminais móveis, e até elevar o consumo de energia do dispositivo, no entanto a predição será mais precisa. Por outro lado, um valor de T_SAMPLE maior exige menos poder de processamento, em contrapartida de uma previsão menos precisa. Esse valor também deve ser menor que o valor de T2 definido pela rede, uma vez que deve realizar a previsão do tráfego em tempo hábil e a desconexão da interface de rádio deve ser feita o quanto antes para evitar desperdício do consumo de energia, caso seja previsto ausência de tráfego até o término de T2.

6.2.2 Modelo II-B - Modelo Completo

Partindo do modelo de simulação da rede real validado (Modelo I), foi inserido o módulo referente ao modelo de previsão baseado no filtro de Kalman com subfiltro (Modelo II-A), como pode ser visto na Figura 6.5.

O novo modelo pode ser representado em 3 grandes blocos:

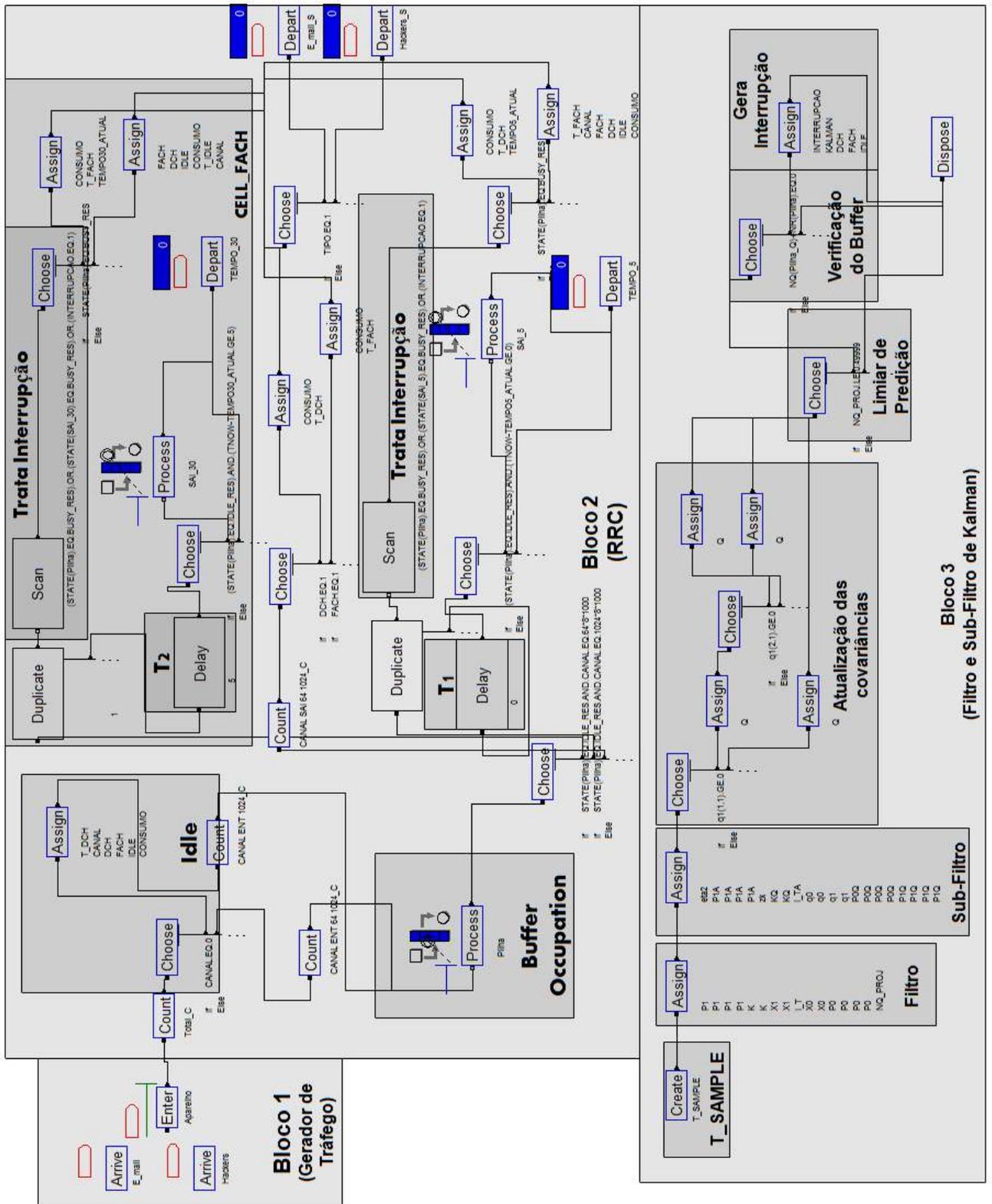


Figura 6.5. Modelo de simulação baseado na rede real incluindo modelo de previsão.

O bloco 1 representa a chegada de pacotes e o bloco 2 a máquina de estados do RRC, assim como apresentadas anteriormente para o Modelo I.

O bloco 3 apresenta o modelo de previsão do filtro e subfiltro de Kalman, também similar ao modelo II-A.

A principal diferença é que os tempos de inatividade não são os únicos responsáveis pela desconexão do RRC. A cada T_SAMPLE segundos a projeção da pilha é calculada, e caso seja detectado que nos próximos T_SAMPLE segundos não chegue nenhum pacote, um *flag* de interrupção é gerado. Essa interrupção só é ativada se:

- Não houver pacotes sendo atendidos na pilha (*Buffer Occupation* vazio) – Verificação do *Buffer*.
- O resultado da previsão é menor do que um limiar pré-estabelecido (O resultado da previsão pode resultar num valor que não é zero, mas bem próximo do que também indica tráfego nulo) – Limiar de Predição.

Essa interrupção é verificada junto aos blocos SCAN, responsáveis por verificar o estouro da temporização de $T1$ e $T2$ dos blocos DELAY. Quando o *flag* é acionado, os temporizadores são interrompidos e o dispositivo vai para o estado *Idle*, o qual o telefone consome menos energia.

Desse modo, é possível realizar dois tipos de simulação:

- O filtro de Kalman atuando em conjunto com os temporizadores $T1$ e $T2$ da rede real, sem alterações nos mecanismos já existentes das redes de terceira geração;
- O filtro de Kalman atuando independente dos temporizadores $T1$ e $T2$, e toda a lógica de desconexão controlada pelo filtro. Para trabalhar dessa maneira, basta configurar no modelo $T1=0$ e $T2=\text{infinito}$. Desse modo, quando não houver mais tráfego no dispositivo, o móvel vai para o estado Cell_FACH (nível de consumo intermediário), e nele permanece até que o filtro decida se o dispositivo deve ir para Idle (não há pacotes para chegar nos próximos T_SAMPLE segundos) ou permaneça em Cell_FACH, evitando uma nova reconexão no RRC, consumindo menos energia (evita-se o estado Cell_DCH), e também o envio de pacotes de sinalização que geram um atraso no recebimento dos pacotes.

6.2.2.1 Simulações e Resultados

Para verificar a eficiência do filtro em relação ao tráfego real monitorado foram realizadas três simulações:

- **Simulação A:** Filtro de Kalman coexistindo com os temporizadores $T1$ e $T2$ com valores de 5 e 30 segundos, respectivamente.
- **Simulação B:** Filtro de Kalman coexistindo com os temporizadores $T1$ e $T2$ com valores de 5 segundos para ambos.
- **Simulação C:** Filtro de Kalman atuando sozinho na decisão de desconexão do RRC.

Os parâmetros utilizados para geração dos pacotes nessa simulação foram obtidos através da monitoração dos pacotes da rede real, apresentados na tabela 5.3.

Na Tabela 6.3 são apresentados os resultados referentes à Simulação A.

TABELA 6.3. SIMULAÇÃO A – FILTRO DE KALMAN ATUANDO JUNTO COM $T1=5S$ E $T2=30S$

Período	CONSUMO (mWh)							
	Modelo I	Modelo II B						
	$T1=5s$ e $T2=30s$	T_SAMPLE (s)						
		1	5	10	15	20	25	Infinito
Total	2,24E+02	1,82E+02	1,87E+02	1,84E+02	1,86E+02	1,85E+02	1,83E+02	2,24E+02
0h às 4h	1,92E+02	1,80E+02	1,80E+02	1,80E+02	1,78E+02	1,78E+02	1,80E+02	1,92E+02
4h às 8h	1,93E+02	1,83E+02	1,84E+02	1,84E+02	1,82E+02	1,83E+02	1,84E+02	1,93E+02
8h às 12h	2,17E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	2,17E+02
12h às 16h	2,16E+02	1,88E+02	1,88E+02	1,88E+02	1,88E+02	1,88E+02	1,87E+02	2,16E+02
16h às 20h	*	*	*	*	*	*	*	*
20h às 24h	2,19E+02	1,83E+02	1,83E+02	1,83E+02	1,82E+02	1,84E+02	1,84E+02	2,19E+02

* Valores não simulados devido à limitação de entradas da versão educacional do software Arena.

As simulações foram executadas levando em consideração diferentes períodos do dia, utilizando o Modelo I validado [TITO, 2011], e utilizando o Modelo II B, onde foram mantidos $T1$ e $T2$ utilizados no Modelo I, e variado o valor de T_SAMPLE. A redução de consumo obtida varia entre 5,6% e 18,6%. Considerando o período total, o valor de T_SAMPLE que foi mais eficiente foi de 1 segundo. Considerando os períodos isoladamente, um T_SAMPLE de 15 segundos foi o que demonstrou melhor desempenho. Quando o Modelo II B é simulado com T_SAMPLE com um valor muito alto (tendendo ao infinito), retrocedemos ao Modelo I, uma vez

que o filtro de Kalman não atua no modelo, e o consumo fica a cargo apenas de $T1$ e $T2$. Nota-se que para este caso os valores obtidos são os mesmos obtidos no Modelo I, validando o Modelo II B a partir de um anteriormente validado.

Na Tabela 6.4 são apresentados os resultados da Simulação B. Mesmo para um $T2$ menor, a atuação do filtro fez com que houvesse a redução no consumo entre 12% e 14,5%. Todos os valores de T_SAMPLE simulados tiveram um desempenho similar no consumo para um mesmo período.

TABELA 6.4. SIMULAÇÃO B – FILTRO DE KALMAN ATUANDO JUNTO COM $T1=5s$ E $T2=5s$

Período	CONSUMO (mWh)							
	Modelo I	Modelo II B						
	$T1=5s$ e $T2=5s$	T_SAMPLE (s)						
		1	5	10	15	20	25	Infinito
Total	2,21E+02	1,91E+02	1,89E+02	1,89E+02	1,90E+02	1,90E+02	1,89E+02	2,21E+02
0h às 4h	2,04E+02	1,80E+02	1,80E+02	1,80E+02	1,79E+02	1,79E+02	1,79E+02	2,04E+02
4h às 8h	2,08E+02	1,83E+02	1,84E+02	1,84E+02	1,82E+02	1,82E+02	1,82E+02	2,08E+02
8h às 12h	2,13E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	2,13E+02
12h às 16h	2,25E+02	1,88E+02	1,88E+02	1,89E+02	1,88E+02	1,88E+02	1,88E+02	2,25E+02
16h às 20h	*	*	*	*	*	*	*	*
20h às 24h	2,08E+02	1,83E+02	1,83E+02	1,83E+02	1,83E+02	1,83E+02	1,84E+02	2,08E+02

*Valores não simulados devido à limitação de entradas da versão educacional do software Arena.

Na Simulação C temos apenas a influência do filtro de Kalman atuando no modelo de decisão de troca de estados. Para isso, $T1$ foi configurado para 0 segundos, a fim de que assim que for cessado o tráfego em Cell_DCH, o modelo vá para Cell_FACH, que é um estado de consumo intermediário. O temporizador $T2$ por sua vez, foi configurado para infinito, fazendo com que o telefone só vá para o estado *Idle* quando o filtro de Kalman detectar que não haverá tráfego. O único temporizador que irá atuar nesse caso será T_SAMPLE . Utilizando a configuração descrita, foi obtido um resultado excelente em relação à redução de consumo, como mostra a Tabela 6.5.

A Tabela 6.6 apresenta o resultado da comparação para as simulações A, B e C. Na primeira coluna (Simulação A), é apresentada a relação de ganho entre o Modelo I com um $T1=5s$ e um $T2=30s$ configurados pela rede e o Modelo II-B que utiliza o filtro de Kalman em conjunto com os parâmetros da rede. A segunda coluna (Simulação B) apresenta a relação de

ganho entre o Modelo I com um $T1=5s$ e um $T2=5s$ configurados pela rede e o Modelo II-B que utiliza o filtro de Kalman conjunto com os parâmetros da rede. Na terceira coluna (Simulação C), é apresentada a relação de ganho entre os menores valores de consumo obtidos no Modelo I para $T2=30s$ e $T2=5s$, configurados pela rede, e o Modelo II-B utilizando apenas o filtro de Kalman para decisão de desconexão do RRC.

TABELA 6.5. SIMULAÇÃO C – FILTRO DE KALMAN ATUANDO SOZINHO NO MODELO ($T1=0s$ E $T2=INFINITO$).

Período	CONSUMO (mWh)						
	Modelo I e II*	Modelo II B					
		T_SAMPLE					
		1	5	10	15	20	30
Total	2,24E+02	1,82E+02	1,87E+02	1,84E+02	1,86E+02	1,85E+02	1,82E+02
0h às 4h	1,92E+02	1,78E+02	1,79E+02	1,79E+02	1,77E+02	1,78E+02	1,78E+02
4h às 8h	1,93E+02	1,80E+02	1,81E+02	1,83E+02	1,81E+02	1,82E+02	1,80E+02
8h às 12h	2,13E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02	1,84E+02
12h às 16h	2,16E+02	1,87E+02	1,87E+02	1,87E+02	1,87E+02	1,87E+02	1,87E+02
16h às 20h	**	**	**	**	**	**	**
20h às 24h	2,08E+02	1,83E+02	1,83E+02	1,83E+02	1,81E+02	1,83E+02	1,83E+02

*Menor consumo para $T2=5s$ ou $T2=30s$

** Valores não simulados devido à limitação de entradas da versão educacional do software Arena.

TABELA 6.6. EFICIÊNCIA DO FILTRO PARA OS CENÁRIOS SIMULADOS

Período	Redução de Consumo (%)		
	Simulação A	Simulação B	Simulação C
Total	-18,52	-14,54	-18,69
0h às 4h	-7,54	-12,60	-13,15
4h às 8h	-5,63	-12,67	-13,24
8h às 12h	-15,01	-13,63	-15,36
12h às 16h	-13,48	-16,72	-17,13
16h às 20h	*	*	*
20h às 24h	-16,93	-12,21	-17,19

* Valores não simulados devido à limitação de entradas da versão educacional do software Arena.

Em nenhum caso houve aumento no consumo de energia. Em todos os casos o filtro de Kalman obteve sucesso na previsão, independente do tipo de tráfego, e com isso não houve desperdício de energia, otimizando o consumo com uma troca de estados coerente. Considerando o período total de tráfego analisado, o modelo utilizando apenas Kalman na decisão da troca de estados do RRC se mostrou mais eficiente, reduzindo em 18,69% o consumo. Um valor de T_SAMPLE de 1 segundo se mostrou mais eficiente para o período total. Analisando os períodos separadamente, valores de 1 e 15 segundos para T_SAMPLE se mostraram mais eficientes.

Considerando o período total de utilização do móvel, o filtro de Kalman atuando sozinho no sistema apresentou um maior ganho na redução de consumo do que atuando em conjunto com $T1$ e $T2$, como pode ser observado na Tabela 6.6.

Na Tabela 6.7 é apresentada a relação de consumo entre a simulação do Modelo I para $T1=5s$ e $T2=30s$ e a simulação do Modelo I para um $T1=5s$ e $T2=5s$ (primeira coluna), e entre a simulação do Modelo I para $T1=5s$ e $T2=30s$ e a simulação do Modelo II-B para $T2$ dinâmico (segunda coluna). A terceira coluna que apresenta $T2_{Dinâmico} - T2_{5s}$ mostra o quanto a mais de redução é obtida utilizando o modelo dinâmico em relação a um $T2$ fixo de 5s. Para um $T2=5s$ fixo, houve um aumento de consumo em relação a um $T2=30s$ para os períodos de 0h às 4h, 4h às 8h e 12h às 16h. Para um $T2$ dinâmico, não houve aumento de consumo em nenhum caso, apenas redução.

TABELA 6.7. COMPARATIVO DE CONSUMO PARA UM $T2=30s$ (MODELO I) CONTRA UM $T2=DINÂMICO$ (MODELO II-B) E UM $T2=5s$ (MODELO I)

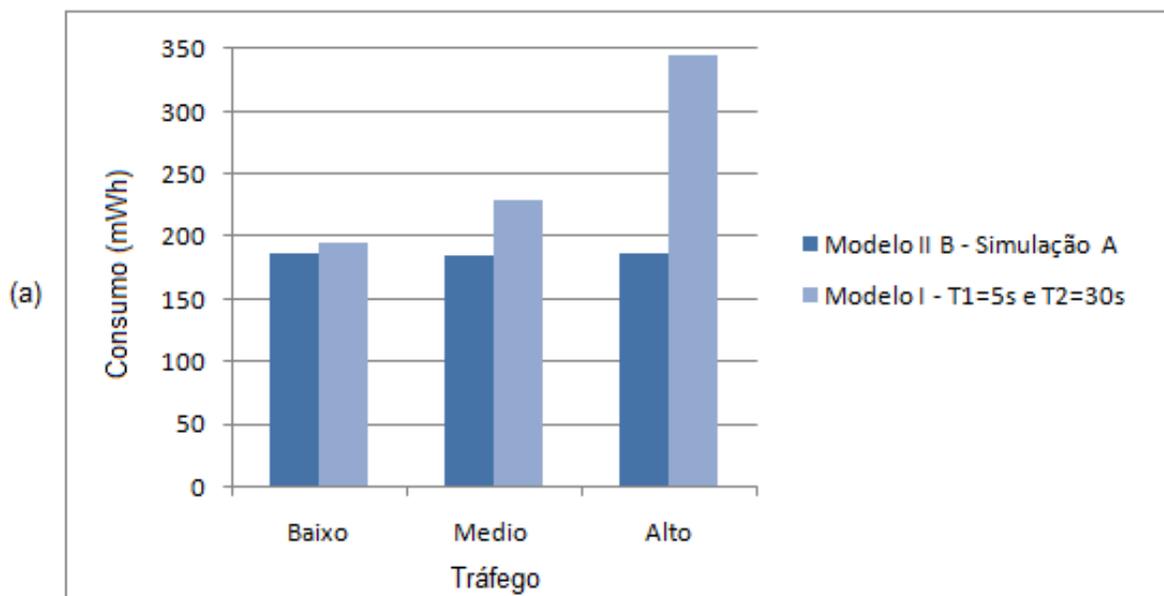
Período	Redução de Consumo (%)		$T2_{Dinâmico} - T2_{5s}$
	$T2=5s$ (Modelo I) em relação a $T2=30s$ (Modelo I)	$T2=dinâmico$ (Modelo II-B) em relação a $T2=30s$ (Modelo I)	
Total	-1,34	-18,69	-17,35
0h às 4h	6,25	-13,15	-19,40
4h às 8h	7,77	-13,24	-21,01
8h às 12h	-1,84	-15,36	-13,52
12h às 16h	4,17	-17,13	-21,30
16h às 20h	*	*	*
20h às 24h	-5,02	-17,19	-12,17

* Valores não obtidos devido à limitação de entradas da versão educacional do software Arena.

Para o período total, tem-se uma redução de aproximadamente 17% a mais do que quando utilizado $T2$ fixo de 5 segundos. Esse ganho na redução de consumo é muito próximo ao valor estimado no Capítulo 3, que era de 20%. No entanto, para alguns períodos isolados, essa relação aproxima-se mais ainda do valor estimado, como no período de 0h às 4h, que é de 19,4%, e chega a ultrapassar o valor estimado, como no caso dos períodos de 4h às 8h e 12h às 16h, que é em torno de 21%.

Partindo dos dados utilizados para validação do Modelo I presentes na Tabela 6.1, foi possível traçar o comportamento do consumo para baixo, médio e alto tráfego utilizando as mesmas premissas das Simulações A, B e C tratadas anteriormente.

Na Figura 6.6(a) é possível observar o comportamento do consumo para os três tipos de tráfego quando se utiliza o filtro de Kalman em conjunto com $T1=5s$ e $T2=30s$ (Simulação A – Modelo II B) e sua relação sem utilizar filtro de Kalman para um $T1=5s$ e $T2=30s$ (Modelo I). Na Figura 6.6(b) é possível observar o comportamento do consumo para os três tipos de tráfego quando se utiliza o filtro de Kalman em conjunto com $T1=5s$ e $T2=5s$ (Simulação A – Modelo II B) e sua relação sem utilizar filtro de Kalman para um $T1=5s$ e $T2=5s$ (Modelo I).



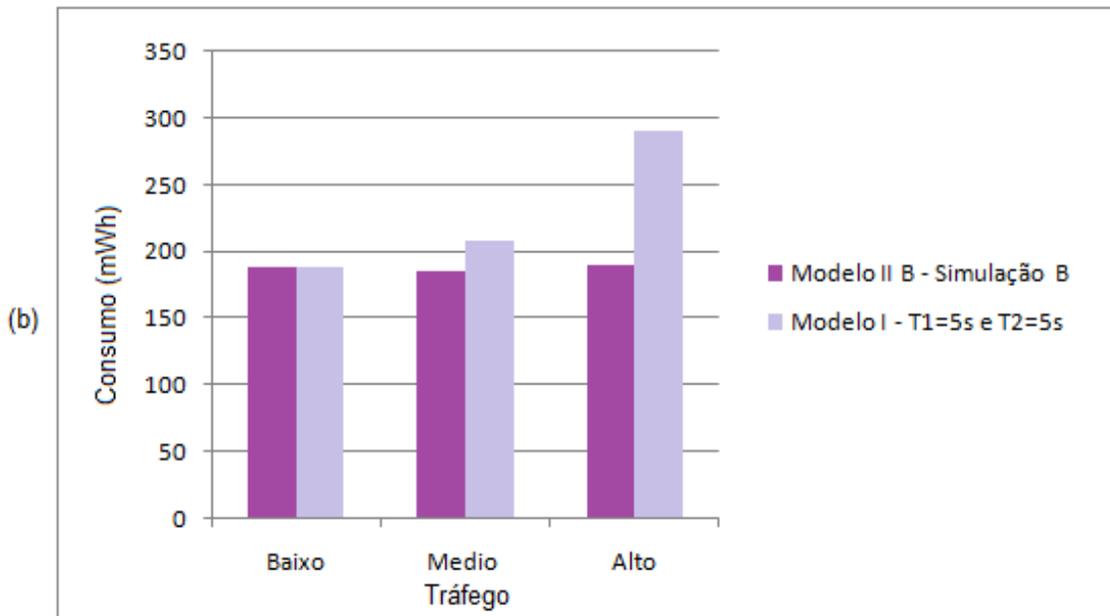


Figura 6.6. Comparação entre o consumo obtido para o modelo utilizando filtro de Kalman (Modelo II B) e limiar de consumo sem filtro de Kalman (Modelo I) para baixo, médio e alto tráfego para (a) $T1=5s$ e $T2=30s$ e (b) $T1=5s$ e $T2=5s$.

Quando o filtro de Kalman é utilizado junto com a temporização, o consumo é praticamente o mesmo, tanto para baixo, médio e alto tráfego. Isso demonstra a eficiência do filtro, e que independente da temporização de $T1$ e $T2$ a qual sofre interferência, o filtro consegue prever o tráfego com precisão e atuar na desconexão adequadamente. No entanto, a redução de consumo é mais significativa quando o filtro atua em um $T2$ de 30 segundos, uma vez que esse valor de temporização faz o dispositivo consumir mais energia.

Quando o Kalman atua junto com os tempos de inatividade da rede, ele terá um ganho maior de redução de consumo quanto maior for o valor de $T2$. Para um $T2$ baixo, ainda haverá um ganho na redução do consumo, porém menor.

Analisando o filtro de Kalman atuando sozinho na desconexão do RRC para um alto, médio e baixo tráfego, seu desempenho é melhor ou igual quando comparado com o Modelo I para um $T1=5s$ e $T2=5s$, que possui um valor de consumo baixo, como mostra a Figura 6.7.

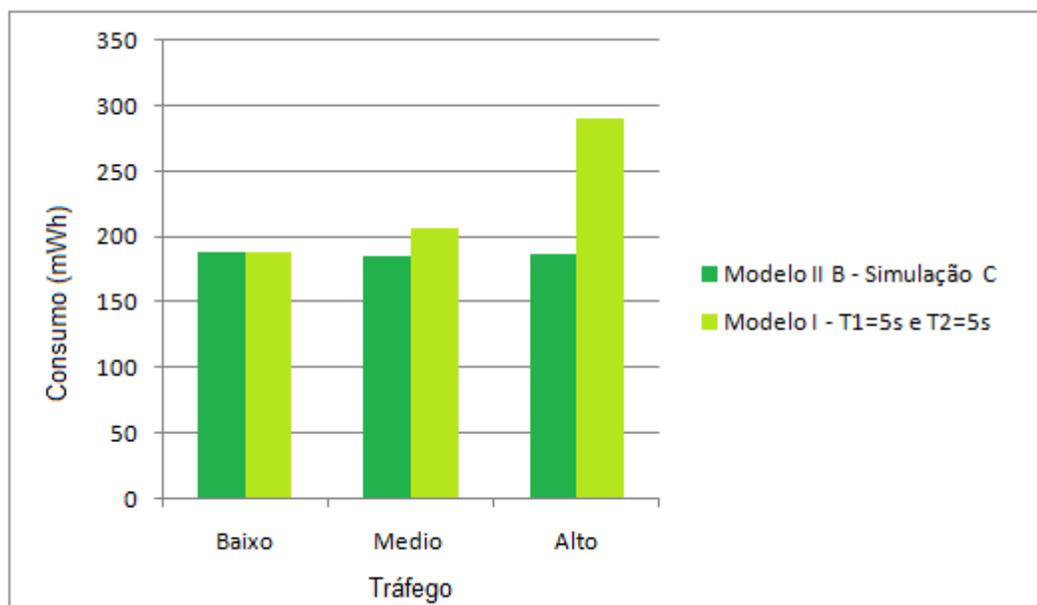


Figura 6.7. Comparação entre o consumo obtido para o modelo utilizando o modelo II B controlado apenas pelo filtro de Kalman e o limiar de consumo sem filtro de Kalman (Modelo I) para baixo, médio e alto tráfego para $T1=0s$ e $T2=infinito$.

6.3 IMPLEMENTAÇÃO DO FILTRO NOS DISPOSITIVOS MÓVEIS

Inicialmente, todos os cálculos teóricos e algoritmo apresentados no Capítulo 4 foram convertidos para a linguagem de programação do software MATLAB, como mostram as Figuras 6.8(a) e 6.8(b), respectivamente.

Depois de testado e validado, o algoritmo foi transposto para uma notação a qual o software ARENA pudesse interpretar. Para isso, todos os cálculos envolvendo matrizes tiveram que ser desenvolvidos manualmente, trabalhando cada índice separadamente, como mostra a Figura 6.8(c). Para validação das equações utilizadas no modelo ARENA elas foram transpostas então para a notação da linguagem C, uma vez que as notações são bem semelhantes, como mostra a Figura 6.8(d). Por último, o programa em C foi modificado para simplificação da notação dos cálculos de matrizes e vetores para ser inserido no dispositivo móvel, como mostra a Figura 6.8(e). Essa simplificação foi apenas na notação da linguagem, ou seja, todos os cálculos e resultados apresentados são os mesmos para os programas C e C simplificado.

(a) Teórica

$$\bar{P}_k = (I - K_k H) P_k^*$$

(b) Matlab

$$P0 = (I - K*H) *P1$$

(c) Arena

```
P0(1,1)=(I(1,1)-K(1,1)*H(1,1))*P1(1,1)+(I(1,2)-K(2,1)*H(1,2))*P1(1,2):
P0(1,2)=(I(1,1)-K(1,1)*H(1,1))*P1(2,1)+(I(1,2)-K(1,1)*H(1,2))*P1(2,2):
P0(2,1)=(I(2,1)-K(2,1)*H(1,1))*P1(1,1)+(I(2,2)-K(2,1)*H(1,2))*P1(1,2):
P0(2,2)=(I(2,1)-K(2,1)*H(1,1))*P1(2,1)+(I(2,2)-K(2,1)*H(1,2))*P1(2,2):
```

(d) C

```
P0[0][0]=(I[0][0]-K[0][0]*H[0][0])*P1[0][0]+(I[0][1]-K[0][0]*H[0][1])*P1[0][1];
P0[0][1]=(I[0][0]-K[0][0]*H[0][0])*P1[1][0]+(I[0][1]-K[0][0]*H[0][1])*P1[1][1];
P0[1][0]=(I[1][0]-K[1][0]*H[0][0])*P1[0][0]+(I[1][1]-K[1][0]*H[0][1])*P1[0][1];
P0[1][1]=(I[1][0]-K[1][0]*H[0][0])*P1[1][0]+(I[1][1]-K[1][0]*H[0][1])*P1[1][1];
```

(e) C simplificado

```
for (i=0;i<2;i++)
    for (j=0;j<2;j++)
        Mat1[i][j] = K[i] * H[j] ;

for (i=0;i<2;i++)
    for (j=0;j<2;j++)
        Mat2[i][j] = I[i][j] - Mat1[i][j] ;

for (i=0;i<2;i++)
    for (j=0;j<2;j++){
        P0[i][j] = 0.0;
        for (k=0;k<2;k++)
            P0[i][j] += (Mat2[i][k] * P1[k][j]) ;
    }
}
```

Figura 6.8. Implementações para o filtro de Kalman em diversas linguagens - validação.

Utilizando um mesmo conjunto de 1000 dados de entrada, foram executadas simulações utilizando os programas feitos em MATLAB, C e C simplificado para obtenção do valor previsto. Os valores obtidos nos dois modelos implementados em C foram semelhantes ao do filtro implementado em MATLAB, que foi utilizado como referência, apresentando um erro quadrático médio zero para os programas em C e C simplificado em relação ao resultado apresentado pelo código executado no MATLAB. Os gráficos comparativos dos valores previstos para as três implementações são apresentadas na Figura 6.9. Como as notações em linguagem C e ARENA são semelhantes, os cálculos utilizados no software ARENA puderam ser validados pelo programa em C.

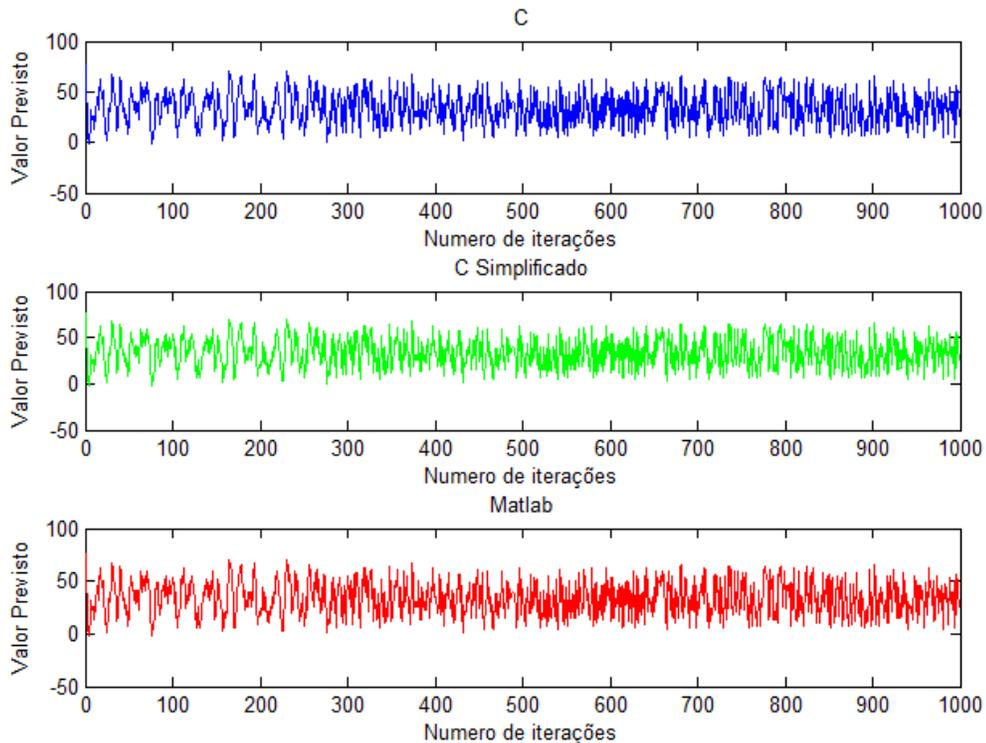


Figura 6.9. Resultado da previsão utilizando filtro de Kalman para as três implementações: C, C simplificado e MATLAB

Utilizando o kit de desenvolvimento (Software Development Kit - SDK) para plataforma Android [ANDSDK], o código em C simplificado foi inserido no “*User Space*”⁶ e compilado para processadores ARM, presentes nos dispositivos móveis os quais o algoritmo foi implantado. O binário final gerado foi inserido no telefone pela interface “ADB” (*Android Debug Bridge*) [ADB], que faz a comunicação entre dispositivo móvel e computador, também presente no SDK Android, pelo comando:

```
adb push <caminho_local>/arquivo <caminho_telefone>
```

Para executar o binário com o código de Kalman no dispositivo móvel, foi obtida uma aplicação do *Android Market* chamada “Terminal”, que consiste de uma interface semelhante a

⁶ O Android é baseado no *Kernel* do Linux, e possui a separação em “*Kernel Space*” e “*User Space*”. O *Kernel Space* é reservado para executar o *Kernel*, suas extensões e drivers de dispositivos, enquanto o *User Space* é a área reservada para aplicações de usuários.

um *shell* do Linux, como mostra a Figura 6.10. O binário foi executado utilizando o comando *time*, apresentado a seguir, para medir o tempo de execução.

```
time ./kalman
```

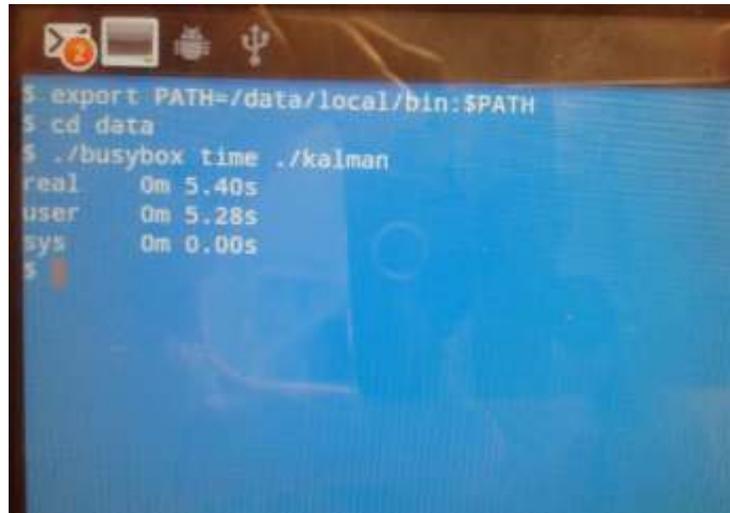


Figura 6.10. Aplicação “Terminal” mostrando a execução do software do Kalman para medição de desempenho.

Foi feita uma comparação do desempenho desse algoritmo executado em um computador pessoal e no dispositivo móvel. A Tabela 6.8 apresenta a configuração dos equipamentos e a Figura 6.11 apresenta os resultados para o tempo decorrido para uma variação de 1 a 10 milhões de execuções do algoritmo. Cada execução corresponde a uma previsão no instante seguinte, por exemplo, 1 execução é equivalente a um T_SAMPLE de 1 segundo, e 10 milhões de execuções são equivalentes a um T_SAMPLE de $100ns$. Pelos resultados obtidos, pode-se observar que o algoritmo é eficiente até um T_SAMPLE de $10\mu s$, o que é mais do que suficiente para uma previsão adequada do tráfego.

As implementações feitas correspondem apenas a parte relacionada à previsão de tráfego para medir o desempenho do filtro de Kalman nos dispositivos. A implementação do modelo completo no dispositivo móvel envolve mudanças na parte de rádio, que é responsável pelo envio de sinalização de desconexão do recurso de rádio, e que por hora não foram implementadas.

TABELA 6.8. CONFIGURAÇÕES DOS DISPOSITIVOS PARA TESTE DE DESEMPENHO DO ALGORITMO

Computador Pessoal	Dispositivo Móvel I	Dispositivo Móvel II
Processador: INTEL Core 2 Duo 2.8 GHz L2 cache 6MB Memória: 4GB RAM – 800MHZ Barramento 1.7GHz	Processador: ARM Cortex-A9 Dual Core 1.2 GHz Memória: 1GB RAM	Processador: ARM Cortex-A8 1.0 GHz Memória: 512 MB de RAM

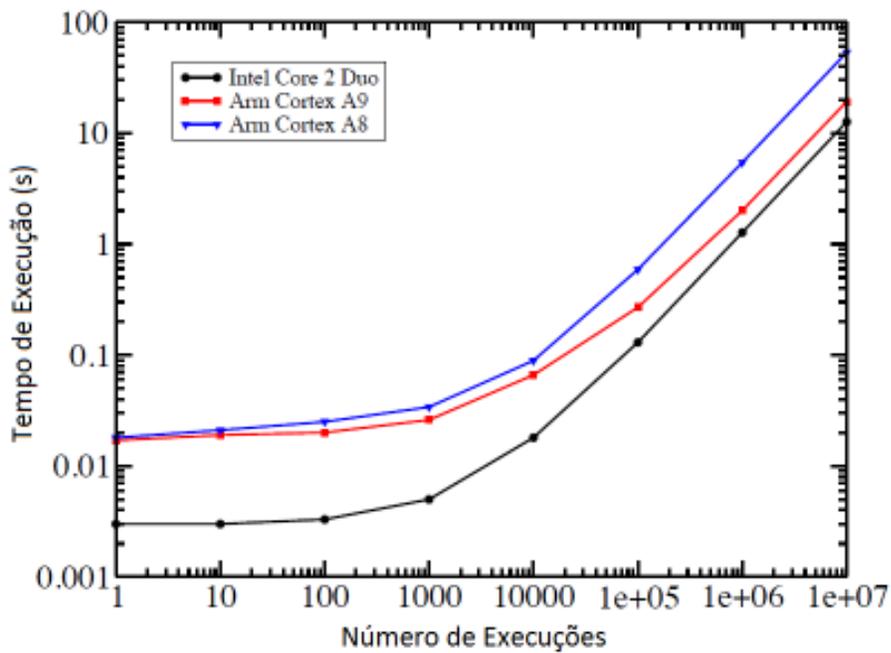


Figura 6.11. Comparação de desempenho do algoritmo em um computador pessoal e nos dispositivos móveis.

7. CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

O aumento da utilização de dados devido ao surgimento dos dispositivos *Always On* e o uso ineficiente dos recursos de rádio da rede de terceira geração celular faz com que os dispositivos móveis 3G tenham um consumo grande de energia.

Algumas técnicas como evitar o recebimento de pacotes desnecessários e aumento do intervalo de pacotes de sinalização apresentam um ganho na redução de consumo de energia do dispositivo. No entanto, esta solução deve ser feita do lado do núcleo da rede ou do lado da aplicação, respectivamente, o que torna a solução difícil de ser viabilizada. Essa inviabilidade ocorre, pois, no lado da rede, envolve mudanças de estrutura e inserção de novos equipamentos e *software* de controle, que podem causar degradação de desempenho caso seja feita de forma incorreta, e, no lado da aplicação pode envolver novos custos para retrabalhar o *software*, deixando os clientes da aplicação insatisfeitos.

A técnica apresentada neste trabalho teve como foco aplicar a solução no dispositivo móvel e, independentemente da rede utilizada e das aplicações que estejam consumindo dados, diminuir o consumo de forma a melhor aproveitar os recursos de rede 3G, individualmente para cada usuário, e não de forma geral como é utilizado atualmente por meio dos tempos de inatividade $T1$ e $T2$. O cálculo de estimativa de melhor tempo para desconexão foi feita do lado do dispositivo móvel, e não na rede da operadora como em [LIERS, 2006]. Com isso, o dispositivo móvel consegue prever a chegada de pacotes de acordo com o perfil de cada usuário sem sobrecarregar a infra-estrutura da rede com cálculos adicionais, atuando nos problemas apontados por [QIAN, 2010]. Para a desconexão realizada pelo lado do positivo, não foi necessária nenhuma modificação das normas internacionais, uma vez que a desconexão dos recursos de rádio originada pelo dispositivo já é uma técnica empregada por alguns fabricantes de dispositivos, mas não de forma dinâmica pelo método conhecido por *Fast Dormancy*.

Neste trabalho, o método de previsão utilizado foi o Filtro de Kalman, devido a sua alta aceitação e desempenho para previsão de tráfego. No entanto, outros tipos de filtro podem ser utilizados ou criados para serem utilizados neste método de redução de consumo, ficando como sugestão para trabalhos futuros.

Utilizando os modelos de simulação foi possível comprovar a eficiência de previsão do algoritmo do filtro de Kalman. Também foi possível simular diversos cenários de tráfego (rede real, obtida através de coleta de dados, e tráfego controlado), utilizando ou não o filtro.

Utilizando o método proposto para um $T2$ dinâmico, porém limitados por $T1$ e $T2$ configurados pela rede, a utilização do filtro se mostrou eficiente independente do valor de $T2$ o qual atua em conjunto.

Utilizando o método proposto para um $T2$ dinâmico sem a interferência de $T1$ e $T2$, foi obtida uma redução do consumo em torno de 17% em média quando comparado com a redução obtida por um $T2$ fixo de menor valor (*Fast Dormancy*). Esse resultado foi próximo aos 20% de redução estimada inicialmente. Tratando os períodos isoladamente, chegou-se a obter uma redução de mais de 20% neste mesmo cenário.

A decisão de desconexão do RRC controlada somente pelo filtro de Kalman (sem interferência de $T1$ e $T2$) apenas foi possível devido ao uso de modelo de simulação. Para que esse método seja possível na rede real, deve-se realizar modificações no núcleo da rede, ou em protocolos de sinalização para que esse controle seja feito 100% pelo dispositivo, ficando como sugestão de estudo para trabalhos futuros.

O método adaptativo utilizado para redução de consumo de energia considerou os tráfegos elástico de e-mail e pacotes não solicitados provenientes de hacker que são os principais pacotes recebidos enquanto o telefone permanece em estado de dormência, no entanto sua aplicação é estendida para todo tipo de pacote de dados. Este método atua na principal dificuldade apontada por [PENG, 2008], que é a redução de consumo para um alto tráfego de chegada de pacotes.

As implementações realizadas no dispositivo móvel foram realizadas visando apenas verificar o desempenho do algoritmo do filtro de Kalman e o poder de processamento exigido do dispositivo móvel. A interação com o módulo de rádio do dispositivo para desconexão do RRC fica como sugestão de trabalho futuro.

A tecnologia de quarta geração celular (4G) começa a ser apresentada ao mundo em caráter experimental. No entanto, publicações sobre o assunto ainda são muito recentes. Devido à falta de terminais 4G no mercado, e de uma rede real ou simulada desta tecnologia, não foi possível medir o consumo de energia real nesses dispositivos, ficando também como sugestão de trabalhos futuros.

REFERÊNCIAS

[Kalman, 1960] KALMAN, Rudolf E. **A new approach to linear filtering and prediction problems.** Trans, ASME (J. Basic Engineering), 82D, 35-45, 1960.

[RIOS, 1982] RIOS, Neto A., KUGA, H. K. **Estimação adaptativa para o estado ruidoso para o Filtro de Kalman.** 4º Congresso Brasileiro de Automática, Campinas, São Paulo, 1982. Vol 1:101-105.

[OKUTANI, 1984] OKUTANI, I. and STEPHANEDES, Y.. **Dynamic Prediction of Traffic Volume through Kalman Filtering Theory.** In Transportation Research B, Vol. 18B, No. 1, pp 1-11, 1984.

[ASRABADI, 1990] ASRABADI, Badiollah. **Estimation in the pareto distribution.** Metrika. ISSN: 0026-1335. Vol. 37. Pg 199-205. 1990.

[JAIN, 1991] JAIN, Raj. **The art of computer systems performance analysis.** John Wiley & Sons, Inc., 1991.

[SMITH, 1994] B.L. Smith, M.J. Demetsky, **Short-term traffic flow prediction: Neural network approach.** In Transportation Research Record No. 1453, Intelligent Transportation Systems: Evaluation, Driver Behavior, and Artificial Intelligence, 98-104, 1994.

[MOORE, 1999] MOORE, D. S. and McCabe, G. P. **Introduction to the Practice of Statistics.** 3rd edition. New York: W. H. Freeman, 1999.

[CASTRO, 2001] CASTRO, Jonathan P. **The UMTS Network and Radio Access Technology: Air Interface Techniques for Future Mobile Systems.** Editora Wiley. 1ª edição. 2001.

[TACHIKAWA, 2002] TACHIKAWA, Keiji. **W-CMDA: mobile communications system.** Editora Willey. 2002.

[SPIEGEL, 2003] SPIEGEL, MURRAY R. **Probabilidade e Estatística.** Editora Pearson. 2003.

[BAVOSA, 2004] BAVOSA, Alan. **GPRS Security Threats and Solution Recommendations.** White Paper. Juniper Networks. 2004.

[QUEIROZ, 2004] QUEIROZ, R. B.; RODRIGUEZ, A. G.; GÓMEZ, A. T. **Estudo comparativo entre as técnicas máxima verossimilhança gaussiana e redes neurais na classificação de imagens IR-MSS CBERS 1"**, WORKCOMP 2004, 2004.

[YEH, 2004] LEE, Chi-Chen; YEH, Jui-Hung; CHEN, Jyh-Cheng. **Performance Analysis of Energy Consumption in 3GPP Networks.** Proc. Wireless Telecomm. Symp., IEEE, May 2004.

[CHEN, 2004] CHEN, Jyh-Cheng; ZHANG, Tao. **IP-Based Next-Generation Wireless Networks. System, Architectures and Protocols**. Wiley-Interscience. 2004.

[LEE-CHEN, 2004] LEE, Chi-Chen; YEH, Jui-Hung; CHEN, Jyh-Cheng. **Impact of Inactivity Timer on Energy Consumption in WCDMA and cdma2000**. in Proc. Wireless Telecommunications Symposium, IEEE 2004.

[HOLMA, 2004] HOLMA, Harri; TOSKALA, Antti. **WCDMA for UMTS. Radio Access for Third Generation Mobile Communications**. Editora Wiley. 3a edição. 2004.

[NETO, 2005] NETO, Durval Dourado, et. al. **Ajuste de modelos de distribuição de probabilidade a séries históricas de precipitação pluvial diária em Piracicaba-SP**. Revista Brasileira Agrometeorologia. V. 13 n.2, p.273-283. 2005. ISSN 0104-1347.

[LI, 2006] LI L., LIN, W.-H., LIU, H.. **Type-2 fuzzy logic approach for short-term traffic forecasting**. In Intelligent Transport Systems, IEE Proceedings Vol.153, 33-40, 2006.

[LIERS, 2006] LIERS, Florian, MITSCHLE-THIEL, Andreas **UMTS Data Capacity Improvements Employing Dynamic RRC timeouts**. 16th International Symposium on Personal, Indoor and Mobile Radio Communications. IEEE, 2006.

[VALE, 2006] VALE, Eduardo Rodrigues. **Técnicas para a melhoria de desempenho da transmissão de voz no sistema UMTS**. Rio de Janeiro: PUC, Departamento de Engenharia Elétrica, 2006. Tese Doutorado.

[ALECRIM, 2006] ALECRIM, Paulo Dias de. **Uma proposta para a provisão de QoS em redes Backbone UMTS**. Universidade Federal de Uberlândia. Faculdade de Engenharia Elétrica. Abril, 2006. Dissertação de Mestrado.

[KAY, 2006] KAY, Steven M. **Intuitive Probability and Random Process Using MATLAB**. Editora Springer, 2006.

[RICCIATO, 2006] RICCIATO, Fabio. **Unwanted Traffic in 3G Networks**. ACM SIGCOMM Computer Communication Review. Volume 36, number 2, April 2006.

[LEE-WOO, 2007] LEE, Patrick P. C.; BU, Tian, and WOO, Thomas. **On the Detection of Signaling DoS Attacks on 3G Wireless Networks**. 26th IEEE International Conference on Computer Communications. 2007.

[HAVERINEN, 2007] HAVERINEN, Henry; SIREN, Jonne; ERONEN, Pasi. **Energy Consumption of Always-On Applications in WCDMA Networks**. In Proceedings of the 65th Semi-Annual IEEE Vehicular Technology Conference (Dublin, Ireland). April 2007.

[ARJONA, 2007] ARJONA, Andres; YLÄ-JÄÄSKI, Antti. **VoIP Call Signaling Performance and Always-On Battery Consumption in HSDPA, WCDMA and WiFi**. In: Proceedings of the

3rd International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2007). Shanghai, China. 21-23 September 2007, pages 2964-2967.

[PENG, 2008] PENG, Mugen. WANG, Wenbo. **An Adaptive Energy Saving Mechanism in the Wireless Packet Access Network**. Proceedings of the IEEE WCNC 2008 - Wireless Communications and Networking Conference, pp. 1536-1540, Las Vegas, USA, Apr. 2008.

[ERONEN, 2008] ERONEN, Pasi. **TCP Wake-Up: Reducing Keep-Alive Traffic in Mobile IPv4 and IPsec NAT Traversal**. NOKIA Research Center NRC-TR-2008-002.

[LEE, 2009] LEE, Chi-Chen; YEH, Jui-Hung; CHEN, Jyh-Cheng. **Comparative Analysis of Energy-Saving Techniques in 3GPP and 3GPP2 Systems**. IEEE Trans Veh Tech, 2009, 58: 432–448.

[PERÄLÄ, 2009] PERÄLÄ, P. H. J.; BARBUZZI, A.PENTIKOUSIS, Kostas. **Theory and Practice of RRC State Transitions in UMTS Networks**. 5TH IEEE Broadband Wireless Access Workshop. July 2009.

[HOLMA, 2010] HOLMA, Harri; TOSKALA, Antti. **WCDMA for UMTS – HSPA evolution and LTE**. 5ª edição. Editora Wiley. 2010.

[QIAN, 2010] QIAN, F., WANG, Z., GERBER, A., MAO, Z. M., SEN, S., and SPATSCHECK O. **Characterizing radio resource allocation for 3g networks**. In Proceedings of the 10th annual conference on Internet measurement, IMC '10, pages 137–150, 2010.

[TITO,2011] OLIVEIRA, Tito R. B.; TIMOTEO, Varese S.; URSINI, Edson L. **Simulation based assessment of energy consumption optimization in 3G mobiles**. In Proceedings of International Conference on Computer and Management - CAMAN. Wuhan, China, May 2011.

[OLIVEIRA, 2011] OLIVEIRA, Tito R. B.; TIMOTEO, Varese S.; URSINI, Edson L. **Simulation based assessment of energy consumption optimization in 3G mobiles**. In Proceedings of 2nd National Conference on Telecommunications – CONATEL. Arequipa, Peru, May 2011.

[PINOTTI, 2011] PINOTTI, L. Fernando, OLIVEIRA, Tito R. B.; TIMOTEO, Varese S.; URSINI, Edson L. **An IP-Based multimedia traffic generator**. In proceedings of the international workshop on telecommunications – IWT/2011. Rio de Janeiro, Brazil, March/ 2011.

[TS25331] ETSI. **Universal Mobile Telecommunications System (UMTS) Radio Resource Control (RRC) Protocol specification**. 3GPP TS 25.331. Version 9.1.0, Release 9.

[UMTS2330] UMTS 23.30: **"Universal Mobile Telecommunications System (UMTS) Iu Principles"**

[TS25303] 3GPP TS 25.303: **"Technical Specification Group Radio Access Network; Interlayer procedures in Connected Mode"**.

[WSHK] Disponível em: <http://www.wireshark.org/>. Acessado em out/2010.

[TCPDMP] Disponível em: <http://www.tcpdump.org>. Acessado em out/2010.

[ADBTCPDMP] Disponível em: <http://prem-android.blogspot.com/2009/03/capture-packets-in-android-tcpdump.html>. Acessado em out/2010.

[ANDSDK] Disponível em: <http://developer.android.com/sdk/index.html>. Acessado em jun/2011.

[ADB] Disponível em: <http://developer.android.com/guide/developing/tools/adb.html>. Acessado em jun/2011.

APÊNDICE A – ARQUITETURA DE REDE DA TERCEIRA GERAÇÃO CELULAR

A.1. Arquitetura do Sistema UMTS

O Sistema UMTS (*Universal Mobile Terrestrial System*) é composto por três principais componentes, como pode ser observado na Figura A.1:

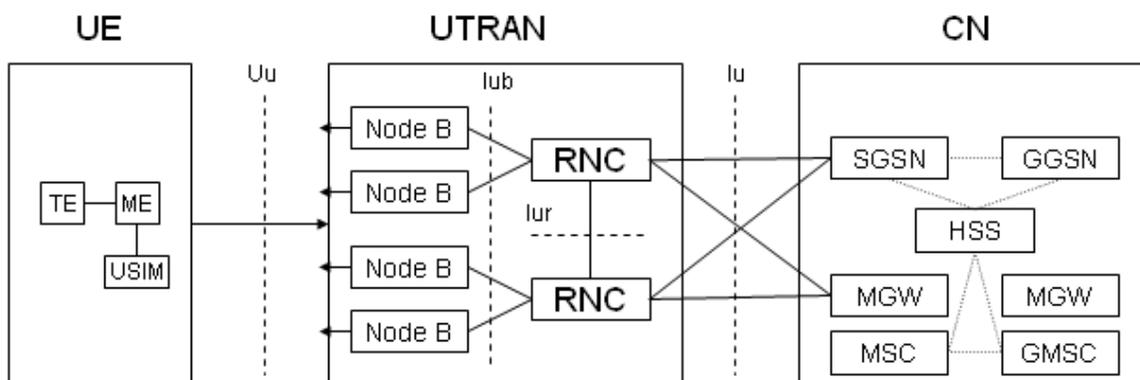


Figura A.1. Estrutura básica do sistema UMTS. [VALE, 2006]

- **UE (User Equipment):** Composto por TE (*Terminal Equipment*), ME (*Mobile Equipment*) e USIM (*Universal Subscriber Identity Module*).

A interface Uu é a responsável pela conexão entre a UTRAN e a UE.

- **UTRAN (UMTS Terrestrial Radio Access Network):** Responsável pelo acesso de rádio da rede UMTS. Composta por:

- **NodeB:** Transmissor e receptor de rádio frequência para a UE.
- **RNC (Radio Network Controller):** Responsável pelo controle de recursos de rádio e gerenciamento de mobilidade de diversos NodeB conectados a ele. É conectado ao CS (*Circuit Switch*) e ao PS (*Packet Switch*) do núcleo da rede, para comunicação de voz e dados, respectivamente.

As interfaces presentes na UTRAN são:

Uu – Interface de rádio entre a UTRAN e o UE.

Iub – Interface interna que conecta o RNC ao NodeB

Iur – Interface interna ou externa, dependendo da arquitetura. Responsável pela conexão entre duas RNCs.

Iu – Interface externa que conecta o RNC ao *Core Network*

➤ **CN (*Core Network*):** É o núcleo da rede. Sua arquitetura é baseada na rede GSM com GPRS (*General Packet Radio Service*). Podemos dividir o núcleo da rede em dois domínios:

- **CS (*Circuit Switched*):** Elementos comutados a circuito. Composto por:
 - **MGW (*Media Gateway*):** Possui as funções de um *gateway* H.248, sendo controlado por um *GMSC Server* (*GPRS Mobile Switching Center Server*).
 - **MSC Server (*Mobile Switching Center Server*):** Opera como um *Media Gateway Controller*, controlando todos os MGWs associados.
 - **GMSC Server:** *Media Gateway Controller* que realiza interface com a RTPC (Rede de Telefonia Pública Comutada).
 - **HSS (*Home Subscriber Service*):** Possui a base de dados dos usuários. Provê suporte a outros servidores para as funções de: Autenticação, Autorização e Localização de usuários.
- **PS (*Packet Switched*):** Elementos comutados a pacotes. Composto por:
 - **SGSN (*Serving GPRS Support Node*):** módulo central na comutação de pacotes. Tem como funcionalidades:
 - Detectar e registrar novos usuários na rede;
 - Criptografia;
 - Manutenção dos registros e localização dentro da área de uso;
 - Gerenciamento de mobilidade - Faz o gerenciamento da sessão e é responsável por providenciar e realizar a comutação para uma outra SGSN em caso de *handover*, em que a área de destino seja gerenciada por outra SGSN;
 - Compressão de dados;
 - Requisição e resposta para o HLR
 - Contagem de pacotes (tarifação)
 - **GGSN (*Gateway GPRS Support Node*):** Responsável pela interconexão entre a rede GPRS e as redes externas de pacote, como Internet e X.25. Recebe dados da rede externa destinada a um equipamento conectado na rede, e transfere o pacote para a

SGSN, que por sua vez irá transferi-lo à UE. Também é responsável por receber pacotes provenientes da UE e fazer o roteamento correto para a rede externa.

O processo de transferência de pacotes entre o GGSN e o SGSN é feito através de tunelamento, onde o GGSN recebe o pacote de dados da rede de pacotes, encapsula e manda para o SGSN através de um *backbone* GPRS, via PDU (*Protocol Data Unit*). O GGSN deve conter o endereço do SGSN o qual o UE está registrado para envio correto dos pacotes.

Para o envio na rede externa, o GGSN converte o pacote proveniente do SGSN para o formato PDP (*Packet Data Protocol*), que pode ser IP ou X.25.

Protocolos:

- **GTP (GPRS Tunneling Protocol):** Utilizado no chaveamento de dados em uma rede GPRS, podendo ser utilizados nas interfaces Gn e Gp para comunicação entre SGSN e GGSN. É responsável pelo encapsulamento dos pacotes IP ou X.25 a serem encaminhados entre os módulos.
- **GTP-C(GPRS Tunneling Protocol-Control):** É a seção de controle do padrão de GTP. O SSGN emite uma mensagem de contexto de ativação GTP-C para o GGSN quando uma ME pede um contexto de PDP. O GGSN responde se a ativação foi efetuada com sucesso ou falhou.
- **GTP-U (GPRS Tunneling Protocol-User):** Baseado em protocolo IP para tunelamento simples, permitindo vários túneis entre cada equipamento fim-a-fim. Quando utilizado cada usuário terá um ou mais túneis, um para cada contexto de PDP ativo.

Interfaces GPRS

Gb: Interface entre a BSS (*Base Station Subsystem*) e o SGSN. Pode utilizar os protocolos *Frame Relay* ou IP para transmissão. Provê transferência de dados e sinalização entre o PCU e o GGSN através do protocolo *Frame Relay*.

Gn: Interface baseada em IP entre um ou mais SGSNs a um GGSN interno em uma mesma PLMN (*Public Land Mobile Network*). Utiliza o protocolo GTP. Este protocolo não foi projetado focado em segurança, não garantindo a integridade e autenticidade dos dados que trafegam entre dois elementos de rede.

Ataques a essa interface podem:

- Provocar queda de desempenho, perda de serviços, até mesmo tirar a rede de operação.
- Confundir o SGSN ou o GGSN, onde o invasor tenta se passar por usuário e tenta enganar o SGSN com IPs que possivelmente façam parte da rede.
- Ataques entre usuários móveis, quando usuários mal intencionados decidem enviar quantidades excessivas de pacotes para outros usuários.

Gp: Interface entre SGSN interno e GGSN externo. Entre eles há um *Border Gateway* (que é um *Firewall*). Também utiliza o protocolo GTP. Os dados que trafegam pela interface estão expostos aos mesmos tipos de ataque que ocorrem na *web*, como vírus, *worms*, cavalos de tróia, etc. Usuários mal intencionados podem alvejar um elemento um elemento da rede como o SGSN ou um usuário do sistema para indiretamente prejudicar a própria operadora no desempenho da rede. A interface Gp também está exposta a ataques como DoS (*Denial of Service*).

A segurança do Gp é feita através do gerenciamento de *Firewalls*, permitindo o provedor utilizar um modelo arbitrário para proteger ataques dentro da rede. Outra maneira consiste em uma inspeção total do *Firewall* e definir a política de bloqueio ou liberação de determinado tipo de tráfego. A política de segurança da interface também inclui proteção relativa à operação do sistema contra invasão do sistema da própria operadora.

Gr: Interface entre o SGSN e o HLR (*Home Location Register*). Mensagens trafegadas nesta interface utilizam o protocolo MAP3 (*Mobile Application Part*).

Gd: Interface entre o SGSN e o SMS (*Short Message Service*) Gateway. Pode utilizar MAP1, MAP2 ou MAP3.

Gs: Interface entre o SGSN e o MSC. Permite paginação e avaliação da estação quando executa transferência de dados. Quando a estação é conectada a rede GPRS, o

SGSN mantém rastreamento de qual RA (*Routing Area*) a estação está conectada. Um RA é uma parte do LA (*Location Area*).

Gi: Interface baseada em IP entre o GGSN e a rede pública de dados (*Public Data Network* - PDN) ou diretamente a Internet, ou através de um WAP (*Wireless Application Protocol*) Gateway. Semelhantes a interface Gp, DoS são os ataques mais comuns nessa rede.

Ge: A interface entre o GGSN e o SCP (*Service Control Point*).

Gx: Interface de policiamento de tempo real entre a GGSN e o CRF (*Charging Rule Function*). Provê serviço de dados baseado nas regras de tarifação.

Gy: Cobrança em tempo real entre o GGSN e o OCS (*Online Charging System*)

Gmb: Interface entre o GGSN e o BM-SC (*Broadcast-Multicast Service Center*) usado para controle de portadoras MBMS (*Multimedia Broadcast Multicast Service*).

Na Figura A.2 é apresentada a disposição na rede das principais interfaces.

- **MRF (*Media Resource Functions*):** Provê funções relacionadas à realização de Conferências Multimídia.
- **MGCF (*Media Gateway Control Function*):** Atua como um controlador de mídia H.248, realizando a interface entre PS e a RTPC.
- **CSCF (*Call State Control Function*):** Funciona como um Proxy Server, em uma configuração SIP (*Session Initiation Protocol*)
Existem três tipos de CSCF:
 - **P-CSCF (*Proxy CSCF*):** Envia e recebe sinalização SIP ao UE ao qual está conectado.
 - **I-CSCF (*Interrogation CSCF*):** Localizado na rede de origem (*Home Network*) do UE. É o elemento de conexão entre o P-CSCF e o S-CSCF.
 - **S-CSCF (*Server CSCF*):** Fornece informações sobre o UE quando solicitadas.
- **PDF (*Policy Decision Function*):** Controla os recursos da rede.

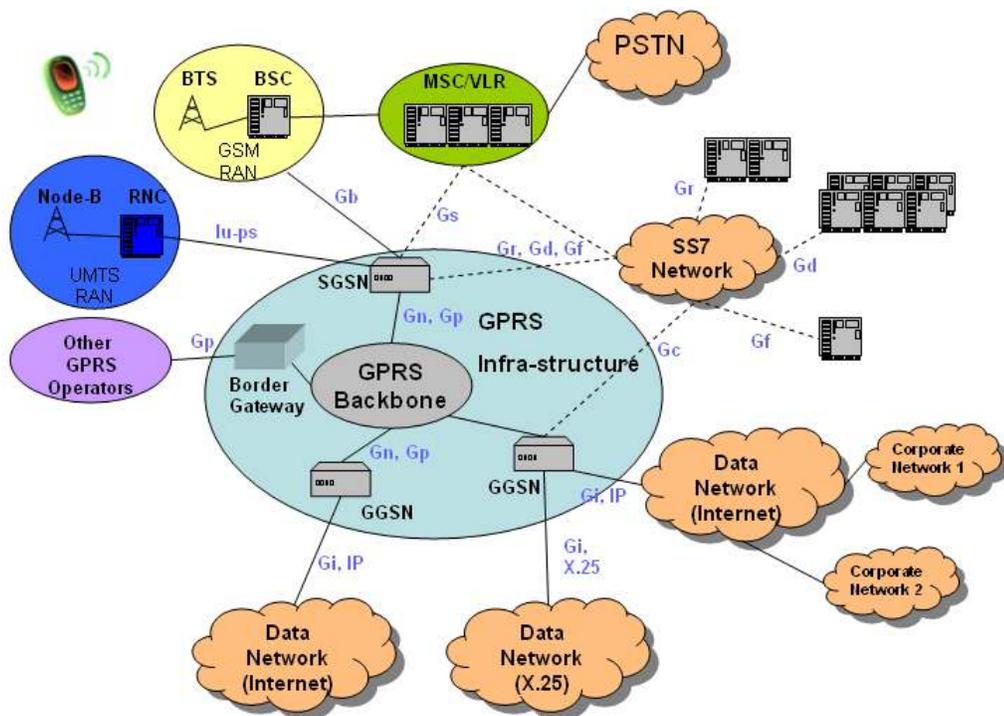


Figura A.2. Principais interfaces e sua ligação com elementos de rede.

Na Figura A.3 são apresentados com mais detalhes os equipamentos do núcleo da rede UMTS.

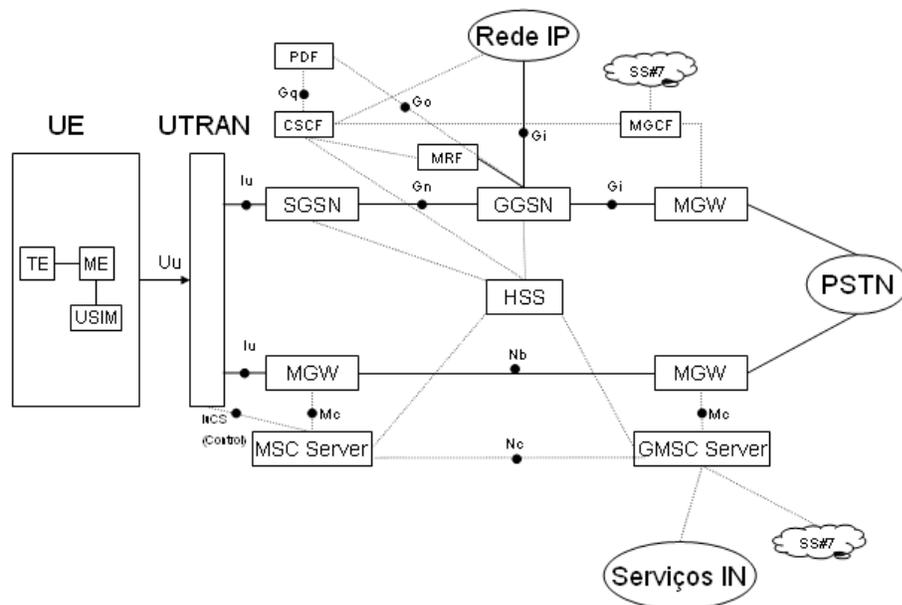


Figura A.3. Detalhamento do núcleo da rede do sistema UMTS [VALE, 2006].

A.1.1. Interface de Protocolos de Rádio

A interface de protocolo de rádio UMTS consiste de duas estruturas: Vertical e Horizontal:

A estrutura Vertical, como pode ser visto na Figura A.4, indica os Planos do Usuário (*U-Plane*), responsável por transmissão de dados do usuário, chamadas de voz e pacotes de dados, e Plano de Controle (*C-Plane*), responsável pela sinalização.

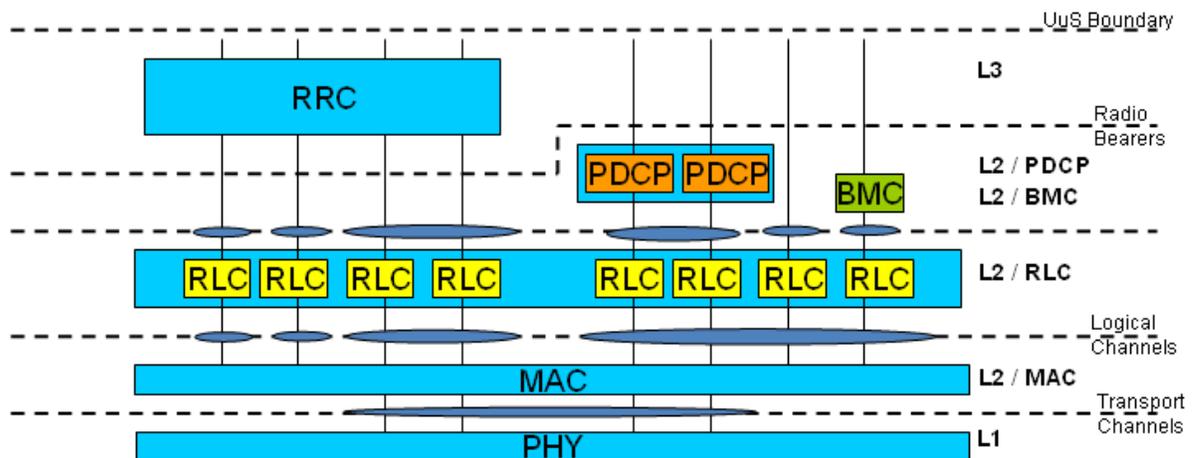


Figura A.4. Estrutura vertical da interface aérea do sistema UMTS.

A estrutura Horizontal indica a representação dos protocolos em 3 camadas (*layers*): *Layer 1* – Camada Física (*Physical Layer*), *Layer 2* – Camada de Link de dados (*Data Layer*) e *Layer 3* – Camada de Rede (*Network Layer*).

A camada física (*Physical layer* - PHY) oferece serviço à camada MAC (*Medium Access Control*) via canais de transporte.

A camada MAC oferece serviço à camada RLC (*Radio Link Control*) por meio de canais lógicos.

A camada RLC oferece serviço às camadas superiores via *Service Access Points* (SAPs) que descreve como o RLC manipula os pacotes de dados e se, por exemplo, um *Automatic Repeat Request* (ARQ) é utilizado. No plano de controle o RLC é usado para transporte de sinalização. No plano do usuário é usado por serviços específicos das camadas de protocolo PDCP (*Packet Data Convergence Protocol*) ou BMC (*Broadcast/Multicast Control*), ou por camadas mais altas do *U-plane*.

A camada *Packet Data Convergence Protocol* (PDCP) existe apenas para serviços de pacotes (*Packet Switched* - PS). Sua principal função é compressão de cabeçalho.

O protocolo BMC (*Broadcast/Multicast Control*) é responsável pelo *broadcast/multicast* de mensagens para os UEs.

O RRC (*Radio Resource Control*) será tratado com mais detalhes no decorrer deste capítulo.

A.1.2. Canais UMTS

Os canais do sistema UMTS são divididos em três níveis, pelos quais são mapeados entre si: Nível Lógico, de Transporte e Físico, como mostra a Figura A.5.

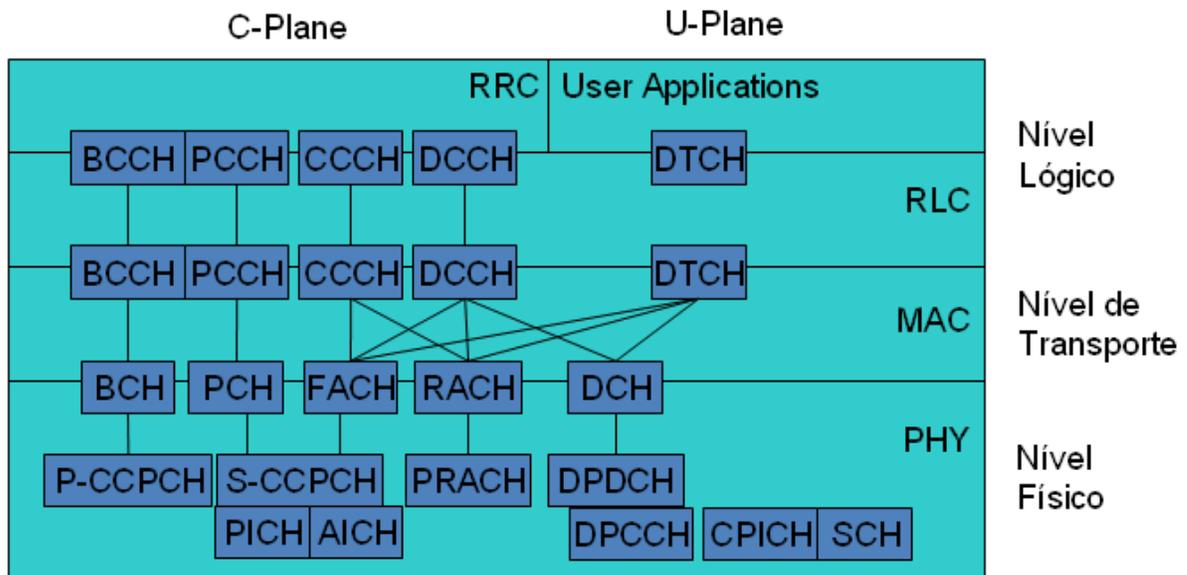


Figura A.5. Planos vertical e horizontal do sistema UMTS.

Canais de Transporte:

Os canais de transporte podem ser classificados em dois grupos: Canais comum e dedicado [CASTRO, 2001]. A Tabela A.1 apresenta um resumo das características dos canais.

TABELA A.1. CANAIS DE TRANSPORTE DO SISTEMA UMTS.

Canal	Grupo	Direção	Descrição
RACH (<i>Random Access Channel</i>)	Comum	<i>Uplink</i>	Usado para transmissão de pequenas quantidades de dados
FACH (<i>Forward Access Channel</i>)		<i>Downlink</i>	Usado para transmissão de quantidade relativamente baixa de tráfego
PCH (<i>Paging Channel</i>)		<i>Downlink</i>	Difusão de informações de controle permitindo modo eficiente de estado de dormência do UE.
BCH (<i>Broadcast Channel</i>)		<i>Downlink</i>	Usado para difusão de informação de sistema em toda a célula
DCH (<i>Dedicated Channel</i>)	Dedicado	<i>Uplink/Downlink</i>	Canal dedicado para recebimento e envio de informações

Canais Lógicos:

A camada MAC provê serviço de transferência de dados nos canais lógicos. Os canais podem ser classificados em dois grupos: Canais de controle (C-Plane) e de tráfego (U-Plane) [CASTRO, 2001]. A Tabela A.2 apresenta um resumo das características dos canais.

TABELA A.2. CANAIS LÓGICOS DO SISTEMA UMTS

Canal	Grupo	Direção	Descrição
BCCH (<i>Broadcast Control Channel</i>)	Controle	<i>Downlink</i>	Difusão de informações de controle do sistema
PCCH (<i>Paging Control Channel</i>)		<i>Downlink</i>	Canal de transferência de informação de paginação
DCCH (<i>Dedicated Control Channel</i>)		<i>Uplink/Downlink</i>	Canal ponto-a-ponto que transmite informação de controle dedicado entre UE e a rede
CCCH (<i>Common Control Channel</i>)		<i>Uplink/Downlink</i>	Transmite controle de informação entre a rede e UE.
DTCH (<i>Dedicated Traffic Channel</i>)	Tráfego	<i>Uplink/Downlink</i>	Canal ponto-a-ponto dedicado a um único UE, para transferência de informação ao usuário.

Canais Físicos:

P-CCPCH (*Primary Common Control Physical Channel*): Transporta o BCH. Necessita ser demodulado por todos os terminais no sistema.

S-CCPCH (*Secondary Common Control Physical Channel*): Transporta dois diferentes canais de transporte comum: o FACH e o PCH. Os dois canais podem compartilhar um mesmo S-CCPCH ou podem usar diferentes canais físicos.

PRACH (*Physical Random Access Channel*): Canal onde o RACH é mapeado. Utiliza fatores de espalhamento 16 e 8.

AICH (*Acquisition Indicator Channel*): Em conjunto com o RACH, é utilizado para indicar da estação base a recepção de uma sequência do RACH. O AICH utiliza uma sequência de assinatura como o RACH em um dos códigos de canalizações do *downlink* da estação base o qual o RACH pertence.

PICH (*Paging Indicator Channel*): O PCH é operado junto como o PICH para prover modo de operação de dormência. O indicador de *paging* ocorre uma única vez por *slot*.

SCH (*Synchronization Channel*): Necessário para busca de célula. Consiste de dois canais, um primário e outro secundário. O P-SCH utiliza uma sequência de espalhamento de 256 *chips* idênticos em cada célula. O S-SCH utiliza uma sequência com diferentes combinações de palavras de códigos representando diferentes grupos de códigos.

DPDCH (*Dedicated Physical Data Channel*): Mapeia o canal DCH para transporte de informação das camadas mais altas, incluindo dados do usuário.

DPCCH (*Dedicated Physical Control Channel*): Mapeia o canal DCH para transporte de informação referente informações de controle.

CPICH (*Common Pilot Channel*): Sua função é auxiliar a estimação de canal no terminal para os canais dedicados e prover a estimação de canal de referência para os canais comuns quando eles não estão associados com os canais dedicados.